Deferred Item and Vehicle Routing within Integrated Networks

Karen R. Smilowitz*

Department of Industrial Engineering and Management Sciences

Northwestern University

Technological Institute, 2145 Sheridan Road, Evanston, IL 60208

Alper Atamtürk

Department of Industrial Engineering and Operations Research

University of California, Berkeley

Etcheverry Hall, Berkeley, CA 94720

Carlos F. Daganzo

Department of Civil and Environmental Engineering

University of California, Berkeley

416A McLaughlin Hall, Berkeley, CA 94720

November 7, 2002

*Corresponding author. E-mail: ksmilowitz@northwestern.edu, Phone: 847-491-4693, Fax: 847-491-8005.

Abstract

This paper studies the possible integration of long-haul operations by transportation mode and service level (defined by guaranteed delivery time) for package delivery carriers. Specifically, we consider the allocation of deferred items to excess capacity on alternative modes in ways that allow all transportation modes to be utilized better. Model formulation and solution techniques are discussed. The solution techniques presented produce solutions for large-scale problem instances with up to 141 consolidation terminals and 17 breakbulk terminals. Allowing deferred items to travel by air reduces long-haul transportation costs. These savings increase with the amount of excess air capacity.

Keywords: Package delivery, distribution and logistics, multicommodity network flow, large-scale optimization

1 Introduction

Rapid growth in the package delivery industry has led carriers to offer a wide range of transportation services (i.e., overnight delivery, two-day delivery, etc.) to capture a larger share of the package delivery market and to utilize resources more efficiently. As a result, new opportunities have emerged in the integration of operations by transportation mode (air, ground) and service level (express, deferred). This paper is part of a larger research project that combines continuous approximation and discrete methodologies to study multimodal package delivery systems (see Smilowitz, 2001). It is motivated by the need to study the various degrees of mode and service level integration within the package delivery industry. Smilowitz and Daganzo (2002) show that integrating operations can produce significant savings in local distribution costs. In addition, with integrated operations, some deferred items (3+ day delivery windows) that normally travel by ground vehicles may be sent by air if excess capacity exists. By intelligently choosing which deferred items to shift to the air network, the ground network can be operated more efficiently and overall network capacity can be better utilized.

In particular, this paper addresses the routing of deferred items and ground vehicles; called here the "deferred item and vehicle routing problem" (DIVRP). Modeling and solution techniques appropriate for large–scale instances of the problem are presented. We perform a computational study to evaluate the economic impact of sharing aircraft capacity by express and deferred packages. Our computational experiments suggest that utilizing the excess capacity on aircraft for deferred package transportation can be an attractive cost–savings option for package delivery carriers serving multiple service levels.

Related literature is discussed in Section 2. Section 3 describes the deferred item and vehicle routing network, and Section 4 presents the model formulation. Section 5 presents the solution method used for large-scale instances. Computational results are discussed in Section 6. Section 7 summarizes the key findings and proposes future work.

2 Related literature

The DIVRP is a special case of the service network design problem, with multiple commodities, transshipment points, vehicle types, and service deadlines. For a review of service network design problems with applications to transportation, see Crainic (2000) and Magnanti and Wong (1984). Service network design problems are typically formulated as multicommodity network design problems. These problems have been studied extensively in the literature, both because of the many applications that exist (especially in transportation) and because of the challenges of solving these large-scale mixed integer network problems.

Solving even the linear programming (LP) relaxation of the problems is a challenge due to the excessive size of formulations even for small instances. The LP relaxation does not often produce good bounds and can be highly degenerate for large instances (see Crainic, 2000). To circumvent these problems, a variety of techniques have been proposed in the literature; a sample of problems and techniques is listed below. Solution methods that decompose the item and vehicle routing have been successful in solving large multicommodity network flow problems (see Crainic and Rousseau, 1986). To route aircraft and packages in express delivery networks, Armacost (2000) introduces composite variables to reduce the problem size. Kim *et al.* (1999) use column generation techniques, combined with cutting planes. Holmberg and Yuan (2000) use Lagrangian relaxation, combined with branch and bound techniques, to solve large-scale capacitated network design problems. Cordeau *et al.* (2000) employ Benders decomposition as an alternative to Dantzig-Wolfe decomposition and Lagrangian relaxation. Yano and Newman (2001) develop a solution procedure for a similar service network design problem with a single vehicle type.

The DIVRP, which considers the simultaneous routing of deferred items and long-haul ground vehicles consistent with delivery time windows and availability of excess capacity, is formulated as a multicommodity network design problem with flow balance constraints on the integer design variables. As such, it exhibits many of the problems cited above, and even obtaining a feasible solution for realistic-size instances is computationally challenging. We develop a solution approach to address these issues, comprised of three tasks.

1) Solve linear programming relaxations using path-based reformulation with column generation to overcome prohibitive memory requirements of the arc based formulations.

2) Implement recently developed cutting plane techniques to eliminate part of the fractional solutions of linear relaxation.

3) Develop effective LP rounding heuristics to find feasible integer solutions.

Each task is described in detail in the paper, with comments on the different approaches tested.

3 Network description

The networks in question include two transportation modes: air and ground, and two service levels: express and deferred. Express items are highly time sensitive; deferred items are not. All regional transportation is conducted by ground vehicles (delivery vans, trucks, etc.), but long-haul transportation can be performed by ground (tractor-trailers) and air. In non-integrated delivery networks, express items are transported by air for long-haul trips due to restrictive time constraints.[1] Deferred items are sent over ground long-haul networks. The DIVRP explores potential savings obtained by integrating the transportation of items with different service levels. Given an aircraft schedule we test the impact of routing some deferred items on excess air capacity rather than the ground network.

The time-space network, $G = (N, A)$ is used to model the system, where each node in the set N represents a physical location and an instant of discrete time within the planning horizon (see Figure 1). Let $C \subset N$ denote the consolidation terminal nodes, $B \subset N$ denote the breakbulk terminals nodes, and $H \subset N$ denote the air hub nodes. Consolidation terminals act as the origins and destinations of items.[2] At breakbulk terminals, items are transferred between ground vehicles for more efficient long-haul transportation. At air hubs, items are transferred between aircraft. An item is transported either through the air hubs by aircraft or through breakbulk terminals by ground vehicles.

The set of arcs, A, that link the nodes of the network is partitioned into three subsets: inventory arcs (IA) for holding items at a node until the next time period, ground arcs (GA) for transporting items by ground vehicle or repositioning ground vehicles, and express (air) arcs (EA) for transporting items by air. A ground arc from (l_1, t_1) to (l_2, t_2) for $l_1, l_2 \in C \cup B$ is included in the network if the ground travel time between l_1 and l_2 is $t_2 - t_1$. Similarly, there is an air arc from (l_1, t_1) to (l_2, t_2) for $l_1 \in C, l_2 \in H$ or $l_1 \in H, l_2 \in C$, for all scheduled departure times t_1 if the air travel time between l_1 and l_2 is $t_2 - t_1$. Perfect information and time-dependent demand are assumed.

