Integrated Inventory Routing and Freight Consolidation for Perishable Goods

Weihong Hu Alejandro Toriello
H. Milton Stewart School of Industrial and Systems Engineering
Georgia Institute of Technology, Atlanta, GA
weihongh *at* gatech *dot* edu, atoriello *at* isye *dot* gatech *dot* edu

Maged Dessouky
Daniel J. Epstein Department of Industrial and Systems Engineering
University of Southern California, Los Angeles, CA
maged *at* usc *dot* edu

March 29, 2017

Abstract

We propose a model that integrates inventory routing and freight consolidation for perishable goods with a fixed lifetime. The problem is motivated by the status quo of logistics in some U.S. agriculture markets, but adapts to other relevant two-echelon supply chains, e.g. combined production planning and distribution. We first identify special cases where solving single-echelon subproblems sequentially yields an optimal solution. For the general case, we propose an iterative framework that consists of a decomposition procedure and a local search scheme. In the decomposition, a freight consolidation subproblem is first solved to obtain crucial shipping decisions, and after fixing these a restrictive model generates the other decisions for the integrated problem. The local search aims at fast identification of good neighborhoods by solving an assignment-style mixed-integer program that matches the consolidation decision with an inventory routing subproblem, and gradually strengthens the incumbent solution pool when executed in an iterative fashion. Experiments based on empirical demand distributions demonstrate that our proposed iterative framework is remarkably efficient compared to a sequential approach, and our approach is effective in solving challenging instances.

1 Introduction

Transportation is frequently the single largest element of total logistics cost, accounting for 50%-65% [22]. A supplier often achieves efficient use of transportation assets by intelligently routing a fleet to serve multiple customers. When transportation decisions are coupled with inventory management, the resulting decisions are modeled by the *inventory routing problem* (IRP). Some

recent examples of successful IRP applications include Walmart's *vendor-managed inventory* (VMI) initiative [6] and ExxonMobil's liquefied natural gas transportation and storage [31].

Nowadays, ever-increasing global competition encourages transportation cost savings through horizontal cooperation within a supply chain echelon, such as when a warehouse consolidates different suppliers' orders to schedule outbound shipments at minimum total freight cost. One concrete example is given by the plight of the California cut flower industry, where growers ship their product individually to U.S. customers, and because of small disaggregated volumes, must usually incur expensive *less-than-truckload* (LTL) or courier (e.g. FedEx, UPS) rates. In the last two decades, this disadvantage has played a major role in California's loss of over 40% of U.S. market share to South American growers, who enjoy favorable *full truckload* (FTL) rates by aggregating products prior to domestic distribution [30]. These authors estimate that California cut flower growers could reduce annual transportation costs by \$6 to \$17 million via consolidation, depending on the number of participating growers.

Vertical coordination across supply chain echelons is an additional avenue to reduce transportation costs. Although consolidation reduces outbound transportation costs with higher utilization of shipping capacity, an excessive emphasis on it could restrict inbound shipping, and indirectly lead to higher total costs. Moreover, it is hard to evaluate the impact of consolidation strategies on inventory control at a facility without considering both echelons together. Therefore, many agricultural sectors are realizing the collaboration imperative for growers, consolidation terminals, sellers, and third-party carriers.

Cooperation and coordination offer benefits in almost every area, from procurement to last-mile delivery. Besides the aforementioned truckload costs, economies of scale prevail when a vendor offers batch ordering discounts, when a factory processes identical jobs on heterogeneous machines, when a liner company ships different customers' cargoes, etc. The ideal performance of the supply chain could entail joint management of cross-functional activities such as production, transportation, consolidation and inventory.

This paper integrates two classical problems that are typically solved independently in these and similar supply chains. As interpreted in agriculture logistics, the first decision includes the shipments and routes from local growers to a consolidation center, which is the short-haul problem and can be modeled as an IRP; trucks are the only form of agriculture transportation in at least 80% of U.S. cities and communities [35], and thus the short-haul problem will generally involve a routing component. The second decision concerns direct shipments from the consolidation center to customers (retailers or wholesalers), the long-haul freight consolidation problem. The two separate models are already difficult and involve real-world complications like time-varying demand, perishability, different truckload costs, routing capacities and duration limits. We propose the integrated problem, formulate it as a *mixed-integer programming* (MIP), and develop an iterative framework with a decomposition procedure and a local search scheme to obtain solutions of high quality in a reasonable amount of time. In the decomposition, a freight consolidation subproblem is

first solved to obtain long-haul FTL schedules, and after fixing these a restrictive model generates the other decisions for the integrated MIP. The local search aims at fast identification of good neighborhoods by solving an assignment-style MIP that matches the consolidation decision with an IRP subproblem, and gradually strengthens the incumbent solution pool when executed in an iterative fashion.

In the agricultural industry, long-haul transportation costs usually dominate the total distribution cost, which intuitively suggests a consolidation-based strategy. Hence, the "standard" approach would be to first solve the long-haul freight consolidation problem, and use its shipping schedule as the demand input for the short-haul IRP. However, our results show that overall costs can be reduced by utilizing a system-wide optimization approach. We also demonstrate the potential of the iterative framework in balancing solution efficacy and computational efficiency. The standard approach tends to exceed 5 hours while solving a moderately-sized IRP subproblem, whereas our proposed method yields near-optimal global solutions by exploring more neighborhoods in the same amount of CPU time. When the planning horizon is longer, or there are more growers or sellers, our approach can provide better solutions than CPLEX upper bounds as well as the standard approach, and the advantage is enhanced as the problem size increases. On the other hand, we find that the standard approach, or the seemingly counterintuitive "reverse" approach (which first optimizes the IRP and uses its schedule as input for the consolidation subproblem), are asymptotically optimal under some assumptions.

The remainder of the paper is organized as follows. We provide a brief literature review below in §1.1. We formally state the problem in §2. We motivate and detail the iterative heuristic in §3. We present experimental results in §4. We discuss potential future research and conclude in §5. The appendix gives additional details not included in the body of the paper.

1.1 Literature Review

Both inventory routing and freight consolidation problems have drawn extensive attention in the operations research community. The IRP simultaneously decides (1) when a central facility dispatches vehicles, (2) which customers a vehicle visits and in what order, and (3) how much demand to fulfill or inventory to maintain at each facility. Over the past thirty years, numerous IRP models have been studied with quite specific problem characteristics, among which the closest variants to our short-haul problem are the class of finite horizon multi-period single-vehicle one-to-many IRP. The solution techniques range from exact methods to meta-heuristics to optimization-based heuristics. We refer the reader to [19] and [6] for comprehensive surveys on state-of-the-art methodologies and industrial applications, respectively.

Freight consolidation addresses the question of how much volume or how many time periods to accumulate before releasing a shipment that leverages economies of scale in transportation costs. Since the pioneering work of [13, 23], quantity-based, time-based and hybrid policies have been investigated; see [16] for a summary. In particular, [29, 30] proposed optimal and near-optimal

algorithms to solve the version of our long-haul problem without inventory aspects. When inventory is taken into account, the long-haul decision can be viewed as a *lot-sizing problem* (LSP) where the ordering cost function is *piecewise linear* (PWL), as depicted in Figure 1. Related research includes the LSP with multiple set-up costs [5, 20, 28] and volume discounts [8, 17], but much of this work usually assumes concave or monotonic properties that do not generalize to our case.

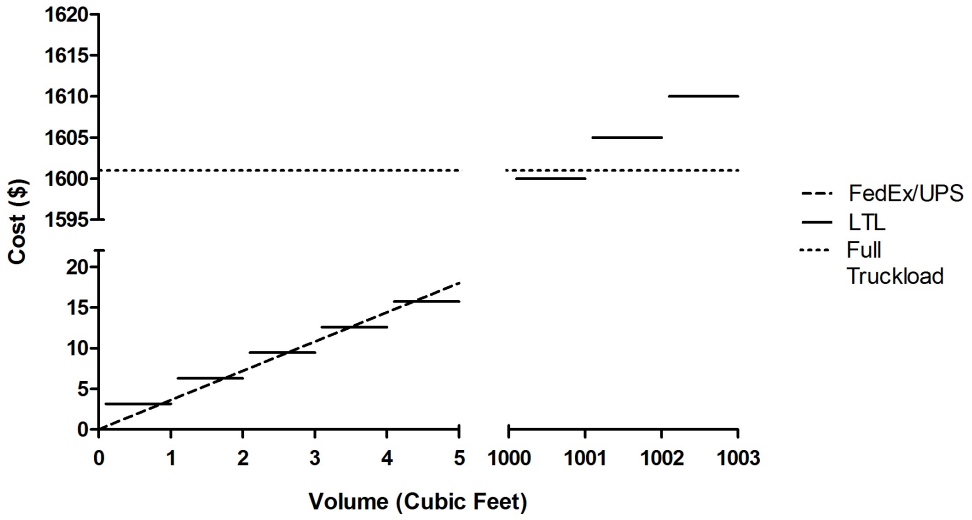


Figure 1: Volume-dependent long-haul shipping costs

The literature on the integration of inventory, routing, and consolidation is comparatively sparse. Representative problems in this vein are the *production routing problem* (PRP) [4] and the *maritime inventory routing problem* (MIRP) [31]. PRP coordinates two core supply chain functions, namely production planning and distribution, which are the origins of LSP and IRP, and thus resemble our long-haul and short-haul decisions, respectively. However, the few publications on this subject concern much simpler settings, e.g. specified inventory replenishment rules [1, 14], uncapacitated production and/or routing [32], fixed-charge or PWL concave production costs [3, 9, 10, 11, 12, 14, 15]. Furthermore, existing PRP models largely consider a homogeneous product from a single plant to customers that do not restrict service time windows. In contrast, we treat the demand for each grower-seller pair as an individual commodity, and impose hard constraints on the time in transit or storage to accommodate the perishable nature of the agricultural product. On the other hand, MIRP typically involves multiple commodities, loading/discharging time slots and intricate cost elements, but fundamentally differs in network topology and does not share our cost structure. Another related but less studied problem is the integration of various transportation modes in the conventional IRP, e.g. when a main route serves a few major customers who then transship the goods to minor customers via direct shipping [18].

