

Deferred Item and Vehicle Routing within Integrated Networks

Karen R. Smilowitz*, Alper Atamtürk[†] and Carlos F. Daganzo[‡]

October 4, 2002

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 efficient solutions for large-scale problem instances. 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

*Department of Industrial Engineering and Management Sciences, Northwestern University

[†]Department of Industrial Engineering and Operations Research, University of California, Berkeley

[‡]Department of Civil and Environmental Engineering, University of California, Berkeley

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). Since aircraft schedules are determined quarterly or yearly for express package delivery, it is reasonable to assume that aircraft schedules are fixed when determining the routes of deferred items and ground vehicles. Modeling and solution techniques to solve dynamic versions of this problem for large networks are presented. We perform a computational study to evaluate the economic impact of sharing aircraft capacity.

Related literature is discussed in Section 2. Section 3 describes the deferred item and vehicle routing network in further detail, 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. Finally, Section 7 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 (see Crainic (2000)). In addition, LP relaxations are highly degenerate for large instances. 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 new 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. Using this method, efficient solutions are found, even for the largest instances in our data set. The gap between feasible integer solutions and the linear relaxation is large in many instances. However, it is still possible to quantify the savings from integration.

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. In particular, 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 arcs (EA) for transporting items by air. A ground arc from (l_1, t_1) to (l_2, t_2) for $l_1, l_2 \in C \cup B$ will be 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 loop, i.e., by wrapping the network on a cylinder and linking the last time

¹Express items with nearby destinations may not travel by air. Such items are ignored in this study.

²Local distribution routes feeding into consolidation terminals are unaffected by decisions to shift deferred freight.

period with the first (see Figure 2). Item demand is specified by origin, arrival date, destination, and due date. Arcs are placed on the cylinder, with some 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. Figure 1 illustrates possible routes from an origin (l_o, t_o) to a destination (l_d, t_d) . Items can be routed between consolidation terminals either by ground through a series of intermediate breakbulk terminals or by air through an air hub. Aircraft schedules are determined quarterly or yearly for express package delivery; it is assumed that aircraft schedules are fixed and excess air capacity is given for the flight legs in the planning horizon. 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 in the model:

Parameters

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

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

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

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

$d_{(l_o, t_o), (l_d, t_d)}$ origin-destination demands with origin times and delivery due dates

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)

Two equivalent arc-based formulations are derived from two commodity definitions for the set K of commodities. In the first, a commodity k is defined by origin-destination pairs, $(l_o, t_o), (l_d, t_d)$. The origin-destination demands are expressed as $d^k > 0$. The origin and destination associated with a commodity k are denoted by \mathcal{O}^k and \mathcal{D}^k , respectively. In the second, demand for each origin is aggregated over destinations and due dates and a commodity, $k \in K$, is defined by (l_o, t_o) only; and \mathcal{D}^k now represents the set of destinations associated with k .

4.1 Formulation I: disaggregated origin-destination formulation

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

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

subject to

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

$$\sum_{k \in K} f_a^k \leq e_a \quad \forall a \in EA \quad (1c)$$

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

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

$$x_a^i \geq 0, \text{ integer} \quad \forall a \in GA, i \in V \quad (1f)$$

$$f_a^k \geq 0 \quad \forall a \in A, k \in K. \quad (1g)$$

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 and ground vehicle variables must satisfy integrality constraints (1f and 1g).

4.2 Formulation II: demand aggregated by destination

As the problem size increases, the sets of variables and constraints in formulation DIVRP-D become quite large as the number of commodities is quadratic in the number of consolidation terminals and the number of constraints (1b) is cubic in the number of consolidation terminals. The aggregated formulation (DIVRP-A) is more compact with fewer flow variables and fewer item flow balance constraints. The remainder of the formulation is the same, including the integer variable count, since the physical network has not changed.

