An Integrated Model and Solution Approach for Fleet Sizing with Heterogeneous Assets

Peiling Wu
Joseph C. Hartman
George R. Wilson
Lehigh University

Report No. 01T-005

An Integrated Model and Solution Approach for Fleet Sizing with Heterogeneous Assets

Peiling Wu*, Joseph C. Hartman† and George R. Wilson

Industrial and Manufacturing Systems Engineering Department, Lehigh University,

Mohler Lab, 200 W. Packer Ave., Bethlehem, PA 18015, USA

August 10, 2001

Abstract

This paper addresses a fleet sizing problem in the context of the truck rental industry. Specifically, trucks which vary in capacity and age are utilized over space and time to meet customer demand. Operational decisions, including demand allocation and empty truck repositioning, and tactical decisions, including asset procurements and sales, are explicitly addressed in order to determine the optimal fleet size and mix. The method utilizes a time-space network, common to fleet management problems, which includes capital cost decisions, as it is assumed that assets of different ages carry different costs, common to replacement analysis problems. A two-phase solution approach is presented to solve large-scale instances of the problem. Phase I allocates customer demand among assets through Benders decomposition with a demand-shifting algorithm assuring feasibility in each subproblem. Phase II improves the solution convergence using Lagrangian relaxation, with initial bounds and dual variables provided from the Phase I analysis. Computational studies are presented to show the effectiveness of the approach for solving large problems within reasonable solution gaps.

1 Introduction

Determining the optimal size of a fleet involves strategic, tactical and operational decisions. Strategic decisions include defining markets and customer service levels. Tactical decisions include asset purchase and sale decisions. Finally, operational decisions include asset allocation, assignment, routing and/or scheduling decisions. These hierarchies require trade-offs between the capital expenses of asset procurement and the operational costs of asset utilization, as well as between customer service and cost minimization goals. Obviously, more assets allow a company to achieve a higher level of customer satisfaction at the expense of

*Now with GM Research.

†Corresponding Author: E-mail: jch6@lehigh.edu, Phone: 610-758-4430, Fax: 610-758-4886

higher capital and holding costs, which are nontrivial. Additionally, retaining trucks in service longer may reduce capital expenditures at the risk of increased operating costs and probabilities of break-downs, hurting customer service. Hence, all three levels of decisions are economically interdependent and must be integrated to find the optimal fleet size.

Fleet sizing has traditionally been studied at the operational level through decisions of vehicle assignment or allocation, routing and scheduling. Vehicle ownership costs are often implied by inventory holding costs; however, capital expenses are generally not considered explicitly. Additionally, assets are generally treated as homogeneous (see [Kirby, 1959], [Wyatt, 1969], [Mole, 1975], [Magnanti, 1981], [Mendiratta and Turnquist, 1982], [Jordan and Turnquist, 1983], [Dejax and Crainic, 1987], [Kincewicz et al., 1990], [Beaujon and Turnquist, 1991], [Powell et al., 1995], [Sherali and Tuncbilek, 1997], [Powell and Carvalho, 1998], [Barnhardt et al., 1998] and [Sherali et al., 1999]).

In contrast to operational studies, parallel replacement analysis provides explicit keep and replace (tactical) decisions for heterogeneous assets that operate in parallel, such as a fleet. Replacement problems consider the trade-offs of retaining older assets with higher operating expenses versus replacing them with newer assets, incurring high capital expenses. These decisions are made under expected asset utilization levels without operational details. The reader is referred to [Vemuganti et al. 1989], [Jones et al., 1991], [Karabakal et al., 1994], [Hartman and Lohmann, 1997], [Chen, 1998], [Hartman, 1999 and 2000] and [Chand et al. 2000].

In this research, we address the fleet sizing problem in the context of the truck rental industry. Vehicles vary in capacity and age, with capital, operating and maintenance costs (O&M) dependent on each. Thus, we are motivated to combine fleet management, which addresses operational details, with replacement analysis, which addresses capital investment decisions, in our analysis.

The problem addressed in this paper can be described as follows. Customers reserve a truck, defined by capacity, to be picked-up at a given location on a given date and taken to another location or returned to the same location a number of days later. If the requested truck size is not available, a larger truck may be substituted. (This is referred to as a "down-grade" policy. From rental industry experience, customers are generally willing to accept such an arrangement.) Non-stationary demand and one-way rentals cause inventory imbalances among locations. Thus, trucks must often be moved empty in order to service demand. It is assumed that empty movement travel time is certain, as it is performed by professional drivers. However, loaded movement travel time is customer dependent as different customers take different amounts of time for a given trip between locations.

Specifically, we consider a time-space network where demand requires loaded truck movements between locations through time. Trucks may be moved empty to facilitate demand requests while the fleet size can be increased or decreased over time with asset purchases and salvages. Each loaded movement represents the allocation of a truck (defined by capacity and age) to a customer demand. The allocation of trucks to meet customer demand involves consideration of the age of the truck, as older trucks tend to be more expensive to operate and maintain, as well as the capacity, as larger capacity trucks can be substituted for smaller trucks originally requested by the customer. Thus, this research, which was motivated through work with a truck rental company, attempts to address fleet sizing with the following considerations:

- The fleet is heterogeneous with trucks defined by capacity and age. The explicit consideration of age coincides with replacement analysis research which generally assumes O&M costs and salvage values of assets are age and usage dependent. This is an important consideration in the rental industry as break-downs and asset performance greatly affect customer satisfaction.
- Capital decisions are explicit, including the purchase and sale of trucks over space and time. Purchase costs and salvage values may be time and space dependent to allow for shipping costs (from truck vendors) or to represent different used truck sale markets. Note that we assume that only new assets can be purchased and only used assets can be salvaged.
- Loaded truck travel time is stochastic. We assume a discrete distribution on the travel time (number of days) to move between two locations. The probabilities can be developed from historical data.

The inclusion of these characteristics complicate the analysis greatly, especially with respect to solving large, practical problems. Our application in the rental truck industry requires that daily truck movements be tracked (as customers are given a limit on the number of days required to move between locations, although they may take less time at no penalty). However, parallel replacement analysis requires long horizons for analysis to consider trade-offs in retaining used assets versus purchasing new assets. Thus, the challenge presented is to provide a solution method that can solve large problem instances.

We provide a mathematical programming formulation for this problem, which we term the Rental Fleet Sizing (RFS) problem, although it may be applied to any fleet sizing problem with heterogeneous assets. We determine operational decisions, including customer demand allocation and empty truck repositioning, along with asset purchase and sale decisions over space and time. We solve the formulation with a two-phase approach based on Benders decomposition and Lagrangian relaxation. Algorithms are provided to assure feasible solutions in each phase.

In order to make the analysis viable, we assume that customer demand is known. Although trucks in the rental business are often reserved well in advance and there is a low no-show rate (unlike the car rental business), it is understood that this is an unrealistic assumption. The inclusion of stochastic demand is reserved for future research.

The paper proceeds as follows. The mathematical formulation of RFS is provided in Section 2 with a discussion of the size of the problem. The Phase I (Benders decomposition) solution procedure is described in Section 3 followed by the Lagrangian relaxation procedure (Phase II) in Section 4. Computational results with discussion are provided in Section 5 with conclusions in Section 6.

2 Rental Fleet Sizing (RFS) Problem Formulation

We utilize the following notation in the formulation of RFS:

Parameters

L = number of locations, $l = 1, 2, ...L$;

H = number of time periods in the planning horizon with $t = 1, 2, ...H$;

K = number of truck types with respect to truck capacity, $k = 1, 2, ...K$;

N = maximum age for trucks to be kept in the fleet, $a = 0, 1, ...N$;

$\lambda_{\max}$ = maximum time periods to move loaded trucks from one location to another, $\lambda = 1, 2, ...\lambda_{\max}$;

$\beta_{ll'}(\lambda)$ = percentage of loaded trucks taking λ time periods from location l to l';

$\omega_{ll'}$ = time periods to move empty trucks from location l to l';

$I_l^{k,a}(0)$ = initial fleet of type k trucks of age a at location l;

$d_{ll'}^k(t)$ = demand for type k trucks to move from location l to l' leaving at time period t.