Because package delivery tends to be periodic in nature (e.g., weekly or daily cycles), an infinite planning horizon can be simulated if one restricts oneself to the exploration of periodic solutions. This is done by introducing periodic boundary conditions, which can be modeled by treating the time dimension as a closed

[1] Express items with nearby destinations may not travel by air. Such items are ignored in this study.
[2] Local distribution routes feeding into consolidation terminals are unaffected by decisions to shift deferred freight.

loop, i.e., by wrapping the network on a cylinder and linking the last time period with the first (see Figure 2). Arcs are placed on the cylinder, with arcs connecting nodes at the end of one rotation to nodes at the beginning of another.

4 Model formulation

The objective of the DIVRP is to minimize marginal item transportation costs by ground and air, vehicle operating costs for ground transportation, and inventory costs. A commodity k, in the commodity set K, is defined by an origin-destination pair, $(l_o, t_o), (l_d, t_d)$. The origin-destination demands are expressed as $d^k > 0$. The origin and destination for a commodity k are denoted by $\mathcal{O}^k$ and $\mathcal{D}^k$, respectively. Items are routed between consolidation terminals either by ground through a series of intermediate breakbulk terminals or by air through an air hub, as shown in Figure 1. Aircraft schedules over the planning horizon are determined exogenously based on express item demand only; it is assumed that aircraft schedules are fixed and excess capacity is given for the flight legs. The routing of ground vehicles (and the resulting capacity on ground arcs) is determined within the DIVRP. The following parameters and decision variables are included:

Parameters

b_a marginal cost of sending an item over arc $a \in A$ (both transportation and inventory arcs)($/item)

c_a cost of sending a vehicle over arc $a \in GA$ ($/vehicle)

v_i ground vehicle capacity for vehicle type $i \in V$ (items)

e_a excess air capacity for arc $a \in EA$ (items)

d^k origin-destination demands with commodity $k \in K$ (items)

Variables

f_a^k amount of commodity $k \in K$ sent over arc $a \in A$, (items, continuous)

x_a^i number of loaded and empty ground vehicles of type $i \in V$ on arc $a \in GA$, (vehicles, integer)

Note that the marginal item cost b_a is the inventory holding cost for $a \in IA$, the air transportation cost for $a \in EA$, and the ground transportation cost for $a \in GA$. Vehicle costs are included only for ground vehicles as aircraft costs are exogenous to DIVRP.

The arc-based formulation of the deferred item and vehicle routing problem (DIVRP) is:

$$min \sum_{k\in K, a\in A} b_a f_a^k + \sum_{a\in GA, i\in V} c_a x_a^i \tag{1a}$$

subject to

$$\sum_{(m,n)\in A} f_{m,n}^k - \sum_{(n,m)\in A} f_{n,m}^k = \begin{cases} -d^k, \text{ if } n = \mathcal{O}^k \\ d^k, \text{ if } n = \mathcal{D}^k \\ 0, \text{ otherwise} \end{cases} \qquad \forall n \in N, k \in K \tag{1b}$$

$$\sum_{k\in K} f_a^k \leq e_a \qquad \forall a \in EA \tag{1c}$$

$$\sum_{k\in K} f_a^k \leq \sum_{i\in V} v_i x_a^i \qquad \forall a \in GA \tag{1d}$$

$$\sum_{(m,n)\in GA} x_{m,n}^i - \sum_{(n,m)\in GA} x_{n,m}^i = 0 \qquad \forall n \in C \cup B, i \in V \tag{1e}$$

$$f_a^k \geq 0 \qquad \forall a \in A, k \in K \tag{1f}$$

$$x_a^i \geq 0, \text{ integer} \qquad \forall a \in GA, i \in V. \tag{1g}$$

We minimize the objective function, which is the sum of vehicle and item transportation costs and inventory holding costs. The flow balance constraints (1b) ensure item flow conservation at each node for all commodities. Built into these equations are delivery time windows implicit in the time-space network: items cannot depart from an origin before the arrival date and must reach the destination by the due date. Vehicle capacity constraints limit allocation to air by excess capacity available (1c) and require sufficient ground vehicles to cover item flow (1d). Flow balance constraints (1e) for ground vehicles create feasible routings between nodes. All decision variables must be non-negative (1f) and ground vehicle variables must satisfy integrality constraints (1g).

Solving a realistic size problem (hundreds or thousands of nodes and arcs) involves an excessive number of variables and constraints and the model presented here cannot be solved without additional refinements. These refinements are presented in the following sections.

5 Solution approach

A two-stage solution approach is proposed.

Stage 1: Lower bound. Solve the linear relaxation of DIVRP, using approaches described in Section 5.1.

Stage 2: Upper bound. Obtain a feasible integer solution from the linear relaxation, using approaches described in Section 5.2.

As explained in these sections, the procedure can be iterated to reduce the gap between bounds.

5.1 Lower bounding techniques

5.1.1 Path-based formulation

The arc–based formulation (1a)–(1g) becomes impractical to implement for large instances due to excessive memory requirements for storing constraints (1b), which are cubic in the number of nodes of the graph. Therefore, a path-based formulation is introduced, where the series of arcs that a commodity traverses from origin to destination is defined as a single path variable. Here the decision variables are the sets of commodity flows over each path and (as with the arc-based formulation) the set of loaded and empty vehicles over the ground arcs. Let $\mathcal{P}^k$ be the set of available origin-destination paths for commodity k. Let λ_P be the fraction of commodity k flowing over path P, comprised of arcs $a \in P$. The path-based formulation is:

$$min \sum_{k\in K} \sum_{P\in\mathcal{P}^k} c_P\lambda_P + \sum_{a\in A, i\in V} c_a x_a^i \tag{2a}$$

subject to

$$\sum_{P\in\mathcal{P}^k} \lambda_P = 1 \qquad \forall k \in K \tag{2b}$$

$$\sum_{k\in K} \sum_{P\in\mathcal{P}^k : a\in P} d^k \lambda_P \le e_a \qquad \forall a \in EA \tag{2c}$$

$$\sum_{k\in K} \sum_{P\in\mathcal{P}^k : a\in P} d^k \lambda_P \le \sum_{i\in V} x_a^i v_i \qquad \forall a \in GA \tag{2d}$$

$$\sum_{(m,n)\in GA} x_{m,n}^i - \sum_{(n,m)\in GA} x_{n,m}^i = 0 \qquad \forall n \in N, \forall i \in V \tag{2e}$$

$$x_a^i \ge 0, integer \qquad \forall a \in GA, \forall i \in V \tag{2f}$$

$$\lambda_P \ge 0 \qquad \forall k \in K, P \in \mathcal{P}^k. \tag{2g}$$

where

c_P is the cost of path P for commodity $k, \forall k \in K, \forall P \in \mathcal{P}^k, c_P = \sum_{a \in P} b_a d^k$

f_a^k is the flow of commodity k on arc a, $f_a^k = \sum_{P \in \mathcal{P}^k : a \in P} \lambda_P d^k$.

The original item flow balancing constraints (1b) from the arc-based formulation are replaced with "convexity constraints" (2b) that ensure the total demand for each commodity is satisfied. The two sets of arc capacity constraints (2c) and (2d) maintain feasible arc flows. Again, the flow of ground vehicles must be balanced (2e) and decision variables must be non-negative and ground vehicle variables must satisfy integrality constraints ((2f) and (2g)).