From the perspective of solution approaches, many exact and heuristic IRP algorithms have been successfully extended to MIRP and IRP with transshipment, whereas the PRP literature centers more on heuristics. Early attempts are primarily meta-heuristics, especially those very

powerful in vehicle routing, e.g. GRASP [14], memetic algorithms [15], tabu search [9, 11]. Metaheuristics are capable of tackling PRP instances up to 200 customers and 20 time periods [14], but the implementation is difficult because of the increased complexity in combined decisions and constraints. Exact methods have also been developed, such as branch-and-price algorithms [12], branch-and-cut algorithms [2, 7, 18, 32] and Lagrangian relaxations [21, 33], which unfortunately are effective only for more basic variants and smaller problem instances. A seemingly promising alternative is thus optimization-based heuristics. For example, [3] propose a hybrid *adaptive large neighborhood search* (ALNS) scheme where upper-level search operators handle binary setup and routing decisions, and lower-level network flow problems yield the corresponding production, inventory and shipping quantities. Recently, [1] introduced a new two-phase scheme which considers a lot-sizing problem with approximated routing costs first and a routing problem subsequently. Both approaches exploit diversification mechanisms to prevent fast convergence to local optima in an iterative fashion, and outperform previous methods for most benchmark instances with 14 to 200 customers and 6 to 20 time periods.

2 Problem Statement

We consider the distribution of a perishable product with a fixed lifetime and deterministic demand over a discretized finite horizon. Figure 2 illustrates the distribution network. The product is moved from local growers to a consolidation center via short-haul routing, and then from the consolidation center to geographically dispersed sellers via long-haul direct shipping. We assume one vehicle available per period for local routing, but allow more expensive direct shipping alternatives for possible excess demand. We differentiate long-haul delivery options by volume-dependent services, including fixed FTL rates per truck, fixed LTL rates per LTL unit (e.g. cubic foot), and linear courier rates per pound. The growers and the consolidation center may keep inventory to delay pickup or delivery at facility-specific holding costs, but all commodities must leave the system before spoilage. We next introduce the model's notation.

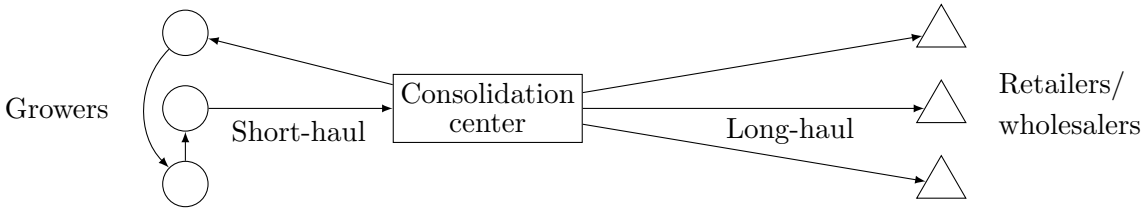


Figure 2: Integrated distribution network

Parameters

T: length of the planning horizon.

θ: product lifetime.

$G, 0, D$: set of growers, the consolidation center, and set of sellers, respectively.

$m = M_{ik}^s$: a commodity tuple of product that is ready for pickup at the beginning of period s and will be moved from grower i to seller k, $\forall i \in G, k \in D, 1 \leq s < T$. $M = \{M_{ik}^s\}$ denotes the set of all commodity tuples, and $\cdot$ indicates all elements with the respective indices. E.g., $M_{i\cdot}^{\cdot}$ is the set of commodities that originate from grower i.

d_m: demand for commodity m.

Q: local vehicle capacity.

l_{ij}: local travel time from facility i to facility j, $\forall i, j \in G \cup \{0\}, i \neq j$, scaled so that a route's maximum duration is one (i.e. one period in the planning horizon).

c_{ij}: local mileage cost from facility i to facility j, $\forall i, j \in G \cup \{0\}, i \neq j$.

B: alternative short-haul direct shipping cost per shipment.

$\{F, L, U\}$: set of long-haul direct shipping modes, where F, L, U represent FTL, LTL and courier services, respectively.

$\mathcal{K}_F, \mathcal{K}_L$: maximum capacities in cubic feet for a FTL truck and a LTL unit, respectively.

c_{kF}, c_{kL}: transportation costs for an FTL truck and an LTL unit, respectively, from the consolidation center to seller k, $\forall k \in D$.

α: conversion factor (lbs. per cubic foot).

c_{kU}: transportation cost (per pound) for a courier shipment to seller k, $\forall k \in D$.

h_i: inventory holding cost per unit product per period at facility i, $\forall i \in G \cup \{0\}$.

The goal is to minimize the total transportation and inventory cost while satisfying all demands before spoilage, and respecting local vehicle capacity, routing duration limits, as well as long-haul shipping capacities for each FTL truck or LTL unit. We propose the MIP formulation below.

Decision variables

$$x_{ij}^t = \begin{cases} 1, & \text{arc } (i,j) \text{ is traversed in period } t \\ 0, & \text{otherwise} \end{cases}, \quad \forall i, j \in G \cup \{0\}, i \neq j, 1 \leq t < T.$$

$$y^t = \begin{cases} 1, & \text{a local trip occurs in period } t \\ 0, & \text{otherwise} \end{cases}, \quad 1 \leq t < T.$$

u_i^t: number of alternative local vehicles used by grower i in period t, $\forall i \in G, 1 \leq t < T$.

q_m^t: short-haul pick-up volume of commodity m in period t, $\forall m \in M_{\cdot}^s, \max\{0, t-\theta\} < s \leq t, 1 \leq t < T$.

v_m^t: volume of commodity m picked up in period t via alternative short-haul direct shipping, $\forall m \in M_{\cdot}^s, \max\{0, t-\theta\} < s \leq t, 1 \leq t < T$.

f_{ijm}^t: short-haul flow volume of commodity m on arc (i,j) in period t, $\forall i,j \in G \cup \{0\}, i \neq j, m \in M_{\cdot}^s, \max\{0, t-\theta\} < s \leq t, 1 \leq t < T$.

z_{mp}^t: long-haul delivery volume of commodity m with mode p in period t, $\forall p \in \{F, L, U\}, m \in M_{\cdot}^s, \max\{1, t-\theta\} \leq s < t, 1 < t \leq T$.

r_{kp}^t: FTL numbers or LTL units sent to seller k in period t, $\forall k \in D, p \in \{F, L\}, 1 < t \leq T$.

I_{im}^t: grower inventory of commodity m at the end of period t, $\forall i \in G, m \in M_{i\cdot}^s, \max\{0, t-\theta\} < s \leq t, 1 \leq t < T$.

I_{0m}^t: central inventory of commodity m at the end of period t, $\forall m \in M_{\cdot}^s, \max\{1, t-\theta\} \leq s < t, 1 < t \leq T$.

Objective function The total distribution cost includes three parts: (1a) the short-haul transportation cost, which equals regular arc routing costs plus alternative direct shipping costs; (1b) the long-haul transportation cost, which is the sum of FTL trucks, LTL units and courier volume multiplied by their respective dispatch cost rates; (1c) the inventory cost, measured with facility-specific holding cost rates.

$$\min \sum_{t=1}^{T-1} \sum_{i,j \in G \cup \{0\}, i \neq j} c_{ij} x_{ij}^t + \sum_{t=1}^{T-1} \sum_{i \in G} B u_i^t \tag{1a}$$

$$+ \sum_{t=2}^{T} \sum_{k \in D} \sum_{p \in \{F,L\}} c_{kp} r_{kp}^t + \alpha \sum_{t=2}^{T} \sum_{s=\max\{1, t-\theta\}}^{t-1} \sum_{k \in D} \sum_{m \in M_{\cdot k}^s} c_{kU} z_{mU}^t \tag{1b}$$

$$+ \sum_{t=1}^{T-1} \sum_{i \in G} h_i \sum_{s=\max\{1, t-\theta+1\}}^{t} \sum_{m \in M_{i\cdot}^s} I_{im}^t + h_0 \sum_{t=2}^{T} \sum_{s=\max\{1, t-\theta\}}^{t-1} \sum_{m \in M_{\cdot}^s} I_{0m}^t \tag{1c}$$

Route definition Arc degree constraints: Equalities (2a) and inequalities (2b) relate the node degrees in the short-haul network to whether a trip occurs in each period. Specifically, the outdegree of the consolidation center equals 1 if there is a trip, and 0 otherwise; the outdegree of a grower is no more than that of the consolidation center. Constraints (2c) balance each facility's indegree and outdegree.

Commodity flow constraints (2d)-(2e): For a commodity originating at a specific grower, the total outflow equals the total inflow plus the short-haul pickup volume in any period; for a commodity originating from other growers, the total outflow equals the total inflow. These constraints also eliminate sub-tours.

Vehicle capacity constraints (2f)-(2g): The total volume carried by a regular vehicle does not exceed its capacity on any arc traversed in each period; the total local direct shipping volume does not exceed the total capacity of alternative vehicles dispatched from a grower in any period.

Duration constraints (2h): The total duration of a tour lasts no more than one period.

$$\sum_{i\in G} x_{0i}^t = y^t, \quad 1 \le t < T \tag{2a}$$

$$\sum_{j\in G\cup\{0\}} x_{ij}^t \le y^t, \quad \forall i \in G, 1 \le t < T \tag{2b}$$

$$\sum_{j\in G\cup\{0\}, j\neq i} x_{ij}^t - \sum_{j\in G\cup\{0\}, j\neq i} x_{ji}^t = 0, \quad \forall i \in G\cup\{0\}, 1 \le t < T \tag{2c}$$

$$\sum_{j\in G\cup\{0\}, j\neq i} f_{ijm}^t - \sum_{j\in G, j\neq i} f_{jim}^t = q_m^t, \quad \forall i \in G, m \in M_{i\cdot}^s, \max\{0, t-\theta\} < s \le t, 1 \le t < T \tag{2d}$$

$$\sum_{j\in G\cup\{0\}, j\neq i} f_{ijm}^t - \sum_{j\in G, j\neq i} f_{jim}^t = 0, \quad \forall i \in G, m \notin M_{i\cdot}^s, \max\{0, t-\theta\} < s \le t, 1 \le t < T \tag{2e}$$

$$\sum_{s=\max\{1,t-\theta+1\}}^{t} \sum_{m\in M_{\cdot}^s} f_{ijm}^t \le Q x_{ij}^t, \quad \forall i, j \in G\cup\{0\}, i \neq j, 1 \le t < T \tag{2f}$$

$$\sum_{s=\max\{1,t-\theta+1\}}^{t} \sum_{m\in M_{i\cdot}^s} v_m^t \le Q u_i^t, \quad \forall i \in G, 1 \le t < T \tag{2g}$$

$$\sum_{i,j\in G\cup\{0\}, i\neq j} l_{ij} x_{ij}^t \le 1, \quad 1 \le t < T \tag{2h}$$