$m^{k,a}(t)$ = inventory costs for idle trucks of type k and age a in period t

$v^{k,a}(t)$ = maintenance cost for loaded trucks of type k and age a in period t

$c^{k,a}(t)$ = operating & maintenance costs for empty trucks, type k, age a in period t

$p^{k,a}(t)$ = purchase prices for each truck of type k and age a in period t

$s^{k,a}(t)$ = salvage values for each truck of type k and age a in period t

$I_l^{k,a}(t)$ = initial fleets of truck type k and age a at location l

Decision Variables

$X_{ll'}^{k,a}(t)$: empty trucks of type k and age a, from location l to l' in period t

$B_l^{k,0}(t)$: number of type k trucks of age a purchased at location l in period t

$S_l^{k,a}(t)$: number of type k trucks of age a sold out at location l in period t

$I_l^{k,a}(t)$: number of type k trucks of age a held in inventory at location l at the end of period t, $t \geq 1$

$\overline{d}_{ll'}^{k,k',a}(t)$: loaded flow with down-grade k', of type k trucks of age a, from location l to l' in period t

The mathematical programming formulation for RFS follows. The objective function minimizes operating and maintenance costs for all trucks (loaded, empty and idle) and purchase costs, less salvage revenues for all the trucks in the fleet within the overall time-space network:

$$min \sum_{t=1}^{H}\sum_{k=1}^{K}\sum_{a=0}^{N}\sum_{(l,l')\in L^2}(c^{k,a}(t)\omega_{ll'}X_{ll'}^{k,a}(t) + v^{k,a}(t)\sum_{\lambda=1}^{\lambda_{\max}}\lambda\beta_{ll'}(\lambda)\sum_{k'=1}^{k}\overline{d}_{ll'}^{k',k,a}(t))$$
$$+\sum_{t=1}^{H}\sum_{k=1}^{K}\sum_{a=0}^{N}\sum_{l=1}^{L}(m^{k,a}(t)I_l^{k,a}(t) + p^{k,a}(t)B_l^{k,a}(t) - s^{k,a}(t)S_l^{k,a}(t))$$

subject to:

Demand allocation with the down-grade policy:

$$\sum_{k'=k}^{K}\sum_{a=0}^{N}\overline{d}_{ll'}^{k,k',a}(t) = d_{ll'}^{k}(t) \quad \forall k=1,2,...K,\ l,l'=1,2,...L,\ t=1,2,...H \tag{1}$$

Flow balance at each location node in time:

$$\begin{aligned}&\sum_{l'=1}^{L}\sum_{\lambda=1}^{\min(t,\lambda_{max})}\beta_{l'l}(\lambda)\sum_{k'=1}^{k}\overline{d}_{l'l}^{k',k,a}(t-\lambda)+\sum_{l'=1}^{L}X_{l'l}^{k,a}(t-\omega_{l'l})^{+}+I_{l}^{k,a}(t-1)\\ &+B_{l}^{k,a}(t)=\sum_{l'=1}^{L}\sum_{k'=1}^{k}\overline{d}_{ll'}^{k',k,a}(t)+\sum_{l'=1}^{L}X_{ll'}^{k,a}(t)+I_{l}^{k,a}(t)+S_{l}^{k,a}(t)\\ &\forall k=1,2,...K,\ a=0,1,...N,\ l=1,2,...L,\ t=1,2,...H\end{aligned} \tag{2}$$

Restricting purchases to only new trucks:

$$B_{l}^{k,a}(t)=0 \qquad \forall k=1,2,...K,\ l=1,2,...,L,t=1,2,...H,a=1,2,..N \tag{3}$$

Restricting sales to only used trucks:

$$S_{l}^{k,0}(t)=0 \qquad \forall k=1,2,...K,\ \ l=1,2,...,L,\ t=1,2,...H \tag{4}$$

And restricting *all variables to be non-negative*:

$$\begin{aligned}&\overline{d}_{ll'}^{k,k',a}(t)\geq 0,\ X_{ll'}^{k,a}(t)\geq 0\ (l\neq l'),\ B_{l}^{k,a}(t)\geq 0,\ S_{l}^{k,a}(t)\geq 0,\ I_{l}^{k,a}(t)\geq 0\\ &\forall (l,l')\in L^{2},\ t=1,2,...H,\ k,k'=1,2,...K,\ a=0,1,...,N\end{aligned}$$

Demand allocation (1) is a characteristic constraint in the RFS model. Given customer demand for a certain type of truck, the loaded truck flows are allocated with respect to vehicle age and the down-grade policy. This complicates solving large problems as the asset allocation is not readily decomposable by truck type.

Note that the flow balance constraints (2) in the RFS model present a sub-network flow structure as shown in Figure 1. The figure illustrates three different types of truck movements: empty (dotted line), local and one-way loaded movements (solid line). Loaded movements are associated with travel time probabilities, which are differentiated by arc boldness. Idle trucks can be treated as inventory (double line) from one time period to the next time period. Additionally, for each time-space node, a purchase node and a salvage node are included to model asset acquisition and disposal decisions. Flow is balanced over time and space, in that all the inflow plus truck inventory (including purchase and salvage) must equal the outflow of trucks and ending inventory in every time period at each location.

We allow non-integer solutions for RFS as our application requires thousands of vehicles. To give an idea of the problem size, a typical RFS model has $O(L^2HK^2N)$ variables and $O(L^2HK)$ constraints, where L,H,K,N represents the number of locations, time periods, and vehicle types, and maximum vehicle age in

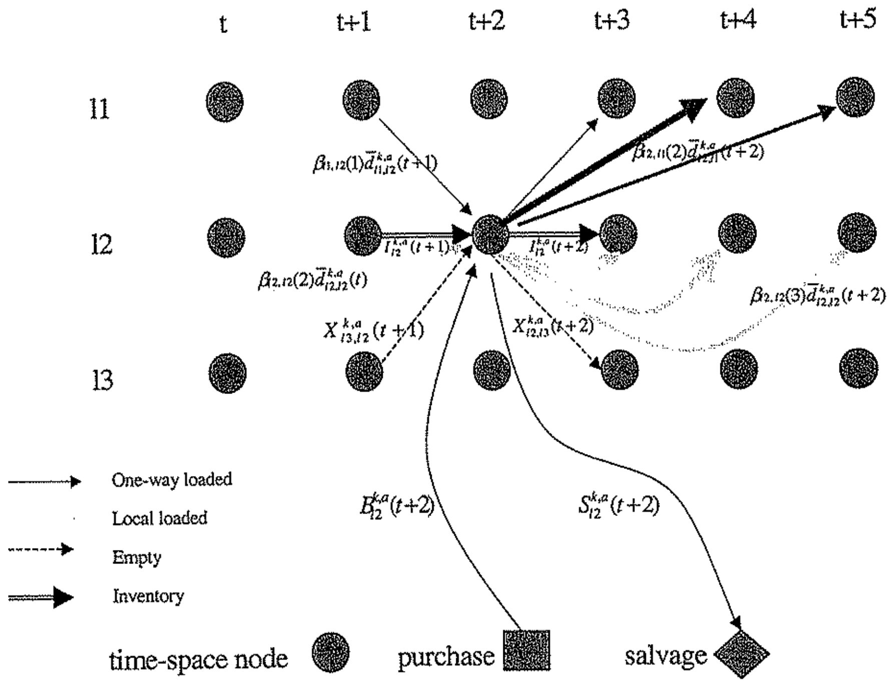


Figure 1: Illustration of flow balance constraint at time-space node $(t+2, l2)$

service. For 60 time periods and 25 locations, a fleet sizing problem with three truck types and a maximum service age of three years has 130,500 constraints and 1,368,000 variables. For realistic problems, RFS cannot be solved directly with linear programming. Also note that the network flow submatrices embedded in the flow balance constraints (2) cannot be exploited when solved as a single problem. Meanwhile, demand allocations with the down-grade policy are very expensive in terms of allocation weight (the numbers of allocation variables versus the total variable numbers). However, if demand allocation variables $\overline{d}_{ll'}^{k,k',a}(t)$ are fixed, allocation constraints can then be separated from the flow balance constraints, which subsequently reduce to pure network flow subproblems. We recognize that this idea can be implemented with Benders decomposition procedure, which we utilize in Phase I of our solution algorithm.

3 Phase-I: Benders Decomposition

Benders decomposition [Benders, 1962], also known as Benders' partitioning or row decomposition regarding coupling variables, has been applied, with success, to both linear and mixed integer programming problems in a variety of applications. These include solving two-stage recourse problems [Van Slyke and Wets, 1969], the multicommodity distribution system design problem [Geoffrion and Graves, 1974], the quadratic assignment problem [Bazaraa and Sherali, 1980], hierarchical production planning problems [Aardal and Larsson, 1990] and the parallel replacement problem [Chen, 1998], among others.