Column generation

The number of constraints of the path formulation (2a)–(2g) is much smaller than the number of constraints of the arc–based formulation (1a)–(1g). However, the number of path variables λ_P is exponential in the number of arcs. For large problems, rather than enumerating all origin-destination paths, the LP relaxation of the formulation with a feasible subset of the paths (master problem) is solved initially. Using column generation, new path variables are added iteratively if the inclusion of such paths in $\mathcal{P}^k$ could reduce the total cost (pricing problem). A candidate path P is obtained by solving the following single-commodity shortest path problem:

$$min \sum_{a \in A} (b_a - \pi_a) f_a^k - \sigma_k \tag{3a}$$

subject to

$$\sum_{(m,n) \in A} f_{m,n}^k - \sum_{(n,m) \in A} f_{n,m}^k = \begin{cases} -d^k, & \text{if } n = \mathcal{O}^k \\ d^k, & \text{if } n = \mathcal{D}^k \\ 0, & \text{otherwise} \end{cases} \qquad \forall n \in N \tag{3b}$$

$$f_a^k \geq 0 \qquad \forall a \in A. \tag{3c}$$

The objective function of the shortest path problem corresponds to the reduced cost of the candidate paths:

$$\bar{c}_P = \sum_{a \in A} (b_a - \pi_a) f_a^k - \sigma_k,$$

where π_a is the dual variable of the arc capacity constraint for arc a (2c) and (2d) and σ_k is the dual variable for the commodity-specific convexity constraint for commodity k (2b). Since a shift of flow to the new path P does not violate the ground vehicle balance constraints (2e) directly, the dual variables of these constraints do not appear. If the reduced cost of candidate path P, $\bar{c}_P$, is negative, then λ_P is added to the formulation and the master problem is resolved.

At each master problem iteration, dual variables are updated and the pricing problem then checks for new columns. Columns with highly positive reduced costs are removed from the master problem to maintain a manageable problem size. This process is repeated until optimality conditions of the linear relaxation are met for all commodities (all paths have non-negative reduced costs).

Periodic boundary conditions can be implemented easily here. When paths that traverse multiple rotations are added to the master problem, they "wrap around" the cylinder.

Smilowitz (2001) compares the path-based formulations with disaggregated commodities, as shown above, with a path-based formulation with aggregated commodities in which demand for each origin is aggregated over all destinations and due dates. While the aggregated formulation requires the solution of fewer pricing problems at each master problem iteration, more master problem iterations are required for convergence. These results are consistent with earlier results on multicommodity network flow problems by Jones *et al.* (1993), even with ground vehicle balancing constraints added. Consequently a hybrid decomposition algorithm is used. Shortest path trees from an origin to all destinations are obtained with aggregated pricing problems. A candidate path P can be obtained for multiple origin-destination pairs by creating shortest path trees from origin nodes (defined by origin location and time) to all destinations (defined by destination locations and due dates). These trees are disaggregated by destination and a column for each origin-destination path with a negative reduced cost is added to the formulation.

5.1.2 Cutting planes

In conventional applications, cutting planes are used to eliminate fractional LP solutions without eliminating integer solutions and the optimal solution of the resulting LP is used as an improved lower bound; sequences of cuts are then generated until an integer (optimal) solution is found (see, for example, Nemhauser and Wolsey, 1999).

Single capacity flow cuts

As shown in Figure 3, flow out of a node (including air and inventory arcs) must satisfy demand at that node. In Figure 3(a), the node under consideration is a consolidation terminal in the first time period. Inbound flow is equal to item flow arriving at the consolidation terminal from local tours for distribution, plus demand from previous days that have wrapped around the cylinder.

For ease of explanation, cuts are described for the arc-based formulation[3]. In the arc-based formulation, the flow balance constraints at the origin consolidation terminal $n = \mathcal{O}^k$ for commodity k are given by (1b). Let A_n^+ be the set of all arcs outbound from that node and let A_n^- be the set of all inbound arcs. With this notation, (1b) can be written:

$$\sum_{a \in A_n^-} f_a^k - \sum_{a \in A_n^+} f_a^k = -d^k \qquad \text{where } n = \mathcal{O}^k \tag{4a}$$

Let S_n^+ be a subset of outbound ground vehicle arcs for node n. Outbound flow can then be decomposed by flow on arcs in S_n^+ and other flow (including air and inventory arcs). Rearranging the terms of (4a), we obtain

$$\sum_{a \in S_n^+} f_a^k + \sum_{a \in A_n^+ \setminus S_n^+} f_a^k = d^k + \sum_{a \in A_n^-} f_a^k \qquad \text{where } n = \mathcal{O}^k \tag{4b}$$

Let K_n be the set of all commodities with origin n and v be the capacity of vehicles serving consolidation terminals. Let $x_a \in A$ be the flow of these vehicles.[4] By summing constraint (4b) over all commodities with origin n (i.e., in K_n) we obtain the following inequality, which applies to a generic origin node:

$$\sum_{a \in S_n^+} x_a v + \sum_{k \in K_n} \sum_{a \in A_n^+ \setminus S_n^+} f_a^k \geq D_n \qquad \forall n \in \mathcal{O}^k \tag{4c}$$

The inequality holds since the last term of (4b), which is non-negative, is dropped and the first term is replaced by an upper bound.

The mixed integer cut-set inequality (Bienstock and Günlük (1996)) can be used, as a valid cutting plane for our problems. It is defined as follows:

$$\sum_{a \in S_n^+} r x_a + \sum_{k \in K_n} \sum_{a \in A_n^+ \setminus S_n^+} f_a^k \geq r \left\lceil \frac{D_n}{v} \right\rceil \qquad \forall n \in \mathcal{O}^k \tag{5}$$

[3]Since the arc-based and path-based formulations are equivalent, the cuts are valid for both.
[4]We omit the superscript i for simplicity.

where $r = D_n - \left(\lceil \frac{D_n}{v} \rceil - 1\right)V$, and $\lceil \frac{D_n}{v} \rceil$ is the minimum number of vehicles needed if all flow is sent by ground arcs.

To generalize (4c), nodes for the same physical origin location are aggregated over multiple time periods, as shown in Figure 3 (b) and (c). These aggregated nodes contain both nodes and inventory arcs between nodes. Flow is then defined as flow inbound to and outbound from the aggregated node. An equation of the same form as (4c) now holds for the aggregated nodes, and cutting planes similar to (5) can be generated.

Let $\{\hat{x}_a, \hat{f}_a^k\}$ be an optimal solution to the LP relaxation found with column generation. A cut of type (5) is found by letting $S_n^+ = \{a \in A_n^+ : \sum_{k \in K_n} \hat{f}_a^k < r\hat{x}_a\}$.

Additional cuts for multiple capacities

The single capacity inequalities (5) are used to cut off fractional vehicle variables for arcs between consolidation terminals and breakbulk terminals. These arcs typically have smaller capacities than the long-haul arcs between breakbulk terminals. This limits the possible improvements to the LP relaxation as higher capacity long-haul arcs have been ignored. By expanding the aggregated nodes shown in Figure 3 to include the breakbulk terminals accessible from the consolidation terminal, long-haul arcs can be incorporated as well.