Demand satisfaction Short-haul (3a): For the short-haul echelon, the total pickup quantity of each commodity, including regular routing and alternative direct shipping volume, equals its demand. Long-haul (3b): For the long-haul echelon, the total delivery quantity of each commodity, including FTL, LTL and courier volume, equals its demand.

$$\sum_{t=s}^{\min\{s+\theta,T\}-1} (q_m^t + v_m^t) = d_m, \quad \forall m \in M_{\cdot}^s, 1 \le s < T \tag{3a}$$

$$\sum_{t=s+1}^{\min\{s+\theta,T\}} \sum_{p\in\{F,L,U\}} z_{mp}^t = d_m, \quad \forall m \in M_{\cdot}^s, 1 \le s < T \tag{3b}$$

Long-haul direct shipping capacity In each period, the total FTL and LTL delivery quantity to a seller does not exceed the total capacity of the dispatched FTL trucks and LTL units.

$$\sum_{s=\max\{1,t-\theta\}}^{t-1} \sum_{m\in M^s_{\cdot k}} z^t_{mp} \le \mathcal{K}_p r^t_{kp}, \quad \forall k \in D, p \in \{F,L\}, 1 < t \le T \tag{4}$$

Inventory conservation Growers (5a)-(5b): The ending inventory of a commodity equals the demand in its ready period or the beginning inventory in other periods, minus the corresponding short-haul pickup volume. Consolidation center (5c): The ending inventory of a commodity equals the beginning inventory plus the short-haul pickup volume and minus the long-haul delivery volume in each period. Boundary condition (5d): The central inventory of a commodity equals zero prior to or in its ready period.

$$I^t_{im} + q^t_m + v^t_m = d_m, \quad \forall i \in G, m \in M^s_{i\cdot}, 1 \le t = s < T \tag{5a}$$

$$I^t_{im} - I^{t-1}_{im} + q^t_m + v^t_m = 0, \quad \forall i \in G, m \in M^s_{i\cdot}, 1 \le s < t < \min\{s+\theta, T\} \tag{5b}$$

$$I^t_{0m} - I^{t-1}_{0m} - q^{t-1}_m - v^{t-1}_m + \sum_{p\in\{F,L,U\}} z^t_{mp} = 0, \quad \forall m \in M^s_{\cdot}, 1 \le s < t \le \min\{s+\theta, T\} \tag{5c}$$

$$I^t_{0m} = 0, \quad \forall m \in M^s_{\cdot}, 1 \le t < s+1 \le T \tag{5d}$$

Variable domains

$$x \in \{0,1\}, y \in \{0,1\}, u \in \mathbb{Z}_+, r \in \mathbb{Z}_+, q \ge 0, v \ge 0, z \ge 0, f \ge 0, I \ge 0 \tag{6}$$

We will refer to (1)-(6) as the *full MIP* model. The problem incorporates the conventional IRP as a special case (when $c_{kF} = c_{kL} = c_{kU} = 0$ for $k \in D$) and is therefore $\mathcal{NP}$-hard. In preliminary tests, we observed that instances with 15 periods, 10 growers and 5 sellers contain over 1,500 binary variables and 100 general integer variables, and usually took 5 hours to solve with CPLEX. In addition, the multi-commodity coefficient array is large and high-dimensional, causing memory shortage issues on a CONDOR system when the number of growers exceeds 15. In summary, the model is already difficult if not intractable for small to moderate instance sizes.

3 Solution Framework

Our approach consists of a decomposition procedure and a local search scheme. For illustrative purposes, we interpret the consolidation center as a common customer of the growers and a common supplier of the sellers. Accordingly, we define *central demand* as the quantity of each commodity

required to arrive at the consolidation center by the end of a period, and let *central supply* be the quantity of each commodity available at the consolidation center at the beginning of a period. These quantities are in fact variable in the full MIP model, but we decompose the model by fixing them in various ways.

3.1 Decomposition

The basic idea in our decomposition is to reduce the computational burden by solving a series of subproblems derived from the full MIP by fixing some decision variables' values. As Table 1 shows, we use three subproblems in different phases of our solution framework. The *direct shipping* (DS) subproblem assumes known central supply by fixing short-haul decisions, and determines the long-haul delivery quantities as well as FTL numbers and LTL units in each period to minimize long-haul transportation and central inventory costs. The IRP subproblem assumes known central demand by fixing long-haul decisions, and determines the short-haul pickup quantities, commodity flows and local routes in each period to minimize short-haul transportation, grower inventory and central inventory costs. The *restricted full MIP* subproblem assumes fixed FTL numbers in each period, and determines all other decisions to minimize the total distribution cost. The subproblems

Table 1: Summary of the three subproblems.

Subproblems	DS	IRP	Restricted full MIP
Fixed information	q	z	$r_{\cdot F}$
Decision variables	$z, r, I_{0\cdot}$	x, y, q, v, u, f, I	$x, y, q, v, u, f, z, I, r_{\cdot L}$
Objective	Long-haul costs	Short-haul costs	Total distribution cost
Constraints	(3b)-(4), (5c)-(6)	(2a)-(3a), (5a)-(5b), (6)	(2a)-(6)

are interrelated in that the output of each problem naturally defines an input neighborhood for the others. Depending on how we combine them, we obtain various decomposition procedures to solve the full MIP. To motivate our work, we first describe two straightforward and mutually complementary approaches.

The first approach is the standard *DS-guided decomposition*, which solves the full MIP in a *backward* fashion. It starts by solving the DS subproblem assuming that the central supply is available at the consolidation center as early as possible, i.e. each commodity's demand is available at the center one period after it is ready at the grower. The long-haul shipping quantities given by this DS subproblem's solution then specify central demand for the short-haul decision. Subsequently, the IRP subproblem is solved with this central demand, giving short-haul shipping quantities and implying central inventory levels, and together with the long-haul shipping quantities from the DS subproblem they constitute a feasible full MIP solution. This approach is intuitive since it prioritizes long-haul shipping and consolidation decisions, the dominant cost component.

The second approach is the alternative *IRP-guided decomposition*, which solves the full MIP in a *forward* fashion. It starts by solving the IRP subproblem assuming that central demand is due as late as possible, i.e. when the product would expire should it stay in the system any longer. The short-haul shipping quantities given by this subproblem's solution then specify the central supply for the long-haul decisions. Subsequently, the DS subproblem is solved with the given central supply, and these decisions together constitute a feasible full MIP solution. This approach is somewhat counterintuitive, as it prioritizes short-haul shipping decisions, but provides another perspective on how to manage system-wide costs.

The following propositions identify scenarios when the above approaches yield asymptotically optimal solutions. For the first result, consider a slight variation of the DS subproblem in which we allow central holding costs to depend on the commodity's grower, so that for any commodity $m = M^s_{ik}$ the central inventory variable I^t_{0m} incurs cost $h_{0,i}$.

Proposition 1. *The DS-guided decomposition yields asymptotically optimal full MIP solutions as the number of sellers increases, $|D| \to \infty$, if the DS subproblem is solved with the grower-specific central holding cost rates $h_{0,i} = \min\{h_0, h_i\}$, $\forall i \in G$.*

Proof. See Appendix A.1. □

Proposition 2. *The IRP-guided decomposition yields asymptotically optimal full MIP solutions as the number of growers increases, $|G| \to \infty$, if the IRP subproblem is solved without central holding cost, i.e. using $h'_0 = 0$.*

Proof. See Appendix A.2. □

Although the DS-guided and the IRP-guided decompositions can be justified in some situations, both resulted in 20% to 30% average optimality gaps for realistic problem instances in preliminary experiments. Example 1 below illustrates a situation where the DS-guided decomposition is ineffective, and there are similar examples where the IRP-guided decomposition yields a poor solution. Such instances motivate a more judicious trade-off between the short-haul and the long-haul decisions to coordinate the two echelons. Therefore, we propose a *partially DS-guided decomposition*, where the FTL numbers instead of all the long-haul shipping quantities are fixed as the DS subproblem is solved under a specific central supply, and then the induced restricted full MIP is solved to obtain a globally feasible solution. Compared with the DS-guided approach, our decomposition still prioritizes long-haul FTL costs (a dominant cost component), but is more flexible since the restricted full MIP relaxes the IRP subproblem by simultaneously deciding the long-haul LTL, courier and short-haul shipping quantities. Compared with the IRP-guided approach, the restricted full MIP gives rise to consolidation opportunities by allowing for limited changes in long-haul shipping quantities. Of course, the restricted full MIP involves more decision variables and constraints, so it appears more computationally demanding than the IRP subproblem. However, we observe that

the overall problem difficulty mainly stems from the routing decisions, and our approach achieves a reasonable balance between solution quality and computational runtime.

Example 1. Consider a network with two growers and one retailer, where the planning horizon length is $T = 3$ days, the product lifetime is $\theta = 2$ days, and the local vehicles as well as the long-haul FTL trucks have identical capacities, i.e. $Q = \mathcal{K}_F$. Suppose the holding cost rates are such that $h_1 = h_2 > h_0 > 0$. Let two orders be ready at the beginning of the planning horizon, $0.5Q$ for one grower and $0.5Q + \epsilon$ ($\epsilon > 0$) for the other. The total shipping volume exceeds the FTL capacity by a small volume of ϵ units, which will be shipped via LTL or courier services in the long-haul echelon. Since $h_0 > 0$ and the long-haul transportation cost depends on the volume rather than the shipping period, the optimal DS solution ships all $\mathcal{K}_F + \epsilon$ units on the second day to minimize central inventory costs. In the consequent IRP subproblem, this results in a violation of the local vehicle capacity on the first day and thus induces a penalty cost of B. On the other hand, the global optimum would postpone ϵ units to the third day as a compromise between the two echelons. The partially DS-guided decomposition attains this optimum since the final LTL and courier schedules are subject to the simultaneous short-haul decisions. The excess total cost of the DS-guided decomposition is then $B - h_2\epsilon$, which can be very high if the grower holding cost rate is small.

The DS subproblem can be solved with CPLEX for moderate instances; the algorithms in [29, 30] can handle large instances (e.g. $T = 300$) assuming just-in-time central supply ($h_i = 0, c_{i,j} = 0, \forall i, j \in G$ and so $I^{\cdot}_{0\cdot} = 0$, which yields the lowest possible long-haul transportation cost). The IRP subproblem and the restricted full MIP subproblem can be solved with CPLEX for small instances; existing heuristics for IRP variants with time windows may also be employed to obtain fast solutions for larger instances.