In general, Benders decomposition works in an iterative fashion. The configurative variables (usually the complicating variables) are determined from the Benders master problem. A myopic optimum is found, with the complicating variables fixed in the subproblems. Optimality cuts are then generated from the subproblem dual variables and added into the master problem. As a coordinator, the master problem re-solves for the complicating variables based upon the dual information obtained from the subproblems. Benders procedures converge in a finite number of iterations, provided that the subproblem(s) and master problem are solved to optimality [Benders, 1962]. The procedure may also be stopped when certain criteria are met, such as the gap between lower and upper bounds and/or a prespecified CPU time.

3.1 Benders Implementation of the RFS model

For the RFS problem, the motivation for our Benders implementation originates from the embedded network flow structure, as well as the difficulty of the demand allocation constraints (1). We observe that if the complicating demand allocation variables, $\overline{d}_{ll'}^{k',k,a}(t)$, are fixed, the RFS problem can be decomposed into a set of minimum cost network flow subproblems, one for each truck type k and age a. Note that the time periods and locations in each of these subproblems are intertwined within the entire transportation network, hence there is no direct way to decompose by time and location. Also, demand allocation constraints are the complicating constraints which contribute drastically to solution time. Given customer demand for a certain type of truck, the loaded truck flows can be allocated based on vehicle age and the down-grade policy. As a result, the demand allocation constraints are separated and solved in the master problem.

The Benders primal network flow subproblems, PS(k, a), are defined as follows:

$$\min \sum_{t=1}^{H} \sum_{(l,l')\in L^2} c^{k,a}(t)\omega_{ll'} X_{ll'}^{k,a}(t)$$
$$+ \sum_{t=1}^{H}\sum_{l=1}^{L}(m^{k,a}(t)I_l^{k,a}(t) + p^{k,a}(t)B_l^{k,a}(t) - s^{k,a}(t)S_l^{k,a}(t))$$

subject to:

$$\sum_{l'=1}^{L} X_{ll'}^{k,a}(1) + I_l^{k,a}(1) - B_l^{k,a}(1) + S_l^{k,a}(1)$$
$$= I_l^{k,a}(0) - \sum_{l'=1}^{L}\sum_{k'=1}^{k} \overline{d}_{ll'}^{k',k,a}(1) \qquad \forall l = 1,2,...L \tag{5}$$

$$\sum_{l'=1}^{L} X_{ll'}^{k,a}(t) + I_l^{k,a}(t) - \sum_{l'=1}^{L} X_{l'l}^{k,a}(t-\omega_{l'l})^{+} - I_l^{k,a}(t-1) - B_l^{k,a}(t) + S_l^{k,a}(t) =$$
$$\sum_{l'=1}^{L}\sum_{\lambda=1}^{\min(t,\lambda_{\max})} \beta_{l'l}(\lambda) \sum_{k'=1}^{k} \overline{d}_{l'l}^{k',k,a}(t-\lambda) - \sum_{l'=1}^{L}\sum_{k'=1}^{k} \overline{d}_{ll'}^{k',k,a}(t) \tag{6}$$
$$\forall l = 1,2,...L,\ t = 2,3,...H$$

$$X_{ll'}^{k,a}(t) \geq 0, \quad B_l^{k,a}(t) \geq 0, \quad S_l^{k,a}(t) \geq 0, \quad I_l^{k,a}(t) \geq 0$$
$$\forall (l, l') \in L^2, \; t = 1, 2, ...H$$

Previously defined Constraints (3) and (4) are also included in the subproblems, as only new assets can be purchased. This limitation on purchasing or leasing used trucks (Constraint (3)) can lead to infeasible subproblems as demand allocated by the master problem, as in the right-hand side of Constraints (5) and (6), may be too great to be handled by the assets in a given subproblem.

To alleviate this concern, valid inequalities [Wu et al. 2000] are derived from the flow balance of the time-space network. They are based on the facts that (1) given an initial fleet of trucks, the demand allocated to each location in the first time period cannot exceed its initial inventory; and (2) in any given time period, the feasible demand allocation cannot be more than the sum of incoming empty trucks, idle trucks from the previous period and returning loaded trucks, up to the current period. The proofs are given in [Wu et al. 2000].

Theorem 1 *Given $k \leq K$ and $a \geq 1$, at each location l,*

$$I_l^{k,a}(0) \geq \sum_{l'=1}^{L} \sum_{k'=1}^{k} \overline{d}_{ll'}^{k',k,a}(1) \quad \forall l = 1, 2, ...L \tag{7}$$

are valid inequalities to the RFS model.

Theorem 2 *Given $k \leq K$ and $a \geq 1$, for all $t > 1$,*

$$\sum_{l=1}^{L} I_l^{k,a}(0) \geq \sum_{(l,l')} \sum_{\bar{t}=1}^{t} \sum_{\lambda=t+1-\bar{t}}^{\lambda_{\max}} \beta_{l'l}(\lambda)(\sum_{k'=1}^{k} \overline{d}_{l'l}^{k',k,a}(\bar{t})) \quad \forall t = 2, 3, ...H \tag{8}$$

are valid inequalities to the RFS model.

Including the derived inequalities, the restricted Benders master problem (PM) is defined as follows:

$$\min \sum_{k=1}^{K} \sum_{a=0}^{N} \sum_{t=1}^{H} \sum_{(l,l') \in L^2} v^{k,a}(t) \sum_{\lambda=1}^{\lambda_{\max}} \lambda \beta_{ll'}(\lambda) \sum_{k'=1}^{k} \overline{d}_{ll'}^{k',k,a}(t) + \sum_{k=1}^{K} \sum_{a=0}^{N} z^{k,a}$$

subject to valid cuts (7) and (8) and:

$$z^{k,a} \geq \sum_{l=1}^{L} \theta_l^{k,a,n} (I_l^{k,a}(0) - \sum_{l'=1}^{L} \sum_{k'=1}^{k} \overline{d}_{ll'}^{k',k,a}(1))$$

$$+ \sum_{l=1}^{L} \sum_{t=2}^{H} \pi_{l,t}^{k,a,n} (\sum_{l'=1}^{L} \sum_{\lambda=1}^{\min(t,\lambda_{\max})} \beta_{l'l}(\lambda) \sum_{k'=1}^{k} \overline{d}_{l'l}^{k',k,a}(t-\lambda) - \sum_{l'=1}^{L} \sum_{k'=1}^{k} \overline{d}_{ll'}^{k',k,a}(t))$$

$$\forall k = 1, 2, ..K, \;\; a = 0, 1, ...N, \;\; n = 1, 2, ...\# \text{ of generated optimality cuts}$$

$$\sum_{k'=k}^{K} \sum_{a=0}^{N} \overline{d}_{ll'}^{k,k',a}(t) = d_{ll'}^{k}(t) \quad \forall k = 1, 2, ...K, \;\; l, l' = 1, 2, ...L, \;\; t = 1, 2, ...H$$

$$\overline{d}_{ll'}^{k,k',a}(t) \geq 0, \;\; z^{k,a} \text{ unrestricted},$$

$$\forall k, k' = 1, 2, ...K, \;\; a = 0, 1, ...N, \;\; l, l' = 1, 2, ...L, \;\; t = 1, 2, ...H$$

Again, $\theta_l^{k,a}$ and $\pi_{l,t}^{k,a}$ are the dual variables associated with (5) and (6). Solving the master problem with the valid inequalities (7) and (8) dramatically decreases the requirement of purchasing used trucks in RFS (the purchases are infeasible). However, the valid inequalities cannot guarantee, in general, feasibility of the subproblems beyond $t = 1$, as the valid inequalities (8) cannot capture all of the dynamics of asset movements.

In [Wu et al. 2000], we proposed a demand-shifting feasibility algorithm which shifts excess demand to other sources of capacities (from the infeasible subproblem to a different subproblem). The algorithm, by construction, guarantees a feasible solution to be found quickly, and no additional computational burden is placed on solving the master problem. It is summarized in the following (for details, the reader is referred to [Wu et al. 2000]):

0) Start from the subproblem with the maximum age group.

1) Solve the subproblem, allowing excess (infeasible) capacity to be acquired.

2) At each time-space node, determine the excess capacity corresponding to excess allocated demand. Eliminate the excess capacity at each node by tracing its path to the end of the horizon following a least-cost driven priority structure of outgoing arcs: salvage, held inventory, empty movement and loaded movement.

3) Remove the entire path of excess demand for a used truck group and shift it to the next subproblem with newer trucks. In essence, this reallocates excess demand to the next subproblem.

4) Re-solve the subproblem of the newer age group with the new allocation; go back to 1).