Consideration shows that inequalities similar to (4c) can be written when n includes nodes served by more than one vehicle type (breakbulk terminals), and that cut-set inequalities in Atamtürk (2002) can be used to generalize (5) for this case. Such cut-set inequalities can be added for arbitrary sets of ground terminals. In addition to these inequalities, we add residual capacity inequalities (Magnanti *et al.*, 1993) as proposed in Atamtürk and Rajan (2002) for each ground arc.

5.2 Upper bounding techniques

In this section, a summary of the rounding techniques explored to obtain feasible solutions is presented. For further details on these rounding techniques, see Smilowitz (2001).

Rounding approach 1

Integer solutions are obtained by solving two auxiliary linear programs after the LP relaxation of DIVRP is solved. First, all fractional ground vehicle variables are rounded up to the smallest integer. These values are used as lower bounds on the arcs of a network flow model for ground vehicles only (the vehicle flow problem) that minimizes the transportation costs of ground vehicles, without considering item flows. Lower bounds on ground vehicles ensure sufficient capacity for the items. Due to the total unimodularity of the network flow matrix and integral right-hand sides, the solution is integer and yields a feasible vehicle routing. The item flow problem allows path flows to be redistributed over the ground network obtained. This approach is described formally below.

(a) Let $\{\bar{x}_a, \bar{f}_a^k\}$ be the optimal solution to the LP relaxation of DIVRP. Round up fractional ground vehicle values: $\ell_a^i = \lceil \bar{x}_a^i \rceil, \ \ \forall a \in GA, \forall i \in V$.

(b) Solve the vehicle flow problem to balance ground vehicle movements

$$min \sum_{a \in GA, i \in V} c_a x_a^i \tag{6a}$$

subject to

$$\sum_{(m,n) \in GA} x_{m,n}^i - \sum_{(n,m) \in GA} x_{n,m}^i = 0 \qquad \forall n \in N, i \in V \tag{6b}$$

$$x_a \geq \ell_a^i \qquad \forall a \in GA, i \in V. \tag{6c}$$

(c) Reflow items over the fixed ground network with integer values $\{\hat{x}_a^i\}$, where $\hat{x}$ is the solution from step (b).

$$min \sum_{k \in K} \sum_{P \in \mathcal{P}^k} c_P \lambda_P + \sum_{a \in A, i \in V} c_a \hat{x}_a^i \tag{7a}$$

subject to

$$\sum_{P \in \mathcal{P}^k} \lambda_P = 1 \qquad \forall k \in K \tag{7b}$$

$$\sum_{k \in K} \sum_{P \in \mathcal{P}^k : a \in P} \lambda_P d^k \leq e_a \qquad \forall a \in EA \tag{7c}$$

$$\sum_{k \in K} \sum_{P \in \mathcal{P}^k : a \in P} \lambda_P d^k \leq \sum_{i \in V} \hat{x}_a^i v_i \qquad \forall a \in GA \tag{7d}$$

$$\lambda_P \geq 0 \qquad \forall k \in K, P \in \mathcal{P}^k. \tag{7e}$$

While this approach guarantees feasibility of capacity constraints and ground vehicle balancing constraints, it could lead to a heavily over-capacitated ground network and a large optimality gap.

Rounding approach 2

Here near-integer variables are rounded to the nearest integer, either up or down in step (a). The vehicle flow problem is run to ensure vehicle flow balancing constraints are met; however, step (c) may be infeasible if there is insufficient capacity since variables may be rounded down in step (a).

Rounding approach 3

A variation of the first approach ignores step (b) and re–runs the LP relaxation of DIVRP with a fraction of ground vehicle variables fixed. This approach is run iteratively, fixing more variables at each step and running the LP relaxation of DIVRP to check to check for feasible, integer solutions. It is implemented as follows:

(a) Define a set GA' of candidate fractional ground vehicle variables to fix.

(b) Fix *some* variables in GA' at integer values.

(c) Re–run the LP relaxation of DIVRP with selected ground variables fixed.

Repeat until no fractional values remain or an infeasible solution is reached.

The candidate set of ground vehicle variables can be defined in several ways (see Smilowitz, 2001). One example, shown below, is based on the near-integrality of variables $\bar{x}_a^i$:

i. $GA' = \{a \in GA : \min\{\lceil \bar{x}_a^i \rceil - \bar{x}_a^i, \bar{x}_a^i - \lfloor \bar{x}_a^i \rfloor\} < \alpha\}$ for $\alpha \in (0, 1]$.

ii. Select a fraction β of candidate variables where $\beta > \rho$ for some random number $\rho \in [0, 1]$ and fix these variables in step (b).

It is critical to choose an appropriate number of fractional variables to fix. Fixing too many (high values of α and β) may lead to an infeasible solution in terms of vehicle balance and item capacity. Fixing too few (low values of α and β) may lead to a highly fractional solutions, much like the linear relaxation itself, requiring many iterations to reach an integer solution.

Rounding approach 4

In a variation of approach 3, near-integer variables are bounded rather than fixed. This approach provides a feedback loop where item and vehicle flows are adjusted to accommodate the new set of bounds. The same options for selecting and rounding candidate variables from approach 3 are used. Iterative solutions to approach 4 may not produce integer solutions; approach 1 can be employed to obtain integer solutions from an existing fractional solution.

Comparison of rounding approaches

The rounding approaches are applied at different stages of the column generation to improve the upper bounds. In Figure 4, the trade-off between improved bounds and computation time is shown. Lines join results for each test case. In particular, test cases ID1 and ID7[5] are highlighted along with the values of α and β, respectively, for approaches 3 and 4. As expected, lower values of α and β often produce better bounds, yet require longer solution times. Fixing variables rather than bounding them often led to infeasible solutions. Rounding approach 3 appears to produce the best bounds; yet solution times are quite high. For larger problems, it may not be practical to run approach 3; approach 1, with shorter solution times, may be preferred.

6 Results and discussion

A series of test cases, introduced in Section 6.1, is used to analyze the solution method and explore savings from integration. Implementation is discussed in Section 6.2. The solution method performance is discussed in Section 6.3.1, focusing on solution quality as a function of problem size. Section 6.3.2 shows the extent to which excess capacity translates into cost savings.

6.1 Test cases

In this section we describe the test cases used in our computational study. The network characteristics of the test cases are provided in Table 1. There is only one main air hub in the network. The total number of nodes and arcs in the time-space graph is given. Arc counts include ground vehicle and air arcs only, since the number of inventory arcs does not significantly impact computation. The final columns show the initial number of rows and columns in the path-based formulation. A typical business day is divided into smaller time units, consistent with pickup and delivery times, 8 am - 11am, 2 pm - 5pm. Each test case includes a three-day cyclic planning horizon with four time intervals per day. A three-day delivery window

[5]Test cases are described in the next section.

is assumed with items arriving each morning for distribution. Distances between nodes are converted into integer multiples of time units based on vehicle travel speeds. Travel times include slack for expected delays.

Cost data and operating statistics are based on consultations with package delivery carriers and work on package delivery from Han (1984), Han and Daganzo (1985), and Kiesling (1995). Company literature from public package delivery companies complement these studies: Federal Express Corporation (1998a), Federal Express Corporation (1998b), and United Parcel Service (2000). See Table 2 for a description of the cost data and operating statistics. Demand information is derived randomly based on estimated from Smilowitz (2001) where customer density is estimated with housing counts from the 1990 United States census as a proxy for customer locations and population for demand.