3.2 Local Search

Our motivation for developing a local search scheme is to escape poor local optima encountered in a single iteration of the partially DS-guided decomposition. Despite the additional flexibility gained over the DS-guided approach by substituting the IRP subproblem with the restricted full MIP, there is no improvement guarantee on the resulting solution, even if the subproblems are solved to optimality, which is unrealistic as the instance size increases. We next propose an optimization-based local search scheme that takes advantage of the inherent incompatibility between the DS and the IRP subproblems. We introduce some terminology before elaboration.

Definition 1 (*Mismatched demand*, MMD)**.** Assume both IRP and DS subproblems are solved simultaneously under given central demand and central supply assignments, respectively. A portion of the demand for a commodity (i, k, s) is said to be mismatched if the short-haul pickup time ι and the long-haul delivery time τ are incompatible, $s < \tau \leq \iota < \min\{s + \theta, T\}$; that is, the short-haul

solution collects this demand from the grower too late for its corresponding long-haul shipment. We denote this mismatched subcommodity by (i, k, s, ι, τ).

Definition 2 (*Service time windows*)**.** For a commodity (i, k, s), we may restrict the time in which it can be picked up at the grower to the time interval $[\underline{\iota}, \tilde{\iota}]$, which we call its *pickup time window*, where $s \le \underline{\iota} \le \tilde{\iota} < \min\{s + \theta, T\}$. Similarly, we may restrict the time in which the commodity can be shipped from the center to the time interval $[\underline{\tau}, \tilde{\tau}]$, which we call a *delivery time window*, where $s < \underline{\tau} \le \tilde{\tau} \le \min\{s + \theta, T\}$.

From the perspective of transportation costs, later central demand benefits short-haul routing whereas earlier central supply facilitates long-haul consolidation. Hence MMD can arise when the IRP subproblem and the DS subproblem are solved separately under the respective central demand and supply assumptions. In particular, the IRP solution with pickup windows $[s, \min\{s + \theta, T\})$ and the DS solution with delivery windows $(s, \min\{s + \theta, T\}]$ tend to be globally infeasible if we piece them together into a full MIP solution. Narrowing service time windows corresponding to MMD can eliminate this global infeasibility; for example, if (i, k, s, ι, τ) is a mismatched subcommodity with $\tau \le \iota$, we can revise the pickup window to be $[s, \tau)$ or the delivery window to be $(\iota, \min\{s + \theta, T\}]$, and the MMD will vanish when the subproblems are solved under the adjusted service time windows. To hopefully balance the objectives of both subproblems, we may use a pickup window $[s, \zeta)$ and a delivery window $[\zeta, \min\{s+\theta, T\}]$ where $\iota < \zeta \le \tau$, or adjust the pickup windows for a portion of the MMD and the delivery windows for the remaining units. The following example illustrates the basic idea.

Example 2. Consider the time-space network in Figure 3. There are three planning periods, two growers and two retailers, i.e. $T = 3$ days, $G = \{i, j\}, D = \{k, \ell\}$. The product has a lifetime $\theta = 2$ days, and the demands are $d^1_{ik} = 5$, $d^1_{i\ell} = 5$, $d^2_{i\ell} = 5$, $d^1_{jk} = 10$, $d^2_{j\ell} = 5$. Suppose local mileage costs are symmetric, holding costs are such that $h_i > h_0 > h_j$, $0 < h_0 < \min\{c_{kL}, c_{\ell L}\}$, and vehicle capacities satisfy $Q = \mathcal{K}_F = 15$. For illustration purposes, the consolidation center is split into two copies, representing the common IRP customer and the common DS supplier, respectively. Assume the central supply is ready for long-haul delivery the day after a demand is ready for pickup at the grower, whereas the central demand is not due until the expiration time or end of the horizon. This encourages both single-echelon subproblems to best utilize transportation capacity.

Suppose we solve the subproblems with the given holding cost rates: The total demand is $\sum_{s=1}^{2}(d^s_{ik}+d^s_{i\ell}+d^s_{jk}+d^s_{j\ell}) = 30 = 2Q$, and the portion ready on the first day is $d^1_{ik}+d^1_{i\ell}+d^1_{jk}+d^1_{j\ell} = 20 = Q + 5$. To avoid expensive direct shipping alternatives, the IRP subproblem tries to fully utilize the local routing vehicle capacity, which means a volume of 5 units will be held in inventory on day 1. Since $h_i > h_0 > h_j$, grower i's demands are prioritized whereas half of the demand for commodity $(j, k, 1)$ is delayed until day 2 for local pickup. On the other hand, the total demands for retailer k and retailer ℓ both equal $\mathcal{K}_F$, and are expected to be ready for long-haul delivery on day 2 and day 3, respectively; hence the optimal DS subproblem solution sends out an FTL on each day.

This solution is outlined in Figure 3(a), where the bold numbers represent the quantities associated with grower j and the others for grower i. The central flow imbalance from day 2 to day 3 indicates that subcommodity $(j, k, 1, 2, 2)$ induces an MMD of 5 units. Therefore, the corresponding full MIP solution would be infeasible if we piece the subproblem solutions together. To remove the MMD, we may keep the IRP solution and narrow the delivery window for commodity $(j, k, 1)$ to day 3 in the DS subproblem, or keep the DS solution and narrow the pickup window to day 1 in the IRP subproblem. As a compromise, we may also take the approach illustrated by Figure 3(b), where the pickup window for 2 units of the mismatched subcommodity $(j, k, 1, 2, 2)$ is narrowed to day 1, and the delivery window for the remaining 3 units is narrowed to day 3, respectively.

MMD essentially provides a guide to attain more compatible central demand and supply assignments for the subproblems. Since the modification of service time windows impacts routing, consolidation and inventory decisions, we propose using a *demand reasignment problem* to determine the time window modifications. The input of this problem includes both relevant full MIP parameters and extra data from the IRP and DS subproblem solutions. Specifically, we calculate mismatched subproblem demands, residual short-haul and long-haul transportation capacities as well as the remaining time allowed for each local route. We also estimate the routing cost and duration changes for each pair of grower and route. If grower i is visited by a regular short-haul vehicle in period t (*route* t), we approximate the routing cost savings of removing i from the route with the amount obtained by joining i's predecessor and successor when it is removed from route t. Meanwhile, if grower i is not visited by route t, we approximate its insertion cost with the cheapest insertion cost of inserting i into the route. We use the same approximations to estimate duration changes.

Additional input

$\hat{m} = \mathcal{M}_{ik}^{s\iota\tau}$: a mismatched subcommodity tuple, where a portion of demand d_m is picked up in period ι for the IRP subproblem and delivered in period τ for the DS subproblem, $m \in \{M_{ik}^s\}, i \in G, k \in D, s < \tau \leq \iota < \min\{s + \theta, T\}$. Define $\mathcal{M}$ with the possible usage of $\cdot$ analogously to M in §2.

$d_{\hat{m}}$: demand for subcommodity $\hat{m}$.

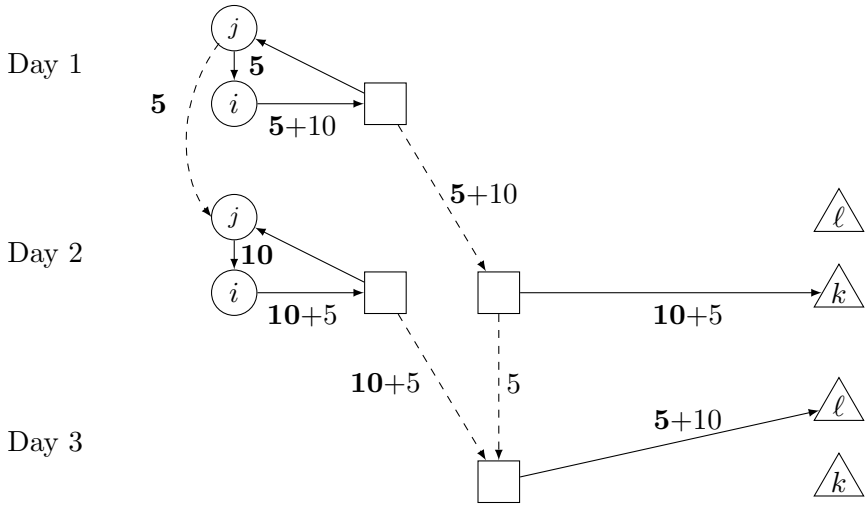
σ_i^t: indicator, equals 1 if grower i is visited by route t, 0 otherwise.

η_{i+}^t: insertion cost for grower-route pair (i, t), equals 0 if $\sigma_i^t = 1$.

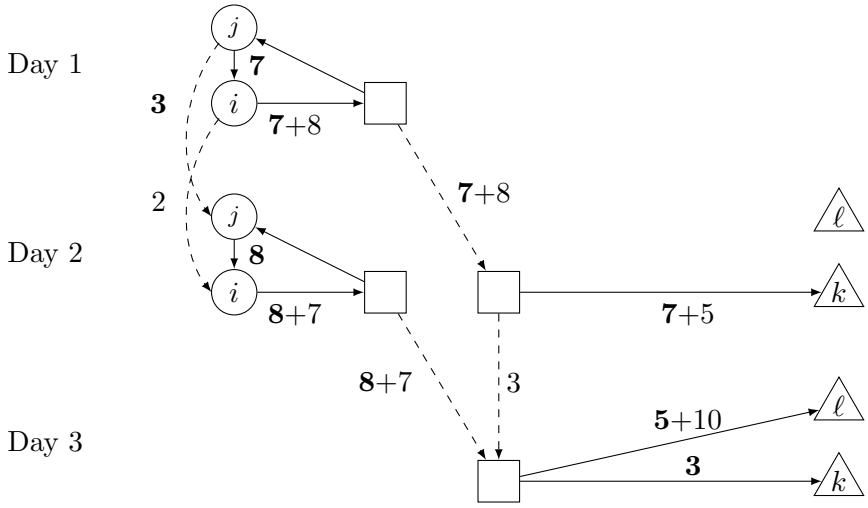
l_{i+}^t: duration increase for route t after inserting grower i, equals 0 if $\sigma_i^t = 1$.

η_{i-}^t: cost savings from removing i from t, equals 0 if $\sigma_i^t = 0$.

l_{i-}^t: duration reduction for route t after removing grower i, equals 0 if $\sigma_i^t = 0$.