Table 1 summarizes our computational experience with DIVRP-A and DIVRP-D for two instances with twenty consolidation terminals (CT's) and four breakbulk terminals (BBT's). The reduction in problem size in DIVRP-A translates to a reduction in solution time for the linear relaxation of DIVRP-A ("CPU time" in the Table 1). Integer solutions for both formulations are obtained with the CPLEX³ 6.5.1 solver, using a branch and bound algorithm. Memory and time constraints are the limiting factors in finding an optimal solution. While the optimality gaps are quite small, solving a realistic size problem (hundreds or thousands of nodes and arcs) with either DIVRP-A or DIVRP-D involves an excessive number of variables and constraints and the models 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-A or D, using approaches described in Section 5.1.

³CPLEX is a trademark of ILOG, Inc.

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 formulations (DIVRP-D and DIVRP-A) become impractical to implement for large problems due to excessive memory requirements. 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 λ_P be the fraction of commodity k flowing over path P , comprised of arcs $a \in P$. Let \mathcal{P}^k be the set of available origin-destination paths for commodity k . We begin with the disaggregated path-based formulation:

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

subject to

$$\sum_{P \in \mathcal{P}^k} \lambda_P = 1 \quad \forall k \in K \quad (2b)$$

$$\sum_{k \in K} \sum_{P \in \mathcal{P}^k: a \in P} d^k \lambda_P \leq e_a \quad \forall a \in EA \quad (2c)$$

$$\sum_{k \in K} \sum_{P \in \mathcal{P}^k: a \in P} d^k \lambda_P \leq \sum_{i \in V} x_a^i v_i \quad \forall a \in GA \quad (2d)$$

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

$$x_a^i \geq 0, \text{ integer} \quad \forall a \in GA, \forall i \in V \quad (2f)$$

$$\lambda_P \geq 0 \quad \forall k \in K, P \in \mathcal{P}^k \quad (2g)$$

where

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

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

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 path variables in formulation (2a) - (2g) is exponential in the number of arcs. For large problems, rather than enumerating all origin-destination paths, the LP relaxation of the formulation that has 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. A candidate path P is obtained by solving the following single-commodity shortest path problem:

Pricing Problem

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

subject to

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

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

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

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

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 iteration of the master problem, dual variables are updated and the pricing problem is run to check 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 an aggregated commodity formulation. 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 then 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)). Our application of cutting planes is unconventional

because solving the new LP's exactly is too time consuming. Instead, we solve the new LPs approximately but quickly (as explained below) and then used the results to improve the upper bounds. The upper bounding method is explained in Section 5.2.

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 formulation⁴. In formulation DIVRP-D, 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:

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

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

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

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.⁵ 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:

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

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:

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

where $r = D_n - (\lfloor \frac{D_n}{v} \rfloor - 1)V$, 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 $S_n^+ = \{a \in A_n^+ : \sum_{k \in K_n} \hat{f}_a^k < r \hat{x}_a\}$.

⁴Since the arc-based and path-based formulations are equivalent, the cuts are valid for both.

⁵We omit the superscript i for simplicity.

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 the cut-set inequalities in Atamtürk (2002) can be used to generalize (5) for this case. Therefore, cut-set inequalities can be written for arbitrary sets of ground terminals. In addition to these inequalities, we also add residual capacity inequalities (as proposed in Atamtürk and Rajan (2002) and Magnanti *et al.* (1993)) for each ground arc.

Implementation of cutting planes

Single and multiple capacity cuts are added iteratively after the column generation phase. Once cutting planes are added, further columns are not added to the formulation.⁶ The cuts produced the largest improvements in the LP relaxations in the first five or six iterations when tested with the cases described in Section 6.1. For the large problems, each iteration can take up to forty minutes.

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-D 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. Thus we have a feasible vehicle routing. Next, 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 an optimal solution to the LP DIVRP-D. 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

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

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

⁶Therefore, cuts are not included in the lower bounds presented in Section 6.

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

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

subject to

$$\sum_{P \in \mathcal{P}^k} \lambda_P = 1 \quad \forall k \in K \quad (7b)$$

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

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

$$\lambda_P \geq 0 \quad \forall k \in K, P \in \mathcal{P}^k \quad (7e)$$

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 entire DIVRP with a fraction of ground vehicle variables fixed. This approach is run iteratively, fixing more variables at each step and running DIVRP-D 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) Rerun DIVRP-D 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 common test cases. In particular, test cases ID1 and ID7 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 is used to analyze the solution method and explore potential cost savings from integration. The test cases are introduced in Section 6.1. The performance of the solution method is discussed in Section 6.2.1, focusing on the quality of the solutions produced by the heuristic as a function of problem size. Section 6.2.2 shows the extent to which excess capacity translates into cost savings in long-haul transportation.