6.2 Implementation issues

The upper and lower bounding techniques described in Sections 5 are implemented using CPLEX callable libraries on a SUN Ultra 10 workstation. Solving even the linear programming (LP) relaxations of the models is a big challenge. Finding optimal solutions to LP relaxations of large scale instances takes over a day of CPU time. In order to produce feasible integer solutions we solve the LP relaxations of large instances approximately. In particular, we stop column generation after a maximum number of iterations is reached (or an acceptable gap) and do not generate additional columns once cuts are added. If an LP relaxation is solved approximately, we report the lower bound on the optimal LP objective value obtained from the dual information and reduced costs.

Since the rounding methods can be applied to any LP solution, we apply them several times during column generation and keep the best upper bound. We add cutting planes to move to a better LP solution, re–apply rounding methods, and repeat until either the maximum number of iterations or the time limit is reached. Single and multiple capacity cuts are added iteratively. Since new columns are not added once cuts are introduced, the lower bounds presented in the tables in this section do not reflect the addition of cuts.

6.3 Computational results

The solution method is applied to the test cases to evaluate potential savings. We look first at the performance of solution method, especially when applied to larger problems.

6.3.1 Solution heuristic performance

The test cases are evaluated at varying levels of excess capacity ranging from 0% to 40%. The "capacity level" is the amount of excess capacity available as a percentage of total deferred demand. The cost savings from excess capacity are approximated by changes in lower bounds Δ_{LB} and upper bounds Δ_{UB}. The problems have different cost structures, vehicle capacities and demand levels, so it is not surprising that the gaps between upper and lower bounds in the following discussion do vary, as some data characteristics favor tighter linear relaxations.

Smaller test cases: ID1 - ID4

For small problems, it is possible to compare the performance of the solution heuristic with CPLEX branch-and-bound techniques to gain insight into the quality of the bounds and the resulting savings measures. Problems ID1 through ID4 are solved both with the solution heuristic and with CPLEX with a time limit of one hour. Results are shown in Table 3. In terms of the gap between bounds, CPLEX significantly outperforms the solution heuristic with problem ID1; however, the difference between the two methods decreases with problem size. Therefore, the solution heuristic is run as a supplement to CPLEX. The gaps obtained with a combination of CPLEX and our solution heuristic are shown in the fifth column of Table 3. In 75% of these test cases, adding the heuristic to CPLEX improves the gap between bounds. The combination of the methods appears to be a good option for the mid-sized problems in our data set.

The test cases performed with CPLEX also reveal that the lower bounds are tighter than the upper bounds for problems ID1 and ID2. The same statement can not be made conclusively for problems ID3 and ID4, although results suggest lower bounds are tighter in these cases as well.

It is also possible to compare the path-based formulation of the linear relaxation with the arc-based formulation for problems ID1 through ID4. The final column of Table 3 presents the optimality gaps obtained with the arc-based formulation, using single capacity cuts to improve lower bounds and CPLEX branch and bound techniques to obtain upper bounds.[6] In these cases, as the problem instances increase in the number of nodes and arcs, the gaps become similar to the gaps obtained with the combined solution heuristic and CPLEX.

[6] A ten hour time limit was imposed. Here cuts can be used to improve the lower bounds since all feasible routes from origin to destination are included.

Larger test cases: ID5 - ID11

For large instances a comparison with CPLEX could not be made, because it was not possible to find feasible solutions with the branch–and–bound method of CPLEX. However, using the upper bounding methods presented in Section 5.2 we were able find many feasible solutions to all instances in the data set.

In Table 4 we report the breakdown of the solution time into phases of the algorithm. The time required to run the column generation is divided into time spent generating new columns with pricing problems and time spent solving the master problem. The total number of master problem iterations (i.e., the number of times the master problem is solved with a new set of columns) is listed. Solution times for rounding and cut generation, as well as total elapsed time are listed. It is clear in Table 4 that resolving the LP relaxations of the master problem in column generation and cut generation phases is the most time consuming part of the overall method.

6.3.2 Savings from integration

For the smaller problems, cost savings measured by either changes in upper or lower bounds appear consistent despite large gaps between bounds. In Figures 5 and 6, the cost savings from test cases ID1 through ID4 (measured by the change in the best feasible solution, Δ_{UB}) are shown when CPLEX is used with the solution heuristic, and when only the heuristic is used, respectively. Cost savings are plotted as a function of the amount of air excess capacity available. The figures show that savings increase significantly with available capacity. The rate of increase depends on the number of ground vehicles that can be saved when air transportation is used. Recall that vehicle costs for aircraft are not considered as deferred items are sent via excess capacity on existing flights. It is expected that networks with higher ground vehicle costs will experience greater savings. Cost savings should also depend on other characteristics of a problem instance, such as the average daily demand per origin-destination pair expressed in truckloads. Given the effort to use realistic parameter values in our examples, the results of Figures 5 and 6 should be fairly representative for networks of similar size.

It is more difficult to quantify savings for larger instances due to larger gaps. Unlike the smaller instances, we obtain different values of savings from integration depending on the bounds used (Δ_{LB} versus Δ_{UB}); see Figure 7. Yet, positive savings and definite trends are found for both the upper and lower bounds. Since the upper bounds in the figure are the best solutions that can be obtained (with existing methods), we can conclude that positive savings and positive trends should be the rule if one uses the best method. The lower bound results, and the results of Figures 5 and 6 suggest that this will continue to be true when improved methods are developed.

7 Conclusions

Shifting deferred items to underutilized aircraft has the potential to increase the operational efficiency of package delivery networks serving multiple service levels. However, the computational requirements to obtain feasible solutions to large–scale instances of the problem are quite high. This paper develops techniques that can find feasible solutions for large problem instances. The results of the computational study allow us to quantify the savings from integration. The test cases illustrate that significant savings can be achieved by integrating transportation modes.

With this solution methodology, package delivery carriers can be more proactive and explore a variety of "what if" scenarios. For example, in express package delivery networks, it is often necessary to add redundancy (in terms of excess air capacity) to meet strict delivery deadlines given uncertain demand. Demand variations for air services increase the availability of excess air capacity. As a result, the benefits of integration may increase significantly when express demand is highly variable.

New techniques to solve larger problems should be explored. Since master problem solution time within column generation is prohibitively large, this should be a major focus of future research. In addition, it may be beneficial to consider ways to decompose the problem, possibly by geographic location, or to aggregate nodes by location. The modeling approach for the DIVRP could be extended to networks with multimodal hubs. Then, items can be served by a combination of modes with the transfer occuring at the hub, allowing for greater flexibility in balancing aircraft loads.

References

Armacost, A. (2000). *Composite Variable Formulations for Express Shipment Service Network Design*. PhD dissertation, Massachusetts Institute of Technology.

Atamtürk, A. (2002). On capacitated network design cut–set polyhedra. *Mathematical Programming*, **92**, 425–437.