(a) Full MIP infeasibility due to MMD



(b) A reassignment strategy by narrowing pickup and delivery windows

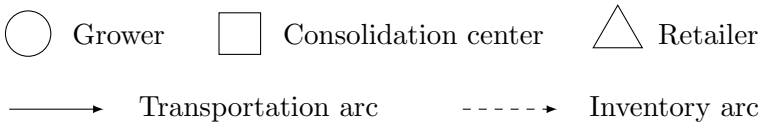


Figure 3: MMD and reassignment strategies

$\mathcal{L}^t$: remaining time allowed for route t, equals 1 minus the duration of route t if it occurs, 1 otherwise.

$\mathcal{Q}^t$: residual capacity for route t, equals Q minus the total short-haul pickup volume for route t if it occurs, Q otherwise.

$\mathcal{V}_k^t$: total FTL and LTL volume sent to seller k in period t in the DS subproblem solution.

Υ_{kp}^t: number of FTL trucks or LTL units sent to seller k in period t in the DS subproblem solution, $p \in \{F, L\}$.

The decision variables of the demand reassignment problem include insertion or removal for each grower-route pair (i, t), short-haul pickup and long-haul delivery quantities for each mismatched subcommodity, extra or saved FTL and LTL numbers, as well as courier volume and inventory level changes in each period.

Decision variables

$$\pi_i^t = \begin{cases} 1, & \text{if grower } i \text{ is inserted into route } t \\ 0, & \text{otherwise} \end{cases}, i \in G, 1 \leq t < T.$$

$$\rho_i^t = \begin{cases} 1, & \text{if grower } i \text{ is removed from route } t \\ 0, & \text{otherwise} \end{cases}, i \in G, 1 \leq t < T.$$

$$\nu^t = \begin{cases} 1, & \text{if the new route } t \text{ exceeds capacity or duration limit} \\ 0, & \text{otherwise} \end{cases}, 1 \leq t < T.$$

$\omega_{\hat{m}}^t \in \mathbb{R}_+$: reassigned short-haul pickup volume of subcommodity $\hat{m}$ to route t, $\hat{m} \in \mathcal{M}_{\cdot}^{s\cdot}, 1 \leq s \leq t < \min\{s + \theta, T\}$.

$\delta_{\hat{m}}^t \in \mathbb{R}_+$: reassigned long-haul delivery volume of subcommodity $\hat{m}$ in period t, $\hat{m} \in \mathcal{M}_{\cdot}^{s\cdot}$, $1 \leq s < t \leq \min\{s + \theta, T\}$.

$r_{kp+}^t, r_{kp-}^t \in \mathbb{Z}_+$: extra and saved FTL or LTL numbers, respectively, dispatched to seller k in period t, $k \in D, p \in \{F, L\}$, $1 < t \leq T$.

$z_{k+}^t, z_{k-}^t \in \mathbb{R}_+$: extra and saved courier volume, respectively, shipped to seller k in period t, $k \in D$, $1 < t \leq T$.

$I_{i\hat{m}+}^t, I_{i\hat{m}-}^t \in \mathbb{R}_+$: inventory increase and reduction, respectively, of subcommodity $\hat{m}$ at grower i in period t, $\hat{m} \in \mathcal{M}_{i\cdot}^{s\cdot}$, $i \in G, 1 \leq s \leq t < \min\{s + \theta, T\}$.

$I_{0\hat{m}}^t \in \mathbb{R}_+$: central inventory of subcommodity $\hat{m}$ in period t, $\hat{m} \in \mathcal{M}_{\cdot}^{s\cdot}$, $1 \leq s < t \leq \min\{s + \theta, T\}$.

We model the demand reassignment problem as a MIP. To allow some flexibility in local direct shipping, we assume that if a new route exceeds the local vehicle capacity or duration limit, one alternative vehicle is sufficient to make the solution feasible.

Reassignment MIP

$$\min \sum_{t=1}^{T-1}\sum_{i\in G}(\eta_{i+}^t\pi_i^t - \eta_{i-}^t\rho_i^t) + \sum_{t=1}^{T-1} B\nu^t + \sum_{t=2}^{T}\sum_{k\in D}\sum_{p\in\{F,L\}} c_{kp}(r_{kp+}^t - r_{kp-}^t) + \alpha\sum_{t=2}^{T}\sum_{k\in D} c_{kU}(z_{k+}^t - z_{k-}^t) + \sum_{t=1}^{T-1}\sum_{i\in G}\sum_{s=\max\{1,t-\theta+1\}}^{t}\sum_{\hat{m}\in\mathcal{M}_{i\cdot}^{s\cdot}} h_i(I_{i\hat{m}+}^t - I_{i\hat{m}-}^t) + \sum_{t=2}^{T}\sum_{s=\max\{1,t-\theta\}}^{t-1}\sum_{\hat{m}\in\mathcal{M}_{\cdot}^{s\cdot}} h_0 I_{0\hat{m}}^t \quad (7a)$$

$$\text{s.t. } \omega_{\hat{m}}^t \le d_{\hat{m}}(\sigma_i^t + \pi_i^t - \rho_i^t), \quad \forall\ \hat{m}\in\mathcal{M}_{i\cdot}^{s\cdot},\ i\in G,\ 1\le s\le t<\min\{s+\theta,T\} \quad (7b)$$

$$\pi_i^t \le 1-\sigma_i^t, \quad \forall\ i\in G,\ 1\le t<T \quad (7c)$$

$$\rho_i^t \le \sigma_i^t, \quad \forall\ i\in G,\ 1\le t<T \quad (7d)$$

$$\sum_{s=\max\{1,t-\theta+1\}}^{t}\left(\sum_{\hat{m}\in\mathcal{M}_{\cdot}^{s\tau\cdot}:\tau\neq t}\omega_{\hat{m}}^t - \sum_{\hat{m}\in\mathcal{M}_{\cdot}^{st\cdot}}(d_{\hat{m}}-\omega_{\hat{m}}^t)\right) \le \mathcal{Q}^t + Q\nu^t, \quad \forall\ 1\le t<T \quad (7e)$$

$$\sum_i l_{i+}^t\pi_i^t - \sum_i l_{i-}^t\rho_i^t \le \mathcal{L}^t + \nu^t, \quad \forall\ 1\le t<T \quad (7f)$$

$$\sum_{t=s}^{\min\{s+\theta,T\}-1}\omega_{\hat{m}}^t = \sum_{t=s+1}^{\min\{s+\theta,T\}}\delta_{\hat{m}}^t = d_{\hat{m}}, \quad \forall\ \hat{m}\in\mathcal{M}_{\cdot}^{s\cdot},\ 1\le s<T \quad (7g)$$

$$\sum_{t=s}^{\tau}\omega_{\hat{m}}^t \ge \sum_{t=s+1}^{\tau+1}\delta_{\hat{m}}^t, \quad \forall\ \hat{m}\in\mathcal{M}_{\cdot}^{s\cdot},\ 1\le s\le\tau<\min\{s+\theta,T\} \quad (7h)$$

$$\mathcal{V}_k^t + \sum_{s=\max\{1,t-\theta\}}^{t-1}\left(\sum_{\hat{m}\in\mathcal{M}_{\cdot k}^{s\cdot\tau}:\tau\neq t}\delta_{\hat{m}}^t - \sum_{\hat{m}\in\mathcal{M}_{\cdot k}^{s\cdot t}}(d_{\hat{m}}-\delta_{\hat{m}}^t)\right) \le \sum_p K_p(\Upsilon_{kp}^t + r_{kp+}^t - r_{kp-}^t)$$
$$+z_{k+}^t - z_{k-}^t, \quad \forall\ k\in D,\ 1<t\le T \quad (7i)$$

$$I_{i\hat{m}+}^t - I_{i\hat{m}-}^t = \sum_{\iota=t+1}^{\min\{s+\theta,T\}-1}\omega_{\hat{m}}^\iota, \quad \forall\ \hat{m}\in\mathcal{M}_{i\cdot}^{s\tau\cdot},\ i\in G,\ 1\le s\le\tau\le t<\min\{s+\theta,T\} \quad (7j)$$

$$I_{i\hat{m}-}^t - I_{i\hat{m}+}^t = \sum_{\iota=s}^{t}\omega_{\hat{m}}^\iota, \quad \forall\ \hat{m}\in\mathcal{M}_{i\cdot}^{s\tau\cdot},\ i\in G,\ 1\le s\le t<\tau<\min\{s+\theta,T\} \quad (7k)$$

$$I_{0\hat{m}}^t = \sum_{\tau=s}^{t-1}\omega_{\hat{m}}^\tau - \sum_{\tau=s+1}^{t}\delta_{\hat{m}}^\tau, \quad \forall\ \hat{m}\in\mathcal{M}_{\cdot}^{s\cdot},\ 1\le s<t\le\min\{s+\theta,T\} \quad (7l)$$

$$r_{kp-}^t \le \Upsilon_{kp}^t, \quad \forall\ k\in D,\ p\in\{F,L\},\ 1<t\le T \quad (7m)$$

$$z_{k-}^t \le \sum_{\hat{m}\in\mathcal{M}_{\cdot k}^{\cdot t}} d_{\hat{m}} - \mathcal{V}_k^t, \quad \forall\ k\in D,\ 1<t\le T \quad (7n)$$

$$\pi\in\{0,1\}, \rho\in\{0,1\}, \nu\in\{0,1\}, r\in\mathbb{Z}_+, \omega\ge 0, \delta\ge 0, z\ge 0, I\ge 0 \quad (7o)$$

The objective (7a) is to minimize the total *net* rerouting and reconsolidation cost. Note that we calculate net inventory cost changes at the growers, but only consider inventory costs after reassignment at the consolidation center. (7b)-(7d) ensure that the binary rerouting variables are correctly updated to form a new route, i.e. pickup can occur only if a grower is visited; insertion can occur only if the grower was not visited in the original IRP subproblem solution; removal can occur only if the grower was visited. (7e)-(7f) are short-haul transportation capacity and duration constraints, i.e. the net increase of reassigned pickup volume does not exceed the residual regular vehicle capacity plus alternative capacity in each period; similarly, the net increase of duration after reassignment does not exceed the residual time of the original regular route plus the length of a possible alternative direct shipping trip (i.e. one period). (7g)-(7h) are demand satisfaction constraints redefined for each MMD, i.e. the total short-haul pickup volume equals the total long-haul delivery volume after reassignment; at any point before the product spoils, the total short-haul pickup volume to date is no less than the total long-haul delivery volume by the next period. (7i) are aggregated direct shipping capacity constraints after canceling out the courier volume, i.e. the net increase of reassigned delivery volume to a seller does not exceed the residual long-haul transportation capacity plus the extra capacity in each period. (7j)-(7l) are inventory balance constraints, i.e. the grower's inventory of any MMD in period t increases by the total later reassigned short-haul pickup volume if it was shipped by period t in the original IRP solution; the grower's inventory of any MMD in period t decreases by the total short-haul pickup volume reassigned earlier than or in period t if it was shipped after that in the original IRP solution; the central inventory of any MMD in period t equals the total short-haul pickup volume that has arrived minus the total long-haul volume that has been delivered to the seller. (7m)-(7n) and (7o) are boundary conditions and the domain, respectively.