6.1 Implementation issues

Attempts are made to obtain realistic parameter estimates. However, as a result of modeling simplifications and imperfect information, there are differences between results obtained here and those found in the industry today. With the methodology developed, package delivery companies can be more proactive and explore a wider range of “what if” scenarios; results provide valuable insight into real world applications. While package delivery carriers have been consulted as part of this research, there has been no direct sharing of private demand and cost information. Therefore, the cost data and operating statistics is based on earlier work on package delivery from Han (1984), Han and Daganzo (1985), and Kiesling (1995). Public company literature from public package delivery companies complement these studies, Federal Express Corporation (1998a), Federal Express Corporation (1998b), and United Parcel Service (2000).

Network descriptions of the test cases are provided in Table 2. 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 Dantzig-Wolfe reformulation. A typical business day is divided into smaller time units, consistent with pickup and delivery times; each test cases includes a three-day cyclic planning horizon with four time intervals per day. A three-day delivery window 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 estimates and operating statistics are presented in Table 3. 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 Computational results

The solution method is applied to the test problems to evaluate potential savings from service and mode integration. We look first at the performance of solution method, especially when applied to larger problems. The solution algorithms are implemented with the CPLEX Callable Library on a Sun Ultra 10 workstation.

6.2.1 Solution heuristic performance

The test cases are compared for varying levels of excess capacity from no excess capacity to upwards of 40% using the solution heuristic. 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 test 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. As problem size increases, this is not possible. Problems ID1 through ID4 are solved both with the solution heuristic and with CPLEX using branch and bound with a time limit of one hour. Results are shown in Table 4. Looking at the gaps 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 4. 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 4 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.⁷ The gaps are significantly higher than the optimality gaps obtained with the arc-based formulation for problems of comparable size shown in Table 1. This can be explained by differences in cost and operating parameters: most importantly vehicle capacities and fixed vehicle costs. 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.

Larger test cases: ID5 - ID11

The CPLEX diagnostic techniques used for smaller problems cannot be used for larger problems since CPLEX cannot even find a feasible solution within one hour. Large gaps between upper and lower bounds are observed for problems ID5 through ID11 in Table 5. Without knowing the true optimal solution (or some estimate from CPLEX), it is difficult to say if the larger gaps are caused by poor upper bounds, poor lower bounds, or both.

Another issue when using the solution heuristic on larger problems is solution time, as shown in Table 5. 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 the rounding approaches and cutting plane methods are listed. The total solution time is provided in the last column. Solution time for linear relaxations within

⁷A 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.

column generation and cut generation appear to be the largest bottleneck in terms of processing time. Rounding approaches 3 and 4 are not used for similar time reasons.

6.2.2 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 excess capacity available. The figures show that savings increase significantly with available capacity. The rate of increase depends most strongly on the relative cost of transporting an item for a mile by air and by truck. (Obviously, if ground transportation were free, relative to air, then the savings would be zero.) Recall that fixed air costs are not considered for these items as deferred items are sent by air via excess capacity on existing aircraft routes. It is expected that networks with higher fixed ground vehicle costs will experience greater savings. Cost savings should also depend on other characteristics of a problem instance, such as the ratio of truck capacity to average airplane excess capacity, and the average daily demand per origin-destination pair expressed in truckloads. An effort has been made to use realistic values for these (and other) parameters in our examples. Therefore, 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 problems because the gaps in the algorithms are larger. Unlike the smaller problems, we obtain different measures of savings from integration depending on the bounds used (Δ_{LB} versus Δ_{UB}). However, Figure 7 shows positive savings and definite trends for both the upper and lower bounds. Since the upper bound 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

For larger problems, the test cases illustrate that savings can be achieved, but the solution method does not reach the necessary level of detail. New solution techniques to solve larger problems should be explored. Since the time required to solve the master problem within column generation is prohibitively large, this should be a major focus of future research. Other relaxations of the problem are possible and could be explored in future work. In addition, it may be beneficial to consider ways to decompose the problem, possibly by geographic location.

