Composite Variable Formulations for Express Shipment Service Network Design

Andrew P. Armacost • Cynthia Barnhart • Keith A. Ware
Department of Management, United States Air Force Academy, Colorado Springs, Colorado 80840
Department of Civil and Environmental Engineering, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139
Operations Research Group, United Parcel Service, Louisville, Kentucky 40223

In this paper we describe a new approach to solving the express shipment service network design problem. Conventional polyhedral methods for network design and network loading problems do not consistently solve instances of the planning problem we consider. Under a restricted version of the problem, we transform conventional formulations to a new formulation using what we term *composite variables*. By removing flow decisions as explicit decisions, this extended formulation is cast purely in terms of the design elements. We establish that its linear programming relaxation gives stronger lower bounds than conventional approaches. We apply this composite variable formulation approach to the UPS Next Day Air delivery network and demonstrate potential annual cost savings in the hundreds of millions of dollars.

The U.S. package delivery industry plays an exceedingly important role in our economy by providing consistent and reliable delivery of a wide range of goods. In 1999 it generated an estimated $52 billion in revenue (see *Standard and Poor's* 2000). The dominant players in this industry are the United Parcel Service (UPS), which is the world's largest package delivery company; Federal Express; and the U.S. Postal Service. In this paper, we focus on express shipment networks, specifically the planning and design of overnight air networks, which are central to providing time-critical expedited delivery service. Improving the design yields significant savings, in terms of both operating cost and the cost of owning aircraft.

When formulated with conventional optimization models, this is a network loading problem with side constraints and is difficult to solve. The linear programming relaxations for these formulations give poor lower bounds, as they tend to select fractional aircraft routes. Advances in the theory of solving network design and network loading problems have focused on strengthening the linear programming relaxation and improving the tractability of formulations. Unfortunately, the massive scale of our network design problem and the inherent difficulty of constraints specific to Express Shipment Service Network Design render these advances ineffective.

For these reasons, we introduce a new approach for solving the express shipment service network design problem. The foundation of this approach is the use of *composite variables*. At their core, the composites are combinations of aircraft routes that implicitly capture package flows. Package flow variables are no longer represented as separate decision variables. The overall result is that the composites prevent many fractional solutions from ever appearing in the LP relaxation. Thus, a composite-based network design formulation is better approximated by its LP relaxation and, therefore, easier to solve than a traditional network design formulation.

Our overall objective is to develop and utilize a practical solution methodology for network design.

0041-1655/02/3601/0001$05.00
1526-5447 electronic ISSN

We make the following specific contributions:

• Develop a *robust solution methodology* for solving the Express Shipment Service Network Design (*ESSND*) problem. Current polyhedral methods for network design and network loading problems are not effective on *ESSND* instances of realistic size. The composite variable formulation provides stronger bounds along with the flexibility to handle operational constraints that make conventional formulations intractable. Computations with this model are fast, making it a useful tool to support network planners.

• Establish the *theoretical foundation* for this method. We show the equivalence of the composite variable formulation with conventional models and prove that using composite variables achieves stronger lower bounds on the optimal integer solution.

• Demonstrate the *practical significance* of the composite variable approach on UPS's Next Day Air network. This instance could not be solved without the new formulation. We demonstrate the potential to save hundreds of millions of dollars in the annual cost of owning and operating aircraft.

The structure of this paper is designed to highlight both the theory and application of composite variable formulations. We begin with a description of express shipment operations and a discussion of conventional network design formulations recently applied to this problem. We then transform this conventional formulation to the composite variable formulation, using an intermediate model to establish the equivalence of the (mixed) integer formulations and to prove that the new formulation's LP relaxation provides a better lower bound on the optimal integer solution. We conclude with a demonstration of this modeling approach on realistic problem instances for the United Parcel Service's Next Day Air network.

1. Background

Express shipment carriers operate vast systems of aircraft, trucks, sorting facilities, equipment, and personnel to move packages between customer locations. The problem we consider involves only overnight air operations. The carrier must determine which routes to fly, which fleet types to assign to those routes, and how to assign packages to those aircraft, all in response to demand projections and operational restrictions. In this section we describe the overnight operations of an express shipment carrier, describe and formulate the planning problem, and highlight previous work in this and related areas.

1.1. Overview of Next Day Air Operations

The Next Day Air (NDA) network consists of *gateway* locations that serve as points at which packages enter (or exit) the air network; *hub* locations, where packages are sorted; and aircraft of multiple fleet types. Consider the simple network shown in Figure 1. Packages arrive from *ground centers* (small squares) to gateways, either on trucks or on small aircraft. Packages enter the main air system through a gateway (e.g., node 1), are loaded onto an aircraft, and are transported to a hub (e.g., node H) no later than a specified time. Aircraft fly to the hub either directly or, as in Figure 1, via a single intermediate gateway (e.g., node 2).

Upon arrival at the hub, packages are unloaded from the aircraft, sorted, and loaded onto aircraft for delivery to their destination gateway. During the sorting process, the inbound planes remain at the hub until they are loaded and ready to start their delivery routes. The aircraft then deliver packages to the gateway locations, which sort the packages and send

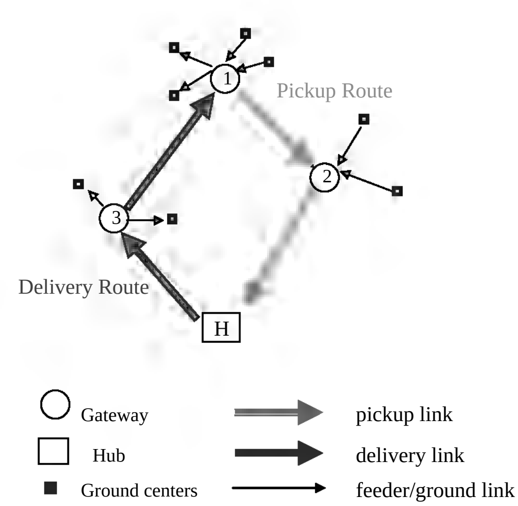

Figure 1 Example Next Day Air (NDA) Routes

them to ground centers via truck or small aircraft. From there, the packages are delivered to customers.