At first glance, the reassignment MIP (7) may appear complicated and similar in structure to the full MIP. However, it exhibits several features that enable efficient optimization: First, decisions for matched demands are fixed so the problem size is smaller than the full MIP; second, combinatorial rerouting costs are approximated linearly; third, complex subtour elimination constraints are circumvented with the introduction of simple binary variables. In our experiments, CPLEX solved it almost instantly in most cases.

3.3 An Iterative Framework

We have set up an optimization problem for local search in hope of eliminating MMD at the lowest cost. We note the following issues:

- We fix the subproblem decisions for matched demands before solving Model (7). If a grower is removed from a tour but only a fraction of the shipment was mismatched, then the matched portion should also be removed from the tour and the residual vehicle capacity should be larger. Model (7) does not capture this.

- The approximate rerouting costs for the reassignment problem may differ from the optimal values. For instance, the effect of rerouting multiple growers in a period is the corresponding TSP tour cost change, which generally is not the summation of cost changes incurred by each single grower. Similarly, the savings approximation of removing a single grower may underestimate the true amount.

- The input quantities that Model (7) inherits from the subproblem solutions may not reflect the full MIP solution induced by the reassigned quantities. For instance, the subproblems start with predefined central demand and supply assignments, but the short-haul pickup or long-haul delivery times are not revealed until the reassigned solution is determined; hence the modeled central inventory levels may differ from the final outcome.

For these reasons, we do not count on finding high-quality full MIP solutions by solving Model (7) once. Instead, we propose to obtain solutions by embedding the decomposition and the local search in an iterative framework (Figure 4). Throughout the procedure, we record the best full MIP solution found so far, and keep a count that determines when to terminate. We partition the given commodities M into two subsets, where $\mathcal{S}$ contains the matched sub-commodities and $\mathcal{M}$ the mismatched subset. Both subsets will be updated from iteration to iteration. To initiate the procedure, we set $\mathcal{S} = M,\ \mathcal{M} = \varnothing$, i.e. we assume all given demands are matched.

We begin with the most flexible transportation for both echelons, which is realized by fixing the pickup windows to $[s, \min\{s+\theta, T\})$ and the delivery windows to $(s, \min\{s+\theta, T\}]$ for each $m \in M^s_{\cdot}$. This gives the latest possible central demand and the earliest possible central supply. We first solve the DS subproblem and obtain the long-haul shipping quantities. Subsequently, we fix the FTL numbers and solve the corresponding restricted full MIP. If a predefined maximum allowable number of iterations N is not yet reached, we then solve the IRP subproblem and calculate all mismatched demands in comparison to the DS solution. If there is any MMD, i.e. $\mathcal{M} \neq \varnothing$ or $\sum_{\hat{m}\in\mathcal{M}} d_{\hat{m}} > 0$, we update the subset $\mathcal{S} = M \setminus \mathcal{M}$, and go to the next iteration. Each new iteration starts by solving Model (7), after which we adjust the mismatched sub-commodity service time windows based on the reassigned quantities. For example, if the demand for $\mathcal{M}^{s\iota\tau}_{ik}$ where $\tau \leq \iota$ is reassigned to period ι' for short-haul shipping and period τ' for long-haul shipping ($\iota' < \tau'$), then we modify the pickup and delivery windows to be $[s, \tau')$ and $(\iota', \min\{s+\theta, T\}]$, respectively. For the matched sub-commodities $\mathcal{S}$, we use the initial service time windows to encourage higher utilization of transportation capacities. Mathematically, this is equivalent to adding constraints $\sum_{t=s}^{\tau'} q^t_m \geq \sum_{t=s}^{\tau'-1} \tilde{\omega}^t_{\hat{m}}$ in the IRP model and $\sum_{t=\iota+1}^{\min\{s+\theta,T\}} \sum_{p\in\{F,L,U\}} z^t_{mp} \geq \sum_{t=\iota+1}^{\min\{t+\theta,T\}} \tilde{\delta}^t_{\hat{m}}$ in the DS model for all $\hat{m} \in \mathcal{M}$, where $\tilde{\omega}$ and $\tilde{\delta}$ are the shipping quantities reassigned by Model (7). The adjusted service time windows induce new central demand and supply, and the iterative process repeats itself until the maximum number of iterations is met or MMD no longer exists, where we output the best full MIP solution and exit.

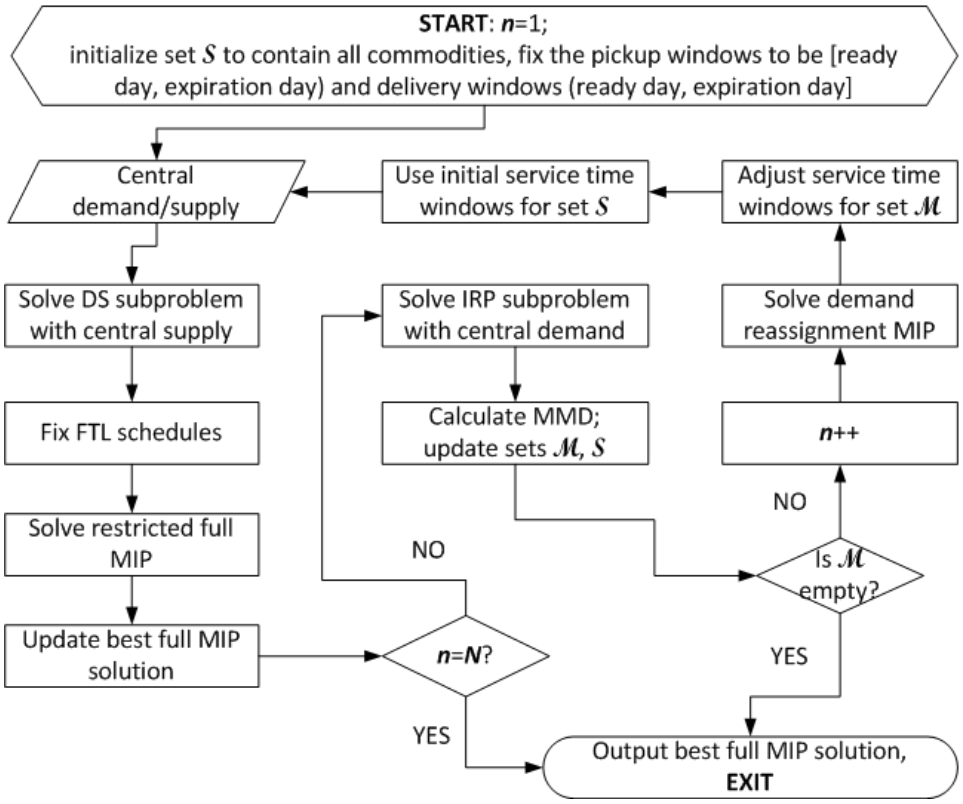


Figure 4: Flowchart of the iterative framework

4 Computational Study

4.1 Experimental Design

We designed experiments based on California cut flower sales data from the year 2010. For each grower-seller pair, we used an empirical demand distribution based on the method described in [29]. We randomly generated 10 instances from the distributions for each of the following cases.

Baseline: $T = 15, |G| = 10, |D| = 5$.

TI: $T = 30, |G| = 10, |D| = 5$.

TII: $T = 45, |G| = 10, |D| = 5$.

GI: $T = 15, |G| = 15, |D| = 5$.

DI: $T = 15, |G| = 10, |D| = 10$.

The baseline instances serve as a standard test bed where we can obtain optimal solutions with a commercial solver. Instances TI and TII examine the performance of solution approaches over longer planning horizons. Instances GI concern the impact of more growers, DI instances that of more sellers. The generated demands were scaled to ensure that the total volume is identical

across profiles, and there are enough local routing vehicles to carry it. For each instance, we further considered three scenarios: $h_0 = 0, 2, 4$, representing low, moderate, and high central holding cost rates, respectively. The common parameter settings are $\theta = 3$ and $h_i = 1, \forall i \in G$.

We used CPLEX 12.6 Concert technology with Visual Studio C++ 2010 for all the MIP models, running on a Linux server with 70 Gs of memory and 16 cores. The CPLEX MIP emphasis parameter is set to *hidden feasibility*, to prioritize the search of high quality feasible solutions over proving optimality of the best incumbent [26]. We ran the proposed iterative partially DS-guided (IPDSG) heuristic, and compare it with both CPLEX and the DS-guided (DSG) benchmarks. For all the instances except GI, we report the IPDSG upper bounds, the CPLEX upper and lower bounds, as well as the DSG upper bounds within five hours of CPU time. The GI instances are more challenging because of the exponentially increasing computational time to solve the IRP subproblems; hence we report the CPLEX upper bounds within 10 hours as an alternative benchmark to the lower bounds, which in many cases remained quite low even when the instances ran on CPLEX for 20 hours.

4.2 Results

We list the baseline results in Table 2, where the numbers in each row are averaged over the 10 instances tested. CPLEX attains the optimal solutions under those demand profiles, utilizing 90% of the 5-hour CPU time. In the IPDSG experiments, we allocated the CPU time as follows: 200 seconds for each DS model, 1500 seconds for each IRP model, 1800 seconds for each restricted full MIP, and 100 seconds for each reassignment MIP. In the DSG experiments, we solved both DS and the induced IRP subproblems to optimality, which typically finished in 200 and 1500 seconds, respectively. Under all the holding cost scenarios, the IPDSG heuristic finds solutions whose objective values are within 2% and 5% of the CPLEX upper and lower bounds, respectively, using 2 to 3 iterations and 65% to 80% CPU time on average. The first iteration solutions achieve less than 10% optimality gaps, and are up to 20% better than the DSG solutions. These results indicate that our decomposition is effective in balancing solution quality and computational time, and the local search scheme is promising in solution improvement by efficiently finding better neighborhoods.