In addition, results from smaller test problems suggest the upper bounds are not tight. Rounding heuristics should be refined to account more explicitly for possibility of shifting items to air. Approaches 3 and 4 should hold promise in reducing gaps; however, solution time is a limiting factor on the use of these approaches. Another possible rounding heuristic would be to consider fractional values in cycles in step (a) of approach 1, rather than as individual arcs. Ground vehicle variables can be adjusted in cycles of the network in such a manner that would ensure that items could flow feasibly over the network and ground vehicle balancing constraints could be maintained.

As these improvements are incorporated, the modeling approach for the DIVRP could be extended to examine networks with multimodal hubs. Multimodal hubs would enable inbound and outbound transportation of deferred items to be modally decoupled at the hubs. Items traveling between a given origin/destination pair could then be served by a combination of modes with the mode transfer occurring at the main hub. This variation would provide greater flexibility to balance loads on aircraft into and out of the hub. Another important area of future research would be to

incorporate demand uncertainty into DIVRP models to make the models more useful for day-to-day planning.

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: transshipments 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.

	Test problem 1		Test problem 2	
	DIVRP-D	DIVRP-A	DIVRP-D	DIVRP-A
Problem size				
CT's	20	20	20	20
BBT's	4	4	4	4
Time periods	8	8	12	12
Constraints	20,676	8,472	31,500	12,834
Total variables	49,164	29,436	74,268	47,736
Integer variables	4,704	4,704	7,008	7,008
Linear relaxation				
CPU time (minutes)	1.4	0.9	0.4	0.3
Simplex iterations	24,231	21,465	15,226	13,878
Objective value	101,840	101,840	152,760	152,760
Integer solution				
CPU time (minutes)	433	600	300	300
Branch and bound nodes	18,892	6,140	70	130
Best integer solution	102,350	102,513	154,141	154,141
Best linear solution	101,840	101,840	152,760	152,760
Optimality gap	0.5%	0.7%	0.9%	0.9%
Limiting factor	Memory	Time	Time	Time

Table 1: Problem size and solution statistics by formulation.

Problem	CT's	BBT's	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 2: Description of test case problems ⁸

Problem	Capacities		Marginal costs				Fixed costs	
	regional ground	long-haul ground	regional ground	long-haul ground	air	holding	regional ground	long-haul ground
	<i>items</i>	<i>items</i>	<i>\$/item</i>	<i>\$/item</i>	<i>\$/item</i>	<i>\$/item</i>	<i>\$/veh.mi.</i>	<i>\$/veh.mi.</i>
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 3: Cost estimates 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 4: Improvement in gaps between upper and lower bounds by combining CPLEX and solution heuristic; smaller test cases: ID1 - ID4.

Test case	Column Generation			Bounding		Total	Gap
	Pricing problem	Master problem	Iterations	Rounding	Cuts		
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 5: Time comparisons for larger test cases: ID5 - ID11 (minutes) ⁹

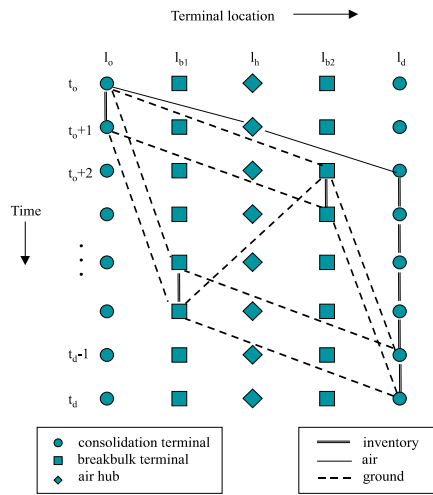


Figure 1: Time-space representation of distribution network.

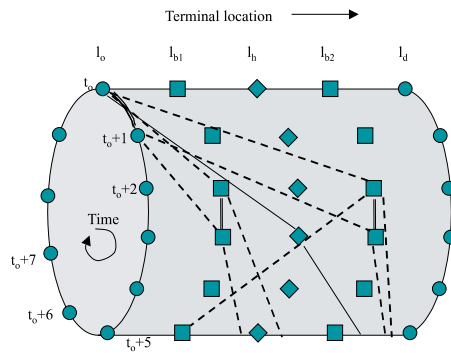


Figure 2: Cyclic network with periodic boundary conditions.