5) The above procedure is repeated until reaching the age 0 (new) group which it is allowed to purchase additional capacity.

It was proved [Wu et al. 2000] that:

Theorem 3 *The demand-shifting subroutine produces a feasible solution to any subproblem PS(k, a), $a > 0$, for any demand allocation $\overline{d}_{ll'}^{k',k,a}(t)$.*

H-L	Master Iter			Ratio Gap			CPU (Sec)		
	Max	Min	Avg	Max	Min	Avg	Max	Min	Avg
30-10	11	3	8	4.97%	4.24%	4.58%	110.9	12.3	58
30-20	2	2	2	4.39%	1.87%	2.84%	20.3	16.7	20
30-30	2	2	2	2.11%	1.82%	1.93%	74.3	63.7	68
40-10	12	3	8	4.96%	4.33%	4.69%	266.0	10.3	107
40-20	3	2	2	3.69%	2.76%	3.08%	372.1	29.2	137
40-30	2	2	2	2.45%	2.10%	2.31%	124.6	105.5	114
50-10	9	3	6	4.98%	3.81%	4.58%	121.4	10.3	64
50-20	3	2	3	4.29%	2.62%	3.33%	360.8	47.5	189
50-30	2	2	3	2.98%	2.60%	2.81%	216.8	153.3	175
60-10	8	2	6	4.90%	4.25%	4.56%	172.6	8.9	94
60-20	4	2	3	4.81%	3.99%	4.50%	1093.2	53.4	338
60-30	3	2	3	3.12%	2.60%	2.79%	1949.0	188.8	557

Table 1: Computational Results for $K = 2$ and $N = 3$

Theorem 4 *The worst-case complexity of the (single or two-step) demand-shifting subroutine is $O(L^2H)$ in one subproblem iteration of the RFS problem.*

3.2 Preliminary Computational Results

As shown in the computational studies of Wu et al. (2000), the demand-shifting algorithm leads to better solution gaps much more quickly than the traditional approach of adding feasibility cuts to the master problem for every infeasible subproblem in each iteration. Computational results for larger sized problems on an IBM PC 300GL personal computer with 733 Mhz Pentium III processor and 512 MB RAM are provided here. In Table 1, performance statistics are summarized for problems varying in horizon length (H) and number of locations (L). All problems assume two asset types (K=2) each with a maximum service of 3 years (N=3). For each combination of H and L, five problems were randomly generated to produce the statistics. (A description of the random parameters utilized in the experiments is given in the appendix.) The algorithm was terminated when the ratio gap fell below 5%.

Results from larger sized instances, with K=3 and N=3 are given in Figure 2. Typically, the ratio gap is reduced below 10% after the first iteration. Each of the four instances shown reaches an optimality gap at or below 2.6% in 10 iterations.

Computational experience shows that the demand-shifting algorithm with valid flow cover inequalities leads to very efficient solutions within a traditional Benders decomposition framework. The benefits of the demand-shifting feasibility algorithm are numerous in that: (i) each subproblem is solved only once in each

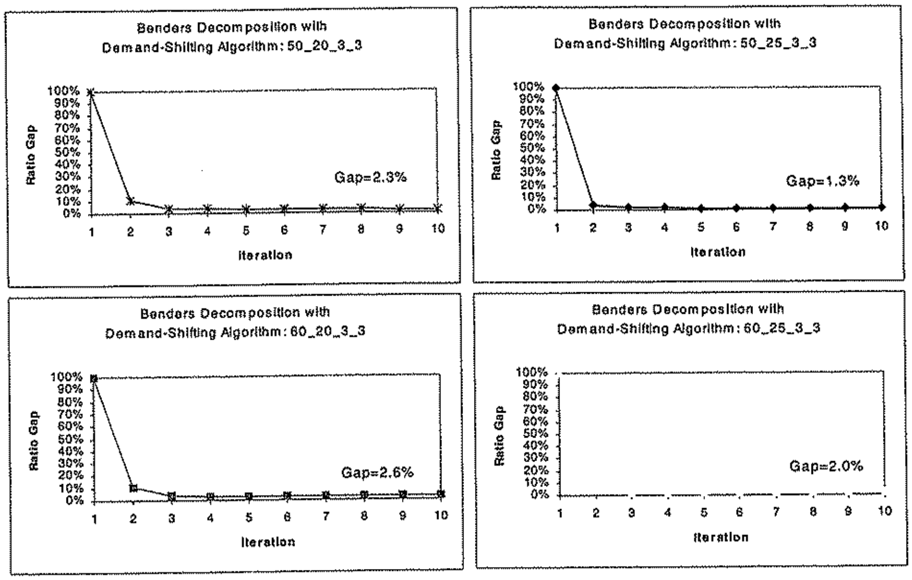


Figure 2: Decreasing solution gap for K=3 and N=3 problem instances.

Benders iteration as feasibility is assured with the demand-shifting algorithm; (ii) no feasibility cuts are added into the master problem; (iii) the master problem need not be reformulated and solved repeatedly to find feasible solutions of the subproblems.

Despite the good computational success, it is well known in the literature that solving the Benders master problem is computationally expensive [Magnanti and Simpson, 1978], [Magnanti and Wong, 1981], [Minoux, 1984], [Aardal and Larsson, 1990], and [Holmberg, 1994] with increasing iterations. Thus, we are motivated to terminate the procedure after a few iterations when it becomes computationally burdensome. We can use the initial bounds and multipliers from Phase I and resume the solution procedure with Lagrangean relaxation and subgradient optimization, which has less intense memory requirements.

4 Phase II: Lagrangean Relaxation

The Lagrangean relaxation technique is a method of solving certain problems with coupling constraints. The idea is to dualize the bundling constraints and absorb them into the objective function with a penalty of Lagrangean multipliers. Hopefully, the resulting Lagrangean dual is easier to solve. Lagrangean relaxation has been applied successfully to a variety of linear and mixed integer linear programs (see [Bararaa and Goode 1979], [Graves 1982], [Fisher 1985], [Reeves 1993], [Karabakal et al. 1994] and [Holmberg et al. 1998], among others).

Essentially, there are three key issues to be addressed: (i) which constraints to relax; (ii) how to find a good starting point such as initial upper bound and/or the Lagrangean multipliers; and (iii) how to update the upper bound and Lagrangean multipliers. In particular, question (iii) is problem-dependent, which may only succeed for some specific structured problems.

4.1 Lagrangean Relaxation of RFS

In the RFS model, we relax the demand allocation constraints (1) and penalize the unmet demand allocation $\sum_{t=1}^{H}\sum_{(l,l')\in L^2}\sum_{k=1}^{K}\mu_{ll't}^{k}(d_{ll'}^{k}(t)-\sum_{a=0}^{N}\sum_{k'=k}^{K}\overline{d}_{ll'}^{k,k',a}(t))$ in the objective function. Observe here that truck types are intertwined because of the down-grade policy. However, the Lagrangean relaxation of the RFS problem is separable by truck age, regarding both its objective function and the constraints. Thus, each Lagrangean dual subproblem, associated with truck age a, is smaller and easier to solve.

L-Dual Subproblem—DS[a]

$$\min\sum_{t=1}^{H}\sum_{(l,l')\in L^2}\sum_{k=1}^{K}(cv^{k,a}(t)\ \omega_{ll'}\ X_{ll'}^{k,a}(t)\quad + v^{k,a}(t)\sum_{\lambda=1}^{\lambda_{\max}}\lambda\beta_{ll'}(\lambda)\sum_{k'=1}^{k}\overline{d}_{ll'}^{k',k,a}(t)$$
$$-\ \mu_{ll't}^{k}\sum_{k'=k}^{K}\overline{d}_{ll'}^{k,k',a}(t))+\sum_{t=1}^{H}\sum_{l=1}^{L}\sum_{k=1}^{K}(m^{k,a}(t)I_l^{k,a}(t)+p^{k,a}(t)B_l^{k,a}(t)-s^{k,a}(t)S_l^{k,a}(t))$$

Note that, if we relax the flow balance constraints, the Lagrangean dual may be unbounded. Hence, we include valid inequalities to assure that the Lagrangean dual problem is bounded. The rationale of the inequalities (9) is that any allocated demand cannot be greater than original customer requests.

$$\overline{d}_{ll'}^{k,k',a}(t)\le d_{ll'}^{k}(t)\qquad \forall(l,l')\in L^2,\ t=1,2,...H,\ k,k'=1,2,...K \tag{9}$$

Constraints (2), (3), and (4) are included, as well, for each age group.