Table 2: Baseline performance

h_0	IPDSG first iteration				IPDSG best solution					IPDSG
	UB	vs. CPLEX		vs. DSG	#iter.	UB	vs. CPLEX		vs. DSG	CPU
		UB	LB				UB	LB		utilization
0	45124	1.03	1.04	0.94	2.8	44756	1.02	1.03	0.94	0.71
2	55808	1.06	1.08	0.82	2.3	53831	1.02	1.05	0.79	0.78
4	54715	1.03	1.06	0.79	1.9	52830	1.00	1.03	0.77	0.65

We next summarize the results for longer planning horizons and more facilities. CPLEX fully utilizes the CPU time in these settings, and DSG also takes more than four hours to optimally solve the most difficult IRP models. On the other hand, if we allow flexibility in CPU time allocation by using different CPU time per iteration under different demand profiles, IPDSG finds good solutions with 3 to 5 iterations within five hours. As Table 3 shows, we can match the CPLEX upper bounds as the first iteration ends, and reduce the optimality gaps by 1% to 3% within the CPU time limit when $T = 30$. The advantage is more evident for $T = 45$, where CPLEX gives weaker lower bounds but IPDSG yields significantly better upper bounds. Compared with the baseline performance, IPDSG also exhibits a growing advantage over DSG as T increases.

Table 3: TI and TII performance

T	h_0	IPDSG first iteration				IPDSG best solution				
		UB	vs. CPLEX		vs. DSG	#iter.	UB	vs. CPLEX		vs. DSG
			UB	LB				UB	LB	
30	0	83362	1.00	1.03	0.90	2.6	82424	0.99	1.02	0.89
	2	105185	0.94	1.09	0.77	2.2	103189	0.92	1.07	0.76
	4	106351	0.99	1.10	0.76	2.6	103358	0.96	1.07	0.74
45	0	125822	0.99	1.04	0.91	1.8	124692	0.98	1.03	0.90
	2	192408	0.86	1.34	0.93	2.2	157083	0.70	1.09	0.76
	4	162140	0.79	1.12	0.76	2.4	156267	0.76	1.08	0.73

Table 4 demonstrates the potential of our heuristic for larger grower sets. Within an average of 3 iterations, IPDSG yields 3% to 15% better solutions than the CPLEX 5-hour upper bounds, and matches or improves upon CPLEX 10-hour upper bounds. The difference between IPDSG and DSG is sometimes smaller than that for the baseline instances, perhaps becaue DSG has more time to solve the IRP subproblem.

Table 4: GI performance

h_0	IPDSG first iteration				IPDSG best solution				
	UB	vs. CPLEX		vs. DSG	#iter.	UB	vs. CPLEX		vs. DSG
		5hr UB	10hr UB				5hr UB	10hr UB	
0	51790	1.10	1.13	1.03	2.8	45933	0.97	1.00	0.91
2	56869	0.85	0.96	0.85	2.7	54939	0.82	0.93	0.82
4	57427	0.88	0.95	0.84	2.7	55355	0.85	0.92	0.81

Table 5 demonstrates the potential of IPDSG for larger seller sets. The first iteration solutions match the CPLEX upper bounds, and the quality is further improved by 1% to 3% as the procedure ends.

Table 5: DI performance

h_0	IPDSG first iteration				IPDSG best solution				
	UB	vs. CPLEX		vs. DSG	#iter.	UB	vs. CPLEX		vs. DSG
		UB	LB				UB	LB	
0	71810	1.00	1.01	0.93	2.2	71634	1.00	1.00	0.93
2	87254	0.99	1.07	0.87	2.2	85233	0.97	1.05	0.85
4	88545	0.99	1.09	0.87	2.0	85590	0.96	1.05	0.84

In all the experiments, DSG perform significantly better when the central holding cost rate is lower. Also, the IPDSG first iteration solutions have relatively smaller optimality gaps when $h_0 = 0$ and/or there are more sellers. Intuitively, lower h_0 encourages both short-haul and long-haul echelons to keep inventory at the consolidation center; hence the initial DS and IRP subproblems are more "compatible". This is in line with Proposition 1, implying that our single-iteration decomposition and the DS-guided decomposition as well yield asymptotically optimal full MIP solutions as $|D| \to \infty$ if $h_0 \leq h_i$, $\forall i \in G$.

We intended to verify Proposition 2 by testing the IRP-guided decomposition, but were unable to carry out the desired experiments because CPLEX could not solve the IRP subproblem within the five-hour limit already for instances with 15 growers. Future work could perhaps employ relevant IRP heuristics within the DSG or IPDSG framework to further investigate this question.

5 Conclusions

In this paper, we study a two-echelon distribution problem integrating an IRP problem and a freight consolidation problem for perishable goods with a fixed lifetime. We propose an iterative solution framework that consists of a decomposition into subproblems and an optimization-based local search. Extensive experiments with empirical demand distributions demonstrate the potential of our solution approach. We also present some theoretical findings on the asymptotic behavior of related decompositions.

The computational results motivate us to develop fast and effective heuristics for the subproblems, in particular the IRP model and the restricted full MIP. We are also interested in designing diversification mechanisms to explore various types of neighborhoods, which seems necessary for larger problem instances where heuristics commonly get stuck in local optima [1, 3]. The optimization-based local search could also give rise to simple and flexible alternatives, e.g. by incorporating other cost components in the objective function (7a) or revising constraints (7g)-(7h), which would result in different reassignment strategies.

Acknowledgements

This work was partially funded by the National Science Foundation under award CMMI-1265616.

References

[1] N. Absi, C. Archetti, S. Dauzere-Peres, and D. Feillet. A two-phase iterative heuristic approach for the production routing problem. *Transportation Science*, 49(4):784–795, 2015.

[2] Y. Adulyasak, J. F. Cordeau, and R. Jans. Formulations and branch-and-cut algorithms for multivehicle production and inventory routing problems. *INFORMS Journal on Computing*, 26(1):103–120, 2014.

[3] Y. Adulyasak, J. F. Cordeau, and R. Jans. Optimization-based adaptive large neighborhood search for the production routing problem. *Transportation Science*, 48(1):20–45, 2014.

[4] Y. Adulyasak, J. F. Cordeau, and R. Jans. The production routing problem: A review of formulations and solution algorithms. *Computers & Operations Research*, 55(0):141–152, 2015.

[5] A. Akbalik and C. Rapine. Polynomial time algorithms for the constant capacitated single-item lot sizing problem with stepwise production cost. *Operations Research Letters*, 40:390–397, 2012.

[6] H. Andersson, A. H. M. Christiansen, G. Hasle, and A. Løkketangen. Industrial aspects and literature survey: Combined inventory management and routing. *Computers & Operations Research*, 37(9):1515–1536, 2010.

[7] C. Archetti, L. Bertazzi, G. Paletta, and M. G. Speranza. Analysis of the maximum level policy in a production-distribution system. *Computers & Operations Research*, 38:1731–1746, 2011.

[8] C. Archetti, L. Bertazzi, and M. G. Speranza. Polynomial cases of the economic lot sizing problem with cost discounts. *European Journal of Operations Research*, 237:519–527, 2014.

[9] V. A. Armentano, A. L. Shiguemoto, and A. Løkketangen. Tabu search with path relinking for an integrated production-distribution problem. *Computers & Operations Research*, 38(8):1199–1209, 2011.

[10] J. F. Bard and N. Nananukul. Heuristics for a multiperiod inventory routing problem with production decisions. *Computers & Industrial Engineering*, 57(3):713–723, 2009.

[11] J. F. Bard and N. Nananukul. The integrated production-inventory-distribution-routing problem. *Journal of Scheduling*, 12(3):257–280, 2009.

[12] J. F. Bard and N. Nananukul. A branch-and-price algorithm for an integrated production and inventory routing problem. *Computers & Operations Research*, 37(12):2202–2217, 2010.

[13] D. E. Blumenfeld, L. D. Burns, D. J. Diltz, and C. F. Daganzo. Analyzing trade-offs between transportation, inventory and production costs on freight networks. *Transportation Research Part B: Methodological*, 19(5):361–380, 1985.

[14] M. Boudia, M. A. O. Louly, and C. Prins. A reactive GRASP and path relinking for a combined production-distribution problem. *Computers & Operations Research*, 34(11):3402–3419, 2007.

[15] M. Boudia and C. Prins. A memetic algorithm with dynamic population management for an integrated production-distribution problem. *European Journal of Operational Research*, 195(3):703–715, 2009.

[16] S. Çetinkaya. Coordination of inventory and shipment consolidation decisions: A review of premises, models, and justification. In G. Joseph, E. Akçali, P. M. Pardalos, H. E. Romeijn, and Z. J. Shen, editors, *Applications of Supply Chain Management and E-Commerce Research*, volume 92 of *Applied Optimization*, pages 3–51. Springer, US, 2005.

[17] L. M. A. Chan, A. Muriel, Z. J. Shen, and D. Simchi-Levi. On the effectiveness of zero-inventory-ordering policies for the economic lot sizing model with a class of piecewise linear cost structures. *Operations Research*, 50(6):1058–1067, 2002.

[18] L. C. Coelho, J. F. Cordeau, and G. Laporte. The inventory-routing problem with transshipment. *Computers & Operations Research*, 39(11):2537–2548, 2012.

[19] L. C. Coelho, J. F. Cordeau, and G. Laporte. Thirty years of inventory routing. *Transportation Science*, 48(1):1–19, 2014.

[20] M. Florian, J. K. Lenstra, and A. H. G. Rinnooy Kan. Deterministic production planning: Algorithms and complexity. *Management Science*, 26(7):669–679, 1980.

[21] F. Fumero and C. Vercellis. Syncronized development of production, inventory, and distribution schedules. *Transportation Science*, 33:330–340, 1999.

[22] M. Goetschalckx. *Supply Chain Engineering*. Springer, London, 2011.

[23] R. W. Hall. Consolidation strategy : Inventory, vehicles and terminals. *Journal of Business Logistics*, 8(2):57–73, 1987.

[24] W. Hu. *Modeling and Solution of Some Multi-Period Supply Chain Optimization Problems*. PhD thesis, Georgia Institute of Technology, 2016.

[25] W. Hu, A. Toriello, and M. Dessouky. Decomposition-based approximation algorithms for the one-warehouse multi-retailer problem with concave batch order costs. Working paper, available at `http://www2.isye.gatech.edu/~atoriello3/owmr_apx.pdf`, 2016.