The aircraft inventory consists of multiple aircraft types. Each aircraft type has operating characteristics that determine the routes it can fly. These include maximum flying range, effective speed, restrictions on the locations at which it can land, and cargo capacity. Landing restrictions include factors such as runway length, physical space on the ramp, and noise restrictions at airports. There are limits on the number of legs a plane can fly on a pickup or delivery route—current UPS operations limit the number of legs to two. In addition to the large jet aircraft, a fleet of small aircraft provides a flexible source of excess capacity used on an as-needed basis.

In this paper, we distinguish between routes, aircraft routes, and paths. A *route* is an ordering of locations (i.e., either gateways or hubs) in the physical network. All pickup routes end at a hub and all delivery routes begin at a hub. An *aircraft route* is the combination of a specific aircraft (fleet) type and a route. Not all fleet types can fly every route. Finally, a *path* is used to denote package flows through the network.

The commodities that move through the network are specified by gateway-to-hub demands on the pickup side and by hub-to-gateway demands on the delivery side. We only consider demands that are to move through the air network; ground movement of packages is external to the problem we consider. Demand volume is specified in terms of packages or in terms of containers, each of which holds hundreds of packages.

To ensure appropriate customer service levels, time boundaries are set for pickup and delivery. For each gateway location, the carrier assigns *level-of-service* (LOS) requirements in the form of an *earliest pickup time* (EPT), which specifies the earliest time an aircraft can depart from that location, and a *latest delivery time* (LDT), which specifies the latest time at which packages can be delivered to the gateway. Timing requirements at hubs are specified by *sort start times* and *sort end times*. Sort start represents the latest time at which planes can arrive at the hub on a pickup route and have their packages sorted. Sort end represents the latest time at which packages may be loaded onto outbound aircraft and, therefore, the earliest time at which planes may depart on delivery routes.

Each aircraft incurs three types of cost. First, *variable operating cost* is based on block hours flown (i.e., flying time plus taxi time) and includes components such as fuel cost. Second, a fixed *cycle cost* is incurred on each leg flown. Third, *ownership cost* is the daily cost of owning the aircraft and is a fixed cost incurred if an aircraft is used. The cost of handling and moving packages is insignificant relative to the cost of owning and operating the aircraft, and is assumed to be zero. The cost components we include in the model depend upon the overall objective. When minimizing total operating cost, we use the variable operating cost and the cycle cost; when minimizing ownership cost, we include only the fixed ownership cost; and when minimizing total cost, we include all three cost components.

1.2. Express Shipment Service Network Design Formulation

We use the following sets to represent components of the system: F is the set of fleet types, H is the set of hubs, G is the set of gateways, R^f is the set of possible routes that can be flown by fleet type $f \in F$, and $\mathcal{K}$ is the set of commodities to be flown. The planning problem faced by express shipment carriers is to design the minimum cost set of routes, aircraft assignments to those routes, and package flows while considering the following operational restrictions:

- limit the number of utilized aircraft (of each fleet type) to the number available, n_f, $f \in F$;
- limit the number of aircraft landing at each hub to the hub's landing capacity, a_h, $h \in H$;
- satisfy level-of-service (LOS) requirements for pickup and delivery;
- arrive at and depart from hub locations according to the sort start and end times;
- ensure that the number of each aircraft type ending at a location equals the number that start there; and
- satisfy the characteristics of each aircraft type, including range, capacity, and flying speed.

The set of all possible aircraft movements defines a time-space network, denoted by $G = (N, A)$. Define the integer decision variable y_r^f to be the number of times we fly route $r \in R^f$ with fleet type $f \in F$. The cost of this aircraft route is denoted by d_r^f and its capacity is denoted by u_r^f. We also define the decision variable x_{ij}^k to be the amount of commodity $k \in \mathscr{K}$ flown on flight arc $(i, j) \in A$. We map each aircraft route (f, r) to the arcs in A with the indicator δ_{ij}^{fr}, which equals 1 when flight arc (i, j) is contained in aircraft route (f, r) and 0 otherwise. We map the arc corresponding the sort at hub h with the indicator δ_{ij}^h, and we map each route to its hub with the indicator δ_h^r. Associated with the start and end of each route is the indicator β_i^r, which equals 1 when i is the route's origin, -1 when i is the route's destination, and 0 otherwise. Each commodity $k \in \mathscr{K}$ has a total volume, denoted by b^k, to be moved from its origin, $O(k)$, to its destination, $D(k)$.

The conventional *ESSND* formulation, introduced in Kim et al. (1999) is given by:

$$\min \sum_{f \in F} \sum_{r \in R^f} d_r^f y_r^f$$

subject to

$$\sum_{k \in K} x_{ij}^k \leq \sum_{f \in F} \sum_{r \in R^f} \delta_{ij}^{fr} u_r^f y_r^f, \quad (i, j) \in A, \tag{1}$$

$$\sum_{j:(i,j) \in A} x_{ij}^k - \sum_{j:(j,i