Atamtürk, A. and Rajan, D. (2002). On splittable and unsplittable capacitated network design arc-set polyhedra. *Mathematical Programming*, **92**, 315–333.

Bienstock, D. and Günlük, O. (1996). Capacitated network flow - polyhedral structure and computations. *INFORMS Journal of Computing*, **8**, 243–259.

Cordeau, J.-F., Soumis, F., and Desrosiers, J. (2000). A Benders decomposition approach for the locomotive and car assignment problem. *Transportation Science*, **34**(2), 133–149.

Crainic, T. (2000). Service network design in freight transportation. *European Journal of Operational Research*, **122**(2), 272–288.

Crainic, T. and Rousseau, J.-M. (1986). Multicommodity, multimode freight transportation: A general modeling and algorithmic framework for the service network design problem. *Transportation Research B*, **20B**(3), 225–242.

Federal Express Corporation (1998a). 1998 SEC Form 10K.

Federal Express Corporation (1998b). Investor relations information packet.

Han, A. F.-W. (1984). *One-to-Many Distribution of Nonstorable Items: Approximate Analytical Models*. PhD dissertation, University of California, Berkeley, Institute of Transportation Studies.

Han, A. F.-W. and Daganzo, C. F. (1985). Distribution of perishable items: transhipments allowed. Research report, University of California, Berkeley, Institute of Transportation Studies.

Holmberg, K. and Yuan, D. (2000). A Lagrangian heuristic based branch-and-bound approach for the capacitated network design problem. *Operations Research*, **48**(3), 461–481.

Jones, K. L., Lustig, I. J., Farvolden, J. M., and Powell, W. B. (1993). Multicommodity network flows: the impact of formulation on decomposition. *Mathematical Programming*, **62**, 95–117.

Kiesling, M. K. (1995). *A comparison of freight distribution costs for combination and dedicated carriers in the air express industry*. PhD dissertation, University of California, Berkeley, Institute of Transportation Studies.

Kim, D., Barnhart, C., Ware, K., and Reinhardt, G. (1999). Multimodal express package delivery: A service network design application. *Transportation Science*, **33**(4), 391–407.

Magnanti, T. and Wong, R. (1984). Network design and transportation planning: Models and algorithms. *Transportation Science*, **18**(1), 1–55.

Magnanti, T., Mirchandani, P., and Vachani, R. (1993). The convex hull of two core capacitated network design problems. *Mathematical Programming*, **60**, 233–250.

Nemhauser, G. and Wolsey, L. (1999). *Integer and Combinatorial Optimization*. Wiley, New York.

Smilowitz, K. (2001). *Design and Operation of Multimode, Multiservice Logistics Systems*. PhD dissertation, University of California, Berkeley, Institute of Transportation Studies.

Smilowitz, K. and Daganzo, C. F. (2002). Cost modeling and solution techniques for complex transportation systems. Under revision in *Transportation Science*. Available at http://www.ce.berkeley.edu/Programs/Transportation/Daganzo/publications.html.

United Parcel Service (2000). Annual report.

Yano, C. and Newman, A. M. (2001). Scheduling trains and containers with due dates and dynamic arrivals. *Transportation Science*, **35**(2), 181– 191.

Figure captions

Figure 1: Time-space representation of distribution network.

Figure 2: Cyclic network with periodic boundary conditions.

Figure 3: Outbound cuts from consolidation terminals: (a) first time period (b) second (c) third.

Figure 4: IP/LP objective value gaps versus computation time.

Figure 5: Cost savings from smaller test cases: ID1 - ID4: solution heuristic with CPLEX.

Figure 6: Cost savings from smaller test cases: ID1 - ID4: solution heuristic only.

Figure 7: Cost savings from larger test cases: ID5 - ID11: solution heuristic only.

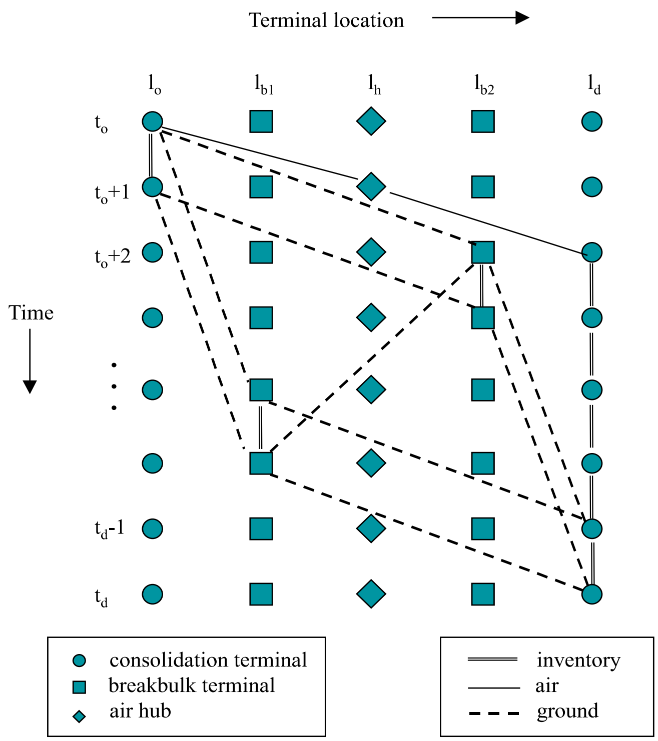

Terminal location
l_o
l_{b1}
l_h
l_{b2}
l_d
t_o
t_o+1
t_o+2
Time
t_d-1
t_d
consolidation terminal
breakbulk terminal
air hub
inventory
air
ground

Figure 1:

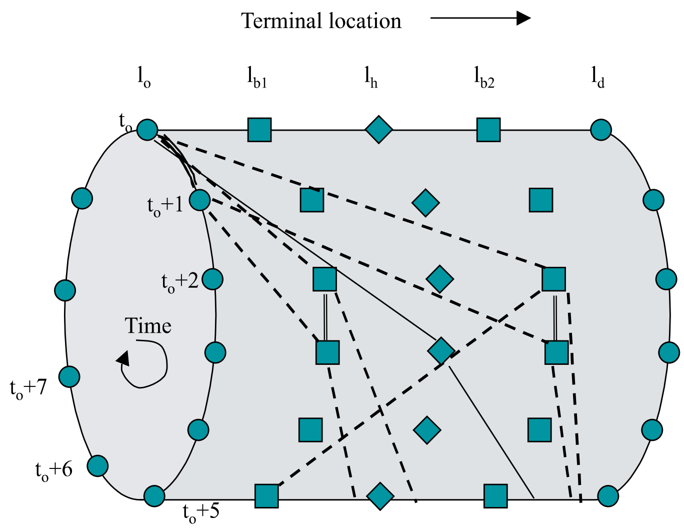


Figure 2:

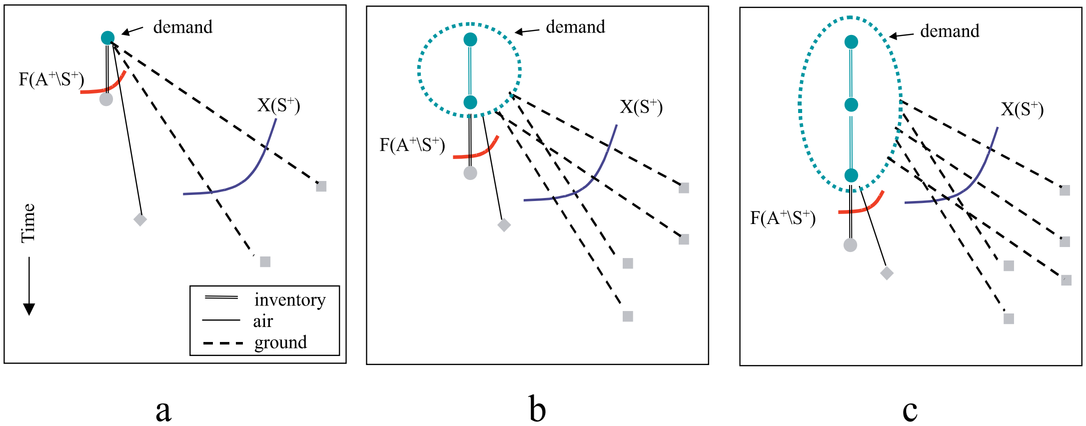
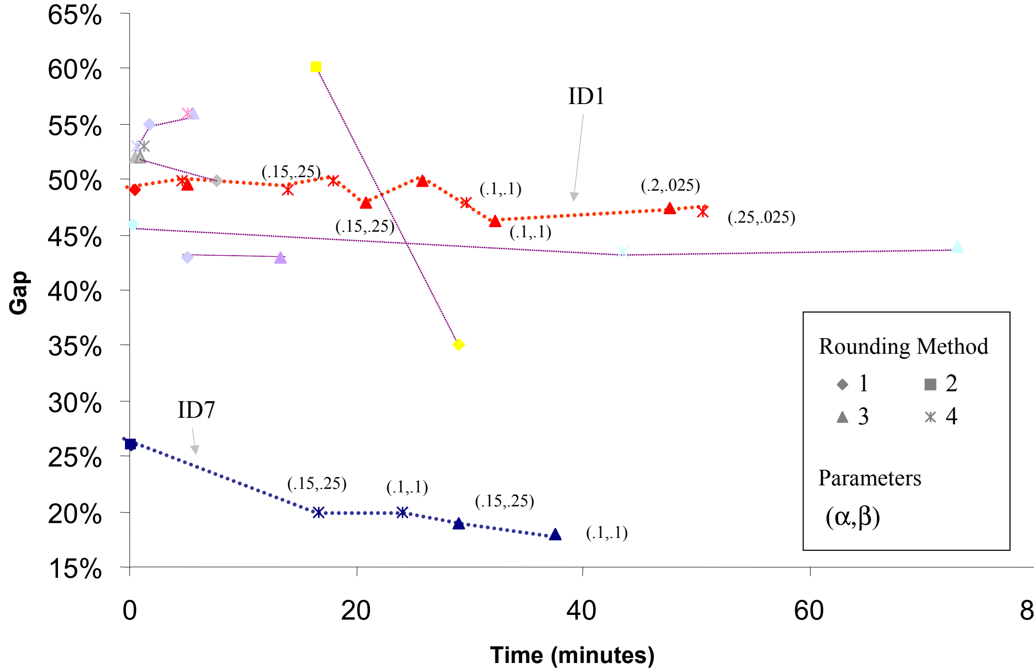

demand
F(A⁺\S⁺)
X(S⁺)
Time
inventory
air
ground
demand
X(S⁺)
F(A⁺\S⁺)
demand
X(S⁺)
F(A⁺\S⁺)
a
b
c

Figure 3:



65%
60%
55%
50%
45%
40%
35%
30%
25%
20%
15%
Gap
0
20
40
60
80
Time (minutes)
ID1
ID7
(.15,.25)
(.15,.25)
(.1,.1)
(.1,.1)
(.2,.025)
(.25,.025)
(.15,.25)
(.1,.1)
(.15,.25)
(.1,.1)
Rounding Method
1
2
3
4
Parameters
(α,β)

Figure 4:

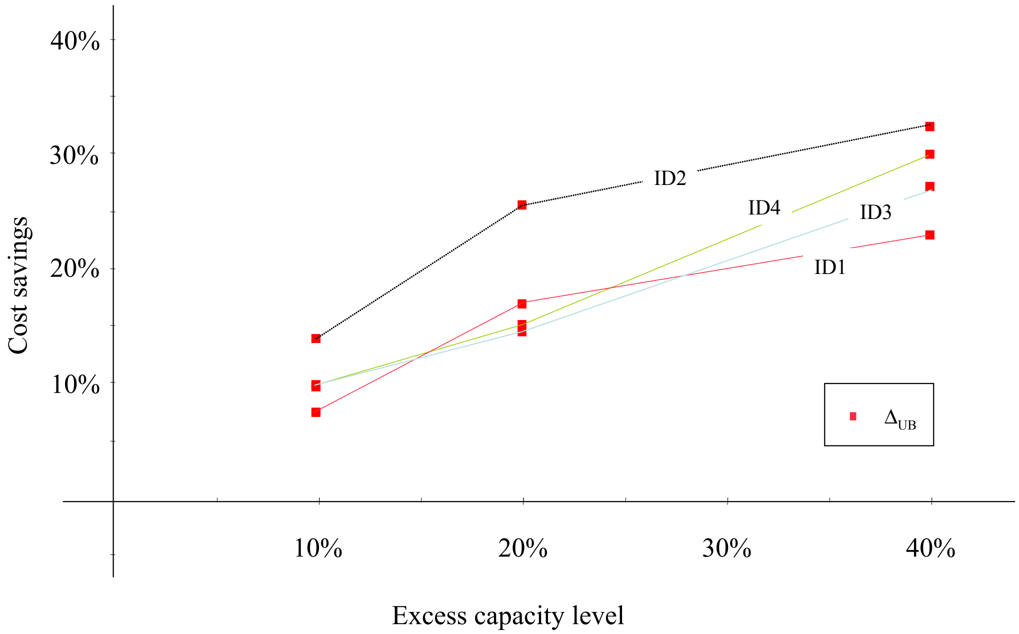

40%
30%
20%
10%
Cost savings
ID2
ID4
ID3
ID1
Δ_UB
10%
20%
30%
40%
Excess capacity level

Figure 5:

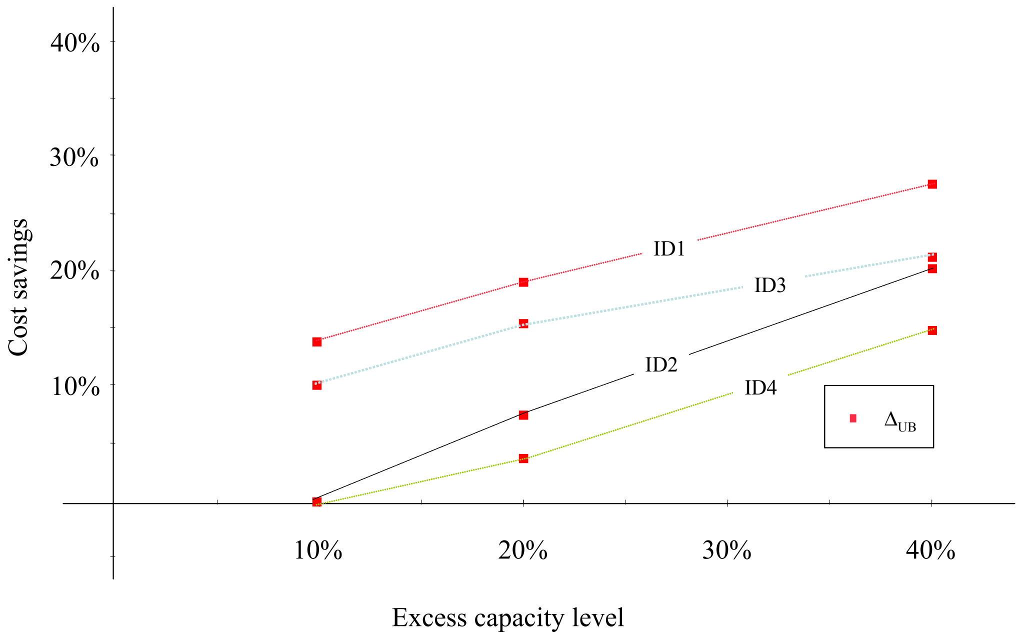

40%
30%
20%
10%
Cost savings
ID1
ID3
ID2
ID4
Δ_{UB}
10%
20%
30%
40%
Excess capacity level

Figure 6:

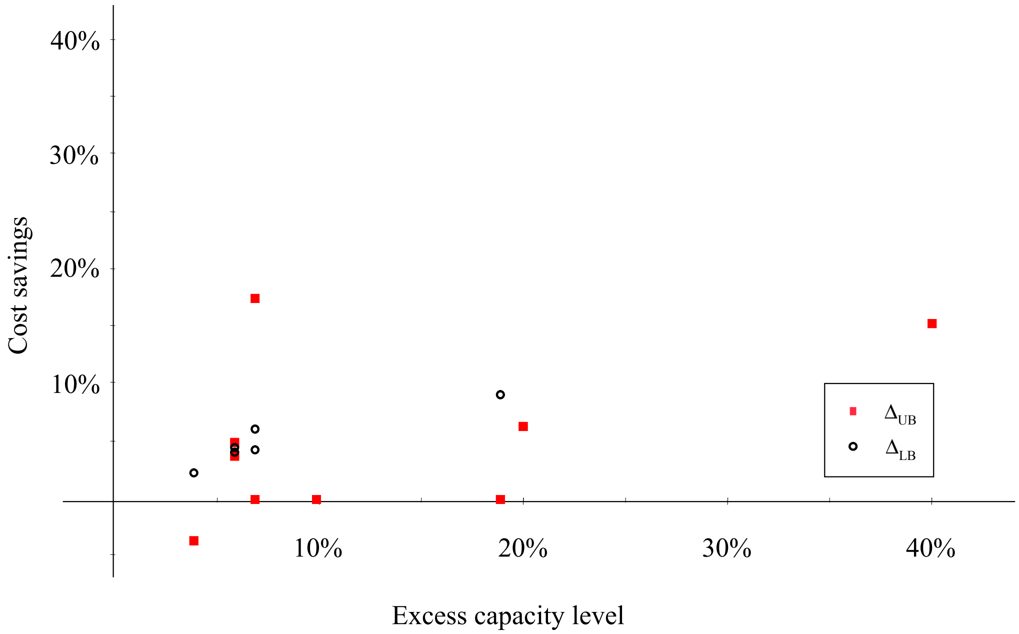

40%
30%
20%
10%
Cost savings
Δ_{UB}
Δ_{LB}
10%
20%
30%
40%
Excess capacity level

Figure 7:

Test case	Consolidation terminals	Breakbulk terminals	Commodities	Nodes	Arcs	Initial rows	Initial columns
ID1	4	2	36	84	240	372	348
ID2	6	4	90	132	720	858	942
ID3	8	4	168	156	1,380	1,224	1,704
ID4	11	5	330	204	2,634	2,208	3,150
ID5	23	8	1,518	384	7,593	6,699	9,450
ID6	38	15	4,218	648	7,425	11,187	12,270
ID7	40	12	4,680	636	5,688	9,744	10,920
ID8	50	15	7,350	792	12,630	16,644	20,742
ID9	62	15	11,346	936	12,327	21,201	24,474
ID10	135	18	54,270	1,848	102,681	94,875	158,274
ID11	141	17	59,220	1,908	35,223	77,751	94,368

Table 1: Description of test cases.[a]

[a] Arc count does not include inventory arcs.

	Vehicle capacities		Marginal item costs				Vehicle operating costs	
Test case	regional ground	long-haul ground	regional ground	long-haul ground	air	holding	regional ground	long-haul ground
	items	*items*	*\$/item*	*\$/item*	*\$/item*	*\$/item*	*\$/veh.mi.*	*\$/veh.mi.*
ID1	50	200	0.1	0.35	0.075	0.001	80	90
ID2	75	400	0.1	0.35	0.075	0.001	80	90
ID3	40	150	0.1	0.35	0.075	0.001	80	90
ID4	100	500	0.1	0.35	0.075	0.001	80	90
ID5	400	750	0.1	0.35	0.075	0.001	80	90
ID6	500	1000	0.05	0.035	0.075	0.001	80	90
ID7	500	1000	0.05	0.035	0.075	0.001	80	90
ID8	500	1000	0.1	0.35	0.075	0.001	80	90
ID9	500	1000	0.1	0.05	0.01	0.0001	55	50
ID10	500	1000	0.05	0.035	0.075	0.001	80	90
ID11	500	1000	0.1	0.05	0.01	0.0001	55	50

Table 2: Cost data and operating statistics.

Test Case	Capacity level	Gap with heuristic	Gap with CPLEX	Gap with CPLEX & heuristic	Gap with Arc-based
ID1	0	40%	9%	6%	4%
	10%	35%	5%	6%	6%
	20%	37%	4%	7%	4%
	40%	34%	13%	6%	NA
ID2	0	29%	31%	26%	11%
	10%	31%	24%	23%	11%
	20%	31%	27%	18%	12%
	40%	37%	35%	23%	NA
ID3	0	18%	12%	11%	18%
	10%	17%	11%	10%	16%
	20%	18%	14%	13%	23%
	40%	24%	21%	11%	NA
ID4	0	27%	23%	26%	19%
	10%	25%	28%	26%	22%
	20%	30%	42%	29%	21%
	40%	39%	33%	34%	NA

Table 3: Gap improvement by combining CPLEX and solution heuristic; smaller test cases.

	Column Generation						
Test case	Pricing problem	Master problem	Iterations	Rounding	Cuts	Total	Gap
ID5	0.03	0.6	11	0.2	14	15	65%
ID6	0.05	6	12	32	36	86	50%
ID7	0.03	4	9	50	5	68	26%
ID8	0.1	17	13	2	250	282	50%
ID9	0.2	41	19	313	797	1,170	44%
ID10	5	1,890	24	145	5,789	7,853	36%
ID11	2	472	25	99	NA	598	55%

Table 4: Time comparisons for larger test cases (minutes).[b]

[b]Cuts not finished for ID10 and ID11 due to time limits.