[26] IBM ILOG. CPLEX Optimization Studio CPLEX Parameters Reference, version 12.6. Available at `http://www.ibm.com/support/knowledgecenter/SSSA5P_12.6.3/ilog.odms.studio.help/pdf/paramcplex.pdf`.

[27] Y. Jin and A. Muriel. Single-warehouse multi-retailer inventory systems with full truckload shipments. *Naval Research Logistics*, 56(5):450–464, 2009.

[28] S. A. Lippman. Optimal inventory policy with multiple set-up costs. *Management Science*, 16(1):118–138, 1969.

[29] C. Nguyen, M. Dessouky, and A. Toriello. Consolidation strategies for the delivery of perishable products. *Transportation Research Part E: Logistics and Transportation Review*, 69:108–121, 2014.

[30] C. Nguyen, A. Toriello, M. Dessouky, and J. E. Moore. Evaluation of transportation practices in the california cut flower industry. *Interfaces*, 43(2):182–193, 2013.

[31] D. J. Papageorgiou, G. L. Nemhauser, J. Sokol, M.-S. Cheon, and A. B. Keha. MIRPLib – A library of maritime inventory routing problem instances: Survey, core model, and benchmark results. *European Journal of Operational Research*, 235(2):350 – 366, 2014.

[32] M. Ruokokoski, O. Solyalı, R. Jans, and H. Süral. Efficient formulations and a branch-and-cut algorithm for a production-routing problem. GERAD Technical Report G-2010-66, HEC Montréal, Canada, 2010.

[33] O. Solyalı and H. Süral. A relaxation based solution approach for the inventory control and vehicle routing problem in vendor managed systems. In S. K. Neogy, A. K. Das, and R. B. Bapat, editors, *Modeling, Computation and Optimization*, volume 6 of *Statistical Science and Interdisciplinary Research*. World Scientific, Singapore, 171-189, 2009.

[34] G. Stauffer, G. Massonnet, C. Rapine, and J.-P. Gayon. A simple and fast 2-approximation algorithm for the one-warehouse multi-retailer problem. In *Proceedings of the Twenty-Second Annual ACM-SIAM Symposium on Discrete Algorithms*, pages 67–79, 2011.

[35] U.S. Department of Agriculture and U.S. Department of Transportation. Study of rural transportation issues. Technical report, Washington DC, April 2010.

Appendix A Remaining Proofs

A.1 Proof of Proposition 1

We bridge the specified DS-guided solution and the optimal full MIP solution with an auxiliary solution, which is constructed as follows:

Step 1. Solve the DS subproblem with $h_{0,i}$ to obtain the long-haul shipping decisions.

Step 2. For each grower $i \in G$, assign the associated DS shipping quantities to the commodity ready periods for local pickup if its holding cost rate exceeds the warehouse's rate; otherwise, assign those quantities to the DS shipping periods. That is, add constraints $q^s_m + v^s_m = d_m$ in the IRP subproblem if $h_i > h_0$, else add $q^{t-1}_m + v^{t-1}_m = \sum_{p\in\{F,L,U\}} \tilde{z}^t_{mp}$, $\forall m \in M^{\cdot}_{i\cdot}$, $t = (s+1)\dots\min\{s+\theta, T\}$, where $\tilde{z}^t_{mp}$ are the values determined in Step 1. This gives the system-wide inventory decisions.

Step 3. Solve the resulting IRP subproblem to obtain the short-haul shipping decisions.

The above procedure yields a feasible full MIP solution. Let the overall costs for this solution, the specified DS-guided solution and the optimal full MIP solution be Z^a, Z^{DSG} and Z^*, respectively. Using the same superscripts, let $V^{\cdot}$, $H^{\cdot}$ and $C^{\cdot}$ be the respective total local vehicle routing and alternative direct shipping cost, total system-wide holding cost, and total long-haul direct shipping cost incurred by the corresponding full MIP solutions, i.e. $Z^{\cdot} = V^{\cdot} + H^{\cdot} + C^{\cdot}$. Assume the demand for each commodity, d_m, is i.i.d. in each period with mean $\bar{d}_{ik} > 0$, $\forall m \in M^{\cdot}_{ik}$. We will show that $\lim_{|D|\to\infty} \frac{Z^{DSG}-Z^*}{Z^*} = \lim_{|D|\to\infty} \frac{Z^a-Z^*}{Z^*} = 0$ with probability 1.

First, $Z^{DSG} \le Z^a$: We respect the DS-guided long-haul shipping plan when constructing the auxiliary solution; hence $C^{DSG} = C^a$. The auxiliary solution implies a feasible solution to the IRP subproblem induced by the DS-guided decomposition under the same long-haul shipping plan; hence $H^{DSG} + V^{DSG} \le H^a + V^a$.

Second, $H^a + C^a \le H^* + C^*$: Step 2 ensures that the total system-wide holding cost and long-haul shipping cost incurred by the auxiliary solution equals the optimal objective value of the DS subproblem under the revised holding cost rates $h_{0,i}$, which is in fact a relaxation of the full MIP where the local transportation decisions x, y, u are removed. The relaxation can be validated by converting any given feasible full MIP solution into a feasible solution with no higher objective value for the specified DS subproblem analogously to Step 2; see [25] Lemma 8.2 and [34] Lemma 5.1 for more details on relevant two-echelon inventory models.

Third, $V^a - V^* \le T|G|(A+B)$, where A is the optimal TSP tour cost when all growers are visited, and B the per alternative local direct shipment cost parameter (see §2). This is true since we allow one routing vehicle in each period, and the alternative direct shipments contain at most one partially-filled truck for each grower in each period.

Fourth, $C^* \ge \sum_{k\in D} \sum_{m\in M^{\cdot}_{\cdot k}} c_{kF} d_m / \mathcal{K}_F$, where the right-hand-side is the long-haul direct shipping cost with the highest possible truck capacity utilization. This value is obtained by assuming

the lowest possible unit rates of $c_{kF}/\mathcal{K}_F$ and the longest possible product lifetime $\theta = T$; hence it is a lower bound for that incurred by any feasible full MIP solution.

Therefore, $\frac{Z^{DSG}-Z^*}{Z^*} \leq \frac{Z^a-Z^*}{Z^*} \leq \frac{V^a-V^*}{C^*} \leq \frac{T|G|(A+B)}{\sum_{k\in D}\sum_{m\in M^-_k} c_{kF}d_m/\mathcal{K}_F} \to 0$ as $|D| \to \infty$. In other words, both the auxiliary solution and the specified DS-guided solution are asymptotically optimal when there are arbitrarily many sellers.

The proposition can be extended to the case of multiple local routing vehicles by slightly modifying the comparison of V^a and V^* assuming $B \leq A$.

A.2 Proof of Proposition 2

We show the result with the following lemmas. Lemma 3 ensures the existence of a special cost function that bounds each seller's long-haul transportation cost function from above, whereas Lemma 4 gives structural properties of an LSP under this upper bound cost function. Both lemmas are proved in [24].

Lemma 3. *Let $C_k(q)$ be the long-haul direct shipping transportation cost when q units are sent to retailer k. There exist functions $\tilde{C}_k(\cdot)$ such that*

$$C_k(q) \leq \tilde{C}_k(q) = \lfloor \frac{q}{\mathcal{K}_F} \rfloor \cdot c_{kF} + \tilde{c}_k(q \bmod \mathcal{K}_F), \quad \forall q \in \mathbb{R}_+, \tag{8}$$

where $\tilde{c}_k(\cdot)$ is a PWL concave function on $[0, \mathcal{K}_F]$ and $\tilde{c}_k(\mathcal{K}_F) = c_{kF}$, $\forall k \in D$.

Lemma 4. *The perishable lot sizing problem with concave batch transportation costs, where the item lifetime is fixed and the cost per batch is an identical concave function of the shipping volume within the batch capacity, has an optimal solution such that in each period,*

i) there is at most one partially filled batch;

ii) the end inventory is lower than the batch capacity.

Let Z^* and Z^{IRPG} be the optimal full MIP objective value and that of the IRP-guided solution, respectively. When the IRP subproblem is solved with $h'_0 = 0$, it minimizes the short-haul transportation costs and the grower holding costs under any input cost parameters; hence its objective value does not exceed the corresponding cost components of the optimal full MIP solution. To analyze the gap between Z^{IRPG} and Z^*, it then suffices to check the central inventory costs and the long-haul transportation costs. Let W^*, W^{IRPG}, and $\tilde{W}^{\text{IRPG}}$ be the respective total values of these cost components incurred by the optimal full MIP solution, the IRP-guided solution, and the optimal solution of the DS subproblem which uses the central supply implied by the IRP solution and some $\tilde{C}_k(\cdot)$ that satisfy (8). Since the long-haul decision for each seller is independent from the others' under the given central supply, the DS subproblem is equivalent to $|D|$ lot sizing problems. Applying a similar argument as Theorem 7.1 in [27], we have

$$\frac{Z^{\text{IRPG}} - Z^*}{Z^*} \leq \frac{W^{\text{IRPG}} - W^*}{W^*} \leq \frac{\tilde{W}^{\text{IRPG}} - W^*}{W^*} \leq \frac{\sum_{k\in D}(h_0\mathcal{K}_F + c_{kF})T}{\sum_{i\in G}\sum_{k\in D}\sum_{m\in M^-_{ik}} c_{kF}d_m/\mathcal{K}_F}, \tag{9}$$

where $W^{\text{IRPG}} \leq \tilde{W}^{\text{IRPG}}$ is implied by Lemma 3; $W^* \geq \sum_{i\in G}\sum_{k\in D}\sum_{m\in M_{ik}^{\cdot}} c_{kF}d_m/\mathcal{K}_F$ since the right-hand-side number is the lowest possible long-haul transportation cost assuming the highest truck capacity utilization, which also gives $\tilde{W}^{\text{IRPG}} - W^* \leq \sum_{k\in D}(h_0\mathcal{K}_F + c_{kF})T$ by Lemma 4. Let the demand for each commodity in each period d_m be i.i.d. with mean $\bar{d}_{ik} > 0$, $\forall i \in G, k \in D, m \in M_{ik}^{\cdot}$; then $\sum_{k\in D}(h_0\mathcal{K}_F + c_{kF}) \sim o(\sum_{i\in G}\sum_{k\in D}\sum_{m\in M_{ik}^{\cdot}} c_{kF}d_m/T\mathcal{K}_F)$ with probability 1, and the right-hand side of (9) converges to 0 as $|G| \to \infty$.