Dual variables $\{\mu_{ll't}^{k}\}$ are associated with the demand allocation constraints, also called Lagrangean multipliers. They are *unrestricted* parameters, which will be determined by subgradient optimization.

4.2 Upper Bound Derivation

The subgradient of the demand allocation constraints reflects the shortage of trucks at each time-space node, as $d_{ll'}^{k}(t)-\sum_{a=0}^{N}\sum_{k'=k}^{K}\overline{d}_{ll'}^{k,k',a}(t)$. This information may be used to generate an upper bound. In order to satisfy the unmet demand, a straightforward way is to buy new trucks to cover shortages. However, as trucks move through the time-space network, it is not necessary to buy new trucks at every time space node with a shortage. The best economic decisions can be determined by solving a minimum cost network flow problem associated with the shortages, termed UBound. The UBound problem has the same structure as a Benders subproblem except that there are no salvage variables for new trucks and there is no initial fleet. (Note that only one network problem must be solved here as truck shortages are only met with new asset purchases.)

In UBound, the demand allocation variables are fixed as the original demand allocation to the new trucks $\overline{d}_{ll'}^{k,k,0}(t)$ plus the shortage of demand allocation $d_{ll'}^{k}(t)-\sum_{a=0}^{N}\sum_{k'=k}^{K}\overline{d}_{ll'}^{k,k',a}(t)$, together referred to

$subgrad^k_{ll'}(t)$. We further define the non-negative variables: new purchases $\hat{B}^k_l(t)$ to meet demand allocation shortages $subgrad^k_{ll'}(t)$, additional empty movements $\hat{X}^k_{ll'}(t)$ and inventory $\hat{I}^k_l(t)$ which are associated with new-purchase ($\hat{B}^k_l(t)$) truck flow in the network.

UBound

$$\min \sum_{t=1}^{H} \sum_{(l,l')\in L^2} c^{k,0}(t)\omega_{ll'}\hat{X}^k_{ll'}(t) + \sum_{t=1}^{H}\sum_{l=1}^{L}(m^{k,0}(t)\hat{I}^k_l(t) + p^{k,0}(t)\hat{B}^k_l(t))$$

subject to:

$$\sum_{l'=1}^{L}\hat{X}^k_{ll'}(1) + \hat{I}^k_l(1) - \hat{B}^k_l(1) = -\sum_{l'=1}^{L} subgrad^k_{ll'}(1) \tag{10}$$
$$\forall l = 1,2,...L$$

$$\sum_{l'=1}^{L}\hat{X}^k_{ll'}(t) + \hat{I}^k_l(t) - \sum_{l'=1}^{L}\hat{X}^k_{l'l}(t-\omega_{l'l})^+ - \hat{I}^k_l(t-1) - \hat{B}^k_l(t) =$$
$$\sum_{l'=1}^{L}\sum_{\lambda=1}^{\min(t,\lambda_{\max})} \beta_{l'l}(\lambda) * subgrad^k_{ll'}(t-\lambda) - \sum_{l'=1}^{L} subgrad^k_{ll'}(t) \tag{11}$$
$$\forall l = 1,2,...L,\ t = 2,3,...H$$

where $\hat{X}^k_{ll'}(t) \geq 0,\ \hat{B}^k_l(t) \geq 0,\ \hat{I}^k_l(t) \geq 0, \forall (l,l') \in L^2,\ t = 1,2,...H.$

This again is a network flow problem, as in the Benders subproblem, but no feasibility issue is involved. That is because we only resolve shortages with new asset purchases. After solving UBound, update the decision variables with $a = 0$ as follows:

$$\bar{d}^{k,k,0}_{ll'}(t) = subgrad^k_{ll'}(t)$$
$$X^{k,0}_{ll'}(t) = X^{k,0}_{ll'}(t) + \hat{X}^k_{ll'}(t)$$
$$B^{k,0}_l(t) = B^{k,0}_l(t) + \hat{B}^k_l(t),$$
$$I^{k,0}_l(t) = I^{k,0}_l(t) + \hat{I}^k_l(t) \qquad \forall (l,l') \in L^2,\ t = 1,2,...H$$

Accordingly, the cost of a feasible solution can be re-calculated. The upper bound is updated if the feasible solution cost is lower than the original upper bound.

5 Two-Phase Algorithm

As we have investigated both Benders decomposition and Lagrangean relaxation, both have advantages and drawbacks in our procedure. Benders decomposition works efficiently before a number of optimality cuts are added, so we apply Bender decomposition for only a given number of iterations and cut if off before the master problem becomes too expensive and slow to solve. Then we restore the upper and lower bounds

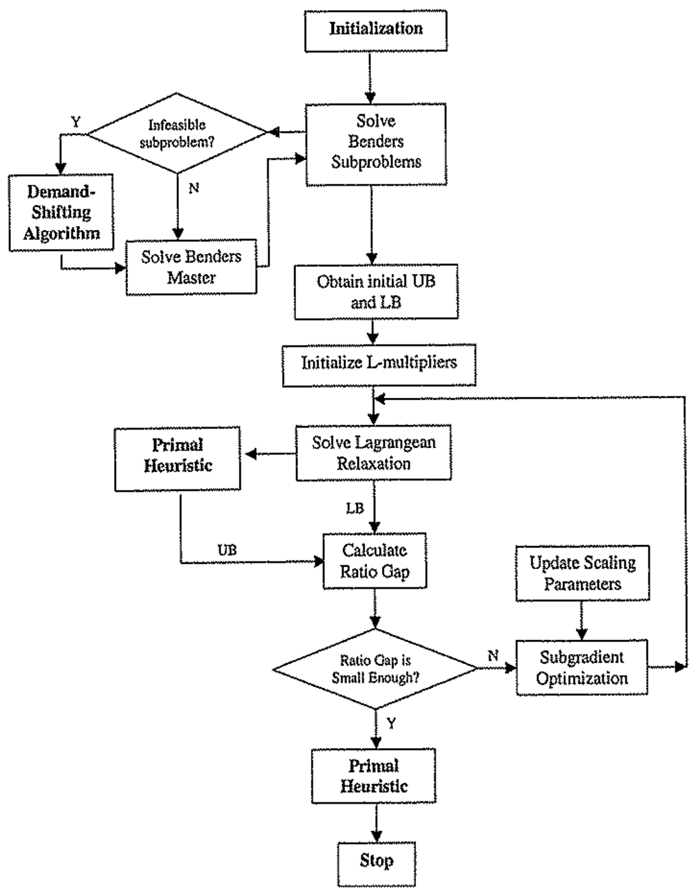


Figure 3: Two-Phase Solution Algorithm Diagram

from Benders decomposition into the Lagrangean relaxation procedure until reaching solutions within a prespecified optimality gap. We describe the algorithm as follows, with the procedure diagram in Figure 3.

Two-Phase Algorithm

1) **Phase I:** Solve the Benders decomposition procedure, as described in Section 4.1, within some preset computing time, iteration count or solution gap. When it reaches the stopping criteria, save:

a) the best upper bound from the Benders subproblems PS(k,a):

$$\text{UB} := \sum_{k=1}^{K}\sum_{a=0}^{N}\sum_{t=1}^{H}\sum_{(l,l')\in L^2} v^{k,a}(t)\sum_{\lambda=1}^{\lambda_{\max}}\lambda\beta_{ll'}(\lambda)\sum_{k'=1}^{k}\overline{d}_{ll'}^{k',k,a}(t) + \sum_{k=1}^{K}\sum_{a=0}^{N} bender\_sub[k,a]$$

b) the best lower bound from the master problem PM:

$$\text{LB} := \sum_{k=1}^{K}\sum_{a=0}^{N}\Big(\sum_{t=1}^{H}\sum_{(l,l')\in L^2} v^{k,a}(t)\sum_{\lambda=1}^{\min(t,\lambda_{\max})}\lambda\beta_{ll'}(\lambda)\sum_{k'=1}^{k}\overline{\mathbf{d}}_{ll'}^{k',k,a}(t) + \mathbf{z}^{k,a}\Big)$$

c) the dual variables $\{\eta_{ll'}^{k}(t)\}$associated with demand allocation constraints (1) in the master problem PM.