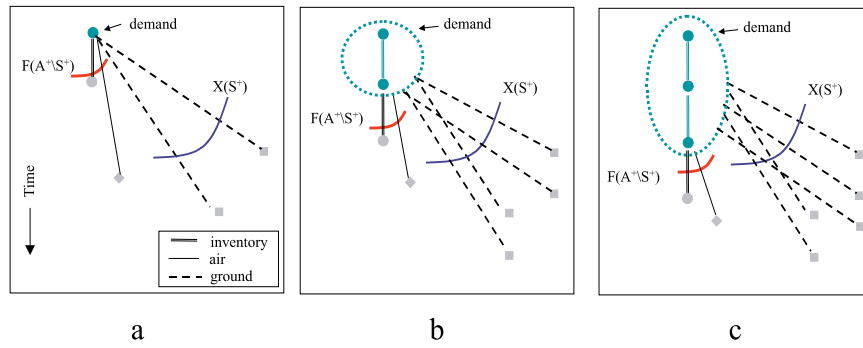


Figure 3: Outbound cuts from consolidation terminals: (a) first time period (b) second (c) third.

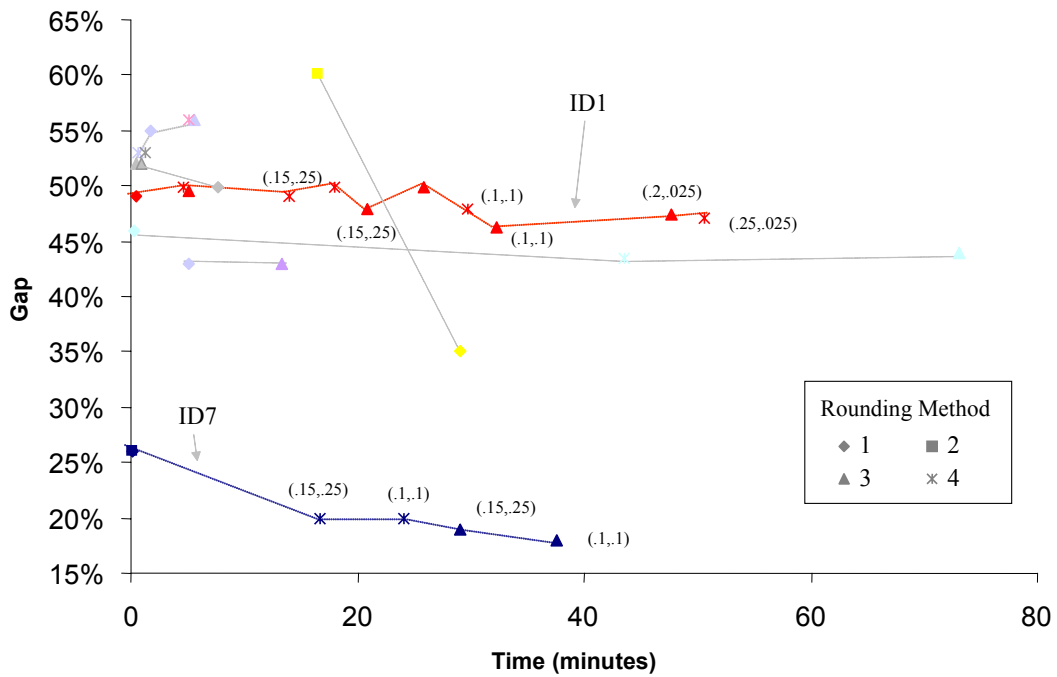


Figure 4: IP/LP objective value gaps versus computation time.