2) **Phase II:** Solve the Lagrangean relaxation procedure with subgradient optimization to provide better lower bounds:

i) Restore UB and LB from Benders procedure and set the initial Lagrangean multipliers $\{\mu_{ll't}^{k}\}$ equal to $\{\eta_{ll'}^{k}(t)\}$, where $(l,l') \in L^2$, $t = 1,2,...H$ and $k = 1,2,...K$

ii) Solve Lagrangean dual subproblems DS[a] with given multipliers $\{\mu_{ll't}^{k}\}$, one for each truck age a.

$$\text{LB} := \sum_{t=1}^{H}\sum_{(l,l')\in L^2}\sum_{k=1}^{K}\mu_{ll't}^{k} d_{ll'}^{k}(t) + \sum_{a=0}^{N} lagrangean\_sub[a]$$

$$\text{Subgradient} := \sum_{t=1}^{H}\sum_{(l,l')\in L^2}\sum_{k=1}^{K}\Big(d_{ll'}^{k}(t) - \sum_{k'=k}^{K}\overline{d}_{ll'}^{k,k',a}(t)\Big)$$

iii) Apply the Lagrangean primal heuristic to improve the upper bound, based on the solutions from the Lagrangean subproblems DS[a], which are most likely infeasible to the original RFS problem.

iv) Update the step size $\theta = \frac{\alpha*(1.05*UB-LB)}{\|subgradient\|^2}$ where $\alpha \in (0,2)$, halve it if Lagrangean dual does not increase after a certain number of iterations

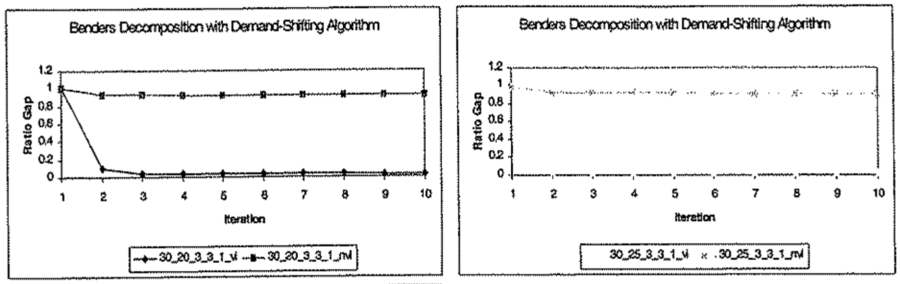
v) Update the Lagrangean multipliers: $\mu_{ll't}^{k\,(n+1)} = \mu_{ll't}^{k\,(n)} + \theta^{(n+1)} * subgradient^{(n)}$, where n is the iteration index.

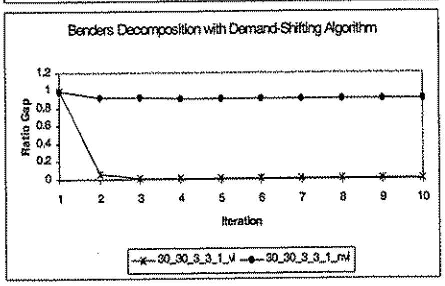
3) Terminate the procedure if iterations reach the prespecified number or the ratio gap is sufficiently small. Then apply the Lagrangean primal heuristic to convert lower-bound solutions to feasible upper bounds. If the Lagrangean heuristic gives a better upper bound than the current upper bound, we save the associated feasible solutions as the near-optimal solutions.

6 Computational Studies

6.1 Implementation Issues

The Two-Phase algorithm is implemented in Microsoft Visual Studio 6.0 C/C++ with the CPLEX 6.5 callable library [ILOG, 1999] on the same computer platform as noted in Section 3.2. Before proceeding, there are a number of practical issues worthy of a note:





Problem Instances	Gap	CPU (Sec)
30_20_3_3_1_vi	2.9%	4989.174
30_20_3_3_1_nvi	91.6%	27624.312
30_25_3_3_1_vi	2.3%	10072.103
30_25_3_3_1_nvi	91.2%	74500.861
30_30_3_3_1_vi	1.6%	5841.81
30_30_3_3_1_nvi	91.0%	226123.01

Figure 4: Solution statistics when solving RFS with (vi) and without (nvi) valid inequalities.

- The network structure of the Benders subproblems is fully exploited through the use of the network simplex method. This assures that the demand-shifting algorithm can be solved quickly and efficiently.

- The valid inequalities (8) included in the Benders master problem tremendously diminish the requirement of purchasing used trucks, therefore reduce the number of iterations of calling the demand-shifting algorithm and shortens solution time. However, once added and then stored in memory, valid inequalities become part of the master problem throughout the Benders procedure. These inequalities increase the master problem size (even before Benders cuts are added). Despite the added overhead with the additional constraints, the valid inequalities (7 and 8) are extremely effective when compared to solving RFS without the valid inequalities as shown by Figure 4.

- An initial demand allocation is required to start the Phase I Benders procedure. The direct way is to assign all demand to new trucks, as it results in a feasible, albeit costly, solution. Another method is to assign all demand to the oldest trucks and then use the demand-shifting algorithm to shift excess allocations towards the next newer-truck group. This shifting proceeds until the newest truck subproblem, where purchases meet any final shortages, providing a feasible initial allocation. As shown in Table 2, the ratio gaps of 10 instances are given for the two allocation schemes. The all-demand-to-oldest allocation scheme works quite well for relatively smaller sized problems; however, the all-demand-to-new demand allocation presents a better ratio gap for larger problems. In all cases, the Benders procedure with the all-demand-to-oldest allocation scheme is more computationally expensive, as it requires use of the demand-shifting algorithm.

	H-L	Ratio Gap		CPU (Sec)	
		Benders_new	Benders_oldest	Benders_new	Benders_oldest
1	30-10	9.5%	6.0%	142.7	219.0
2	30-15	6.0%	2.6%	402.4	595.1
3	30-20	3.1%	1.9%	1590	2167.2
4	30-25	2.5%	1.5%	3623.5	6291.5
5	40-20	2.5%	3.2%	1667.8	2673.6
6	40-25	2.5%	1.9%	3336.9	4245
7	50-20	3.3%	3.1%	3550	6556.7
8	50-25	1.3%	2.2%	4305.3	13093.4
9	60-20	2.6%	3.3%	3203.8	14898.5
10	60-25	2.0%	2.2%	8696.8	27666.9

Table 2: Ratio Gap: Different Initial Demand Allocations ($K = 3$ and $N = 3$)

6.2 Computational Results

Figure 5 illustrates change in the optimality gap per iteration when solving two large-scale instances of RFS (($H\_L\_K\_N$) are 50_20_3_3 and 60_20_3_3) using strictly Benders decomposition (top two figures) and the two-phase algorithm (bottom two figures). These problems are solved within 2.3% and 2.6% ratio gaps in 7,463 and 12,274 CPU seconds, respectively, in 10 iterations of the Phase I procedure. For these same two instances, if we terminate the Benders procedure after five iterations (ratio gaps of 3.3% and 2.63%) and take advantage of the initial bounds and the dual variables of demand allocation constraints from Phase I, we obtain solution gaps of 0.91% and 1.47% through Phase II. The attraction of the two-phase approach is that we can concentrate on achieving good solutions as the problem sized is fixed in the Lagrangian relaxation implementation phase.

Numerical results are summarized in Table 3 for various instances with $K = 3$ and $N = 3$. Different horizon lengths and number of locations are tested. Similar to the experiments discussed in Table 1, five instances are randomly generated for each sized problem. The Phase I Benders procedure is terminated when either one of the criteria is met: ratio gap falls below 3% or 10 iterations; then the Phase II Lagrangean relaxation procedure proceeds for 20 iterations. Performances including the ratio gap of the bounds and CPU time in seconds are measured by the statistics maximum, minimum and average for both Benders decomposition and Lagrangean relaxation.

As shown by the computational results, the RFS problem can be solved in a reasonable amount of time with the two-phase algorithm, given the decision-maker's acceptable solution gap and the fact that fleet-sizing is not a real-time decision. It should be clear that the number of locations has the largest impact on solution time. It is expected that some aggregation must occur to solve even larger instances.

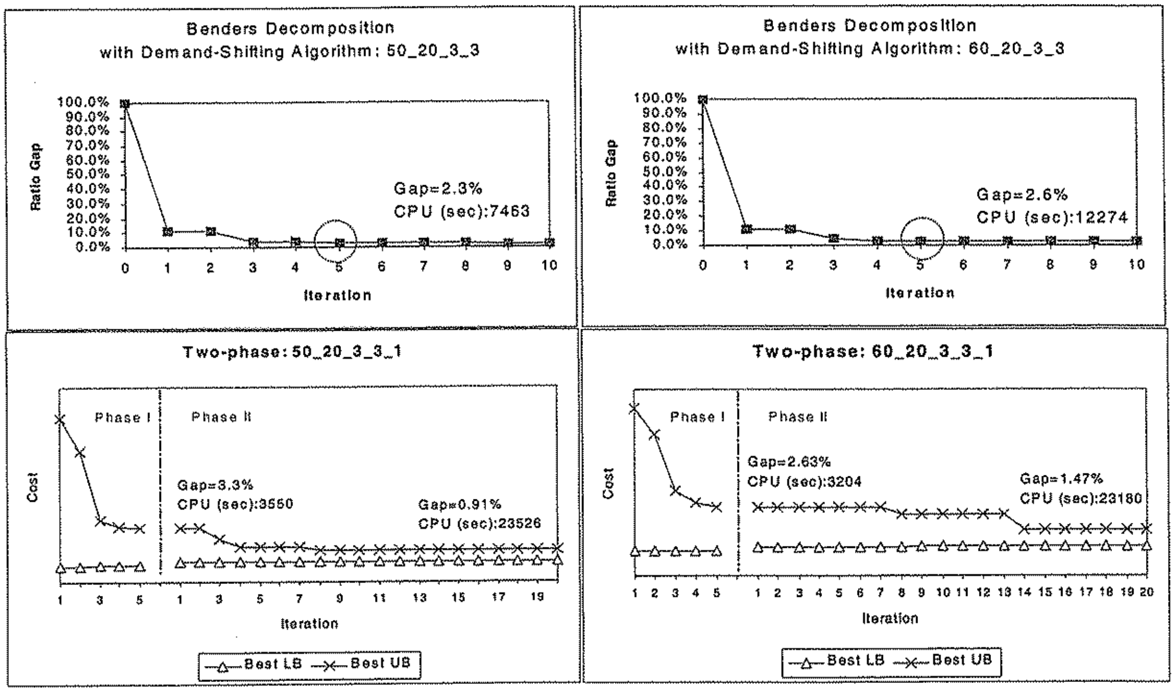


Figure 5: Benders decomposition and two-phase approach for two problem instances.

H-L	Benders Ratio Gap (%)			Lagn Ratio Gap (%)			CPU (Sec)		
	Max	Min	Avg	Max	Min	Avg	Max	Min	Avg
30-10	12.17	8.04	9.5	2.43	1.85	2.0	1044	966	1006
30-20	4.51	2.92	3.5	3.83	2.55	3.06	14006	11053	12294
30-25	2.97	2.79	2.87	2.68	1.97	2.42	27241	21578	24772
40-10	8.34	6.28	7.4	6.01	4.5	5.54	2334	1768	2007
40-20	3.52	2.91	3.29	2.81	2.09	2.54	17448	13900	15752
40-25	3.31	2.89	3.08	2.89	1.25	1.91	42306	28901	36770
50-10	8.06	6.5	7.08	7.1	5.98	6.42	2810	2365	2592
50-20	4.5	3.08	3.77	2.61	0.91	1.7	34382	22995	27634
60-10	11.97	6.92	8.86	6.9	4.5	5.9	4000	2512	3185
60-20	4.23	3.64	3.97	3.33	1.03	1.92	43081	26769	34152

Table 3: Computational Results for $K = 3$ and $N = 3$

The decisions provided by the model include demand allocation (including substitutions of larger assets for smaller assets) and empty truck repositioning as well as purchase and sales of assets over space and time. The inclusion of these multiple decisions lead to lower overall cost solutions when compared to models which focus on allocation and repositioning decisions. Additionally, the explicit capital decisions eliminate the need to re-solve the model for different fleet sizes as the fleet size itself is dynamic in the model.

The model is expected to provide support in planning fleet requirements, in terms of size and mix. However, the information supplied also provides support for operational decisions. The asset movement decisions are defined according to asset age and type. Although these decisions are not for specific trucks (by model number), they may be used in aggregate to determine assignment policies. For example, an analysis of allocation policies at different locations over time may illustrate that newer trucks are sent from Cleveland to Chicago while older trucks are used from Cleveland to Columbus. From these model based allocations, an assignment policy or policy structure for assignments may be designed.

An analysis of the solutions may also aid in making revenue management decisions. Excessive empty truck movements to, or purchases at, a given location may signal a need to raise prices for demand leaving the area whereas excessive empty movements from, or sales at, a given location may lead to a lowering of prices accordingly.

7 Conclusions and Future Research

This research addresses a rental fleet sizing problem (RFS) in the context of the truck rental industry, subject to uncertain customer travel time, and non-stationary customer demand that is dependent on geographical location, time, and the economic cycle of the industry. We integrate tactical (asset purchases and sales) and operational (empty truck movement and vehicle assignment) decisions, with the explicit incorporation of an asset age factor, to achieve lower cost solutions. The age factor is required to differentiate the costs attributed to assets as they age and deteriorate.

The unique model has led to a two-phase algorithm solution approach to deal with large-scale RFS problems. Phase I Benders decomposition allocates customer demand among available assets by asset type and age. In dealing with possible subproblem infeasibilities, a demand-shifting feasibility algorithm shifts excess demand to other sources of capacity such that a feasible solution can be found quickly without placing additional computational burden on the master problem, as feasibility cuts are not required. Valid inequalities, integrated into the master problem, help reduce subproblem infeasibilities and improve computational time. In Phase II, by taking advantage of the initial bounds and dual variables from the Benders procedure, Lagrangean relaxation further improves the overall solution quality. Computational studies on the Benders procedure, demand-shifting feasibility algorithm and the two-phase approach demonstrate the algorithmic efficiency.

In summary, the contributions of this research are:

- An integrated RFS model which incorporates operational and capital investment decisions:
  1. Integration leads to better solutions in terms of lower investment costs and inventory and less empty movement;
  2. Heterogeneous fleet defined by capacity and age allows for more accurate cost modeling, especially considering deterioration and age-based asset performance;
  3. RFS model has multiple uses for planning, as well as operational decisions.
- An effective solution approach:
  1. Two-Phase approach captures benefits of both Benders decomposition and Lagrangean relaxation;
  2. Demand-shifting feasibility algorithm speeds Benders decomposition approach and is computationally superior (solution gap and time) to traditional feasibility cuts;
  3. Primal heuristic converts Lagrangean relaxation solution to upper bound with the solution of one network flow problem.

There are a number of remaining issues that we think are very interesting problems for future research. While we assumed that demand was deterministic in order to achieve tractability, it is in fact stochastic. This assumption is clearly not valid when solving problems with long horizons. The formulation could be altered to include chance constraints or a two-stage recourse stochastic programming could be employed such that the first-stage variables make here-and-now decisions (buy and sell) without full information of demand, while second-stage variables are wait-and-see recourse actions (empty movement or inventory) to hedge against the uncertainty of demand after a demand scenario is realized. Obviously, the inclusion of stochastic demand greatly complicates the solution procedure.

Also, the current RFS model does not differentiate between assets with different levels of cumulative utilization, which is measured by truck mileage. While the inclusion of mileage will better describe the status of the assets, its incorporation will pose a greater solution challenge as the number of subproblems could increase dramatically.

We believe that this work represents an initial step in an important research area: integrating fleet management and replacement analysis decisions. As these decisions are dependent on each other, lower overall cost solutions are achievable when they are examined simultaneously. Replacement decisions coincide with capital budgeting decisions and thus require long time horizons (generally measured in quarters or years). However, fleet management decisions require short time buckets for accurate analysis (measured in days in this paper). Integrating tactical decisions with their impact on operational decisions, at least in the short term, poses an interesting line of research. Solutions may lie in rolling horizon methods or hybrid models which shift from operational decisions in the near terms to tactical decisions over the long term.

8 References

Aardal, K. and Larsson, T. (1990), "A Benders Decomposition Based Heuristic for the Hierarchical Production Planning Problem", *European Journal of Operational Research*, No.45, pp.4-14.

Bararaa, M.S. and Goode, J.J. (1979), "A Survey of Various Tactics for Generating Lagrangean Multipliers in the Context of Lagrangean Duality", *European Journal of Operational Research*, No.3, pp.322-338.

Barnhart, C., Boland, N.L., Clarke, L.W., Johnson, E.L., Nemhauser, G.L., and Shenoi, R.G. (1998), "Flight String Models for Aircraft Fleeting and Routing", *Transportation Science*, Vol.32, pp.208-.

Bazaraa, M.S., Sherali, H.D. (1980), "Benders' Partitioning Scheme Applied to a New Formulation of the Quadratic Assignment Problem", *Naval Research Logistics Quarterly* 27(1), pp.29-41.

Beaujon, G.J. and Turnquist, M.A. (1991), "A Model for Fleet Sizing and Vehicle Allocation", *Transportation Science*, Vol.25, No.1, pp.19-45.