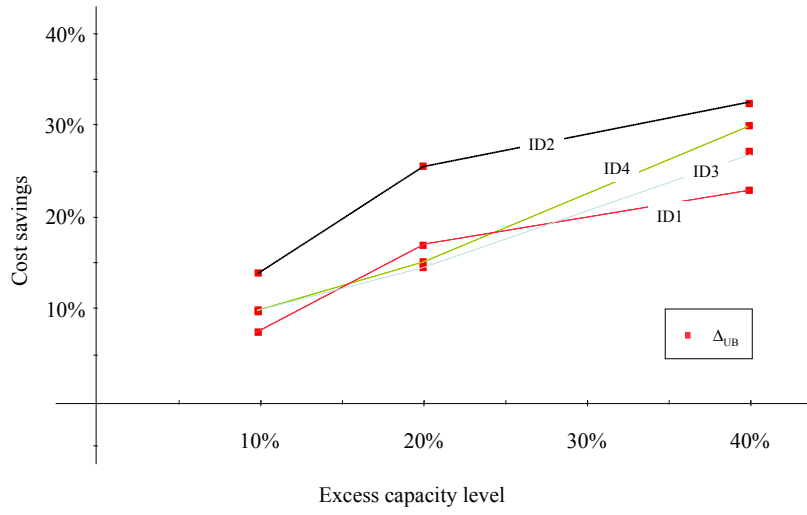


Figure 5: Cost savings from smaller test cases: ID1 - ID4: solution heuristic with CPLEX.

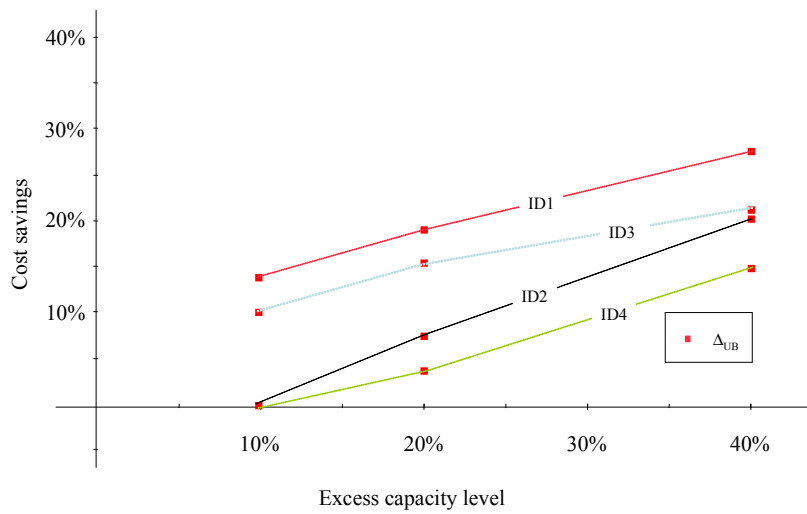


Figure 6: Cost savings from smaller test cases: ID1 - ID4: solution heuristic only.

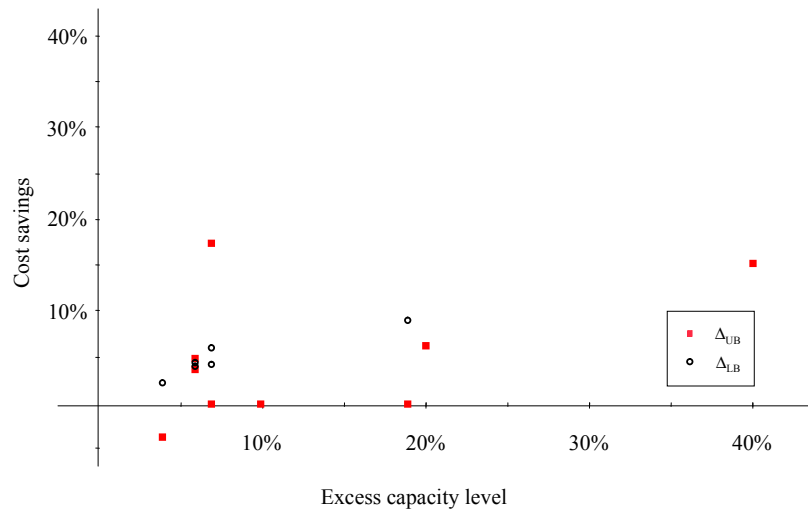


Figure 7: Cost savings from larger test cases: ID5 - ID11: solution heuristic only.