Benders, J.F. (1962), "Partitioning Procedures for Solving Mixed-Variables Programming Problems", *Numerische Mathematik,* No.4, pp.238-252.

Chand, S., McClurg, T. and Ward, J. (2000), "A Model for parallel Machine Replacement with Capacity Expansion", *European Journal of Operational Research*, No.121, pp.519-531.

Chen, Z. (1998), "Solution Algorithm for the Parallel Replacement Problem under Economy Scale", *Naval Research Logistics,* Vol.45, pp.279-295.

Dejax, P. J. and Crainic, T.G. (1987), "A Review of Empty Flows and Fleet Management Models in Freight Transportation", *Transportation Science*, Vol.21, No.4, pp.227-247.

Fisher, M.L. (1985), "An Applications Oriented Guide to Lagrangean Relaxation", *Interface*, Vol.15, No.2, pp.10-21.

Geoffrion, A.M., Graves, G.W. (1974). "Multicommodity Distribution System Design by Benders Decomposition", *Management Science,* Vol.20, No.5, pp.822-844.

Graves, S.C. (1982), "Using Lagrangean Techniques to Solve Hierarchical Production Planning Problems", *Management Science*, Vol.28, No.3, pp.260–275.

Hartman, J. C. (2000), "The Parallel Replacement Problem with Demand and Capital Budgeting Constraints", *Naval Research Logistics*, No.47, pp.40-56.

Hartman, J.C. (1999), "A General Procedure for Incorporating Asset Utilization Decisions into Replacement Analysis", *The Engineering Economist*, Vol.44, No.3, pp.217-238.

Hartman, J.C. and Lohmann, J. (1997), "Multiple Options in Parallel Replacement Analysis: Buy, Lease or Rebuild", *The Engineering Economist*, Vol.42, No.3, pp.223-248.

Holmberg, K. (1994), "On Using Approximations of the Benders Master Problem", *European Journal of Operational Research*, No.77, pp.111-125.

Holmberg, K and Hellstrand, J. (1998), "Solving the Uncapacitated Network Design Problem by a Lagrangean Heuristic and Branch-and-Bound", *Operations Research,* Vol.46, No.2, pp.247–259.

ILOG, 1999. ILOG CPLEX 6.5, Reference Manual.

Jones, P., Zydiak, J. and Hopp, W. (1991), "Parallel Machine Replacement", *Naval Research Logistics,* Vol.38, pp.351-365.

Jordan, W.C. and Turnquist, M.A. (1983), "A Stochastic, Dynamic Network Model for Railroad Car Distribution", *Transportation Science*, Vol.17, No.2, pp.123-145.

Karabakal, N., Lohmann, J. and Bean, J. (1994), "Parallel Replacement under Capital Rationing Constraints", *Management Science*, Vol.40, No.3, pp.305-319.

Kirby, D. (1959), "Is Your Fleet the Right Size?", *Operational Research Quarterly*, Vol.10 pp.252.

Klincewicz, J.G., Luss, H. and Pilcher, M.G. (1990), "Fleet Size Planning When Outside Carrier Services Are Available", *Transportation Science*, Vol.24 No.3, pp.169-182.

Magnanti, T.L. and Simpson, R.W. (1978), "Transportation Network Analysis and Decomposition Methods", U.S. Department of Transportation, DOT-TSC-RSPD-78-6.

Magnanti, T.L. (1981), "Combinatorial Optimization and Vehicle Fleet Planning: Perspectives and Prospects", *Networks*, Vol.11, pp.179-213

Magnanti, T.L. and Wong, R.T. (1981), "Accelerating Benders Decomposition: algorithmic Enhancement and Model Selection Criteria", *Operations Research,* Vol. 29, No.3, pp.464-484.

Mendiratta, V.B. and Turnquist, M.A. (1982), "Model for Management of Empty Freight Cars", *Transportation Research Record* 838, pp.50-55.

Minoux, M. (1984), "Subgradient Optimization and Benders Decomposition for Large Scale Programming", *Mathematical Programming*, Cottle, R.W., Kelmanson, M.L. and Korte, B. (Editors).

Mole, R.H. (1975), "Dynamic Optimization of Vehicle Fleet Size", *Operational Research Quarterly*, Vol.26, No.1(I), pp.25-34.

Powell, W.B., Carvalho, T.A., Godfrey, G.A. and Simao, H.P. (1995), "Dynamic Fleet Management as A Logistics Queueing Network", *Annals of Operations Research,* Vol.61, pp.165-188.

Powell, W.B., Carvalho, T.A. (1998), "Dynamic Control of Logistics Queueing Networks for Large-Scale Fleet Management", *Transportation Science*, Vol.32, No.2, pp.90-109.

Reeves, C.R., *Modern Heuristic Techniques for Combinatorial Optimization,* Wiley and Sons, 1993.

Sherali, H. and Tuncbilek, C. (1997), "Static and Dynamic Time-Space Strategic Models and Algorithms for Multilevel Rail-car Fleet Management", *Management Science*, Vol.43, No.2, pp.235-250.

Sherali, H., Al-Yakoob, S.M. and Hassan, M.M. (1999), "Fleet Management Models and Algorithms for an Oil-Tanker Routing and Scheduling Problem", IIE Transactions, No.31, pp.395-406.

Van Slyke, R., Wets, R. J-B, (1969). "L-Shaped Linear Programs with Application to Optimal Control and Stochastic Programming", *SIAM Journal on Applied Mathematics* 17, pp.638-663.

Vemuganti, R., Oblak, M. and Aggarwal, A. (1989), "Network Models for Fleet Management", *Decision Sciences,* 20, pp.182-197.

Wu, P., Hartman, J.C. and Wilson, G.R. (2000), "A Demand-Shifting Feasibility Algorithm for Benders Decomposition", Department of Industrial & Manufacturing Systems Engineering, Lehigh University, Technical Paper No. 00T-008, 2000. (accepted to appear in *European Journal of Operational Research*).

Wyatt, J.K. (1969), "Optimal Fleet Size", *Operational Research Quarterly*, Vol.12, pp.186-187.

9 Acknowledgements

This research was supported in part by the National Science Foundation grant No. DMI-9713690.

10 Appendix

The following data was generated for the computational studies. The function RAND(a,b) describes a the generation of a random number where a and b are the lower and upper bounds.

- Network: Given the number of locations, the network was generated by randomly generating empty movement travel times (RAND(1,3))between locations (assumed symmetric). Note that customer travel time may be longer, as described below.
- Initial fleets: Generated uniformly by age group n and location l as follows. We assume a larger mix of larger assets to allow for the downgrade policy to have an effect.
  - $I_l^{1,a}(0) = 1 + \text{RAND}(10,20)$
  - $I_l^{2,a}(0) = 10 + \text{RAND}(10,20)$
  - $I_l^{3,a}(0) = 20 + \text{RAND}(10,20)$
- Travel times: Local rentals (retuned to the same location) were assumed to be uniformly distributed as RAND(1,3). For one-way rentals, the distribution on the number of days to make the trip was determined from the network. Given the empty movement travel time of t, a distribution of RAND(t,t+2) was used for customer travel time as it was assumed that they would go no faster than the professional drivers.
- Cost functions:

Inventory or holding cost: $m^{k,a}(t) = 0.3*(20.0 + (k+1)*10 + a*20 + \text{RAND}(1,6))$

Empty movement cost:

$c^{k,a}(t) = (35.0+(k+1)*10+a*5+\text{RAND}(1,6))*\text{travel days}$

Loaded movement cost:: $v^{k,a}(t) = 0.5*(c^{k,a}(t))$

Purchase price: $p^{k,a}(t) = 20000.0 + (k+1)*3500 - a*2000 + (t+1)*10 + \text{RAND}(10,100)$

Salvage value: $s^{k,a}(t) = 0.15*(18000.0 + (k+1)*3500 - a*2000 - (t+1)*50 + \text{RAND}(10,80))$

- Demand: The demand was generated to simulate what occurs in the truck rental business: strong demand is expected on weekends (Fridays and Saturdays) and at the end of each month. For a typical weekday: RAND(2,6), RAND(3,10) and RAND(3,12) were used to generate imbalanced demand across the network. For Fridays and Saturdays, these were strengthed with 6 + RAND(1,3), 10 + RAND(1,3) and 12 + RAND(1,4) (distributed across the network). The last days of the month were further strengthened with 10 + RAND(1,2), 14 + RAND(1,3) and 16 + RAND(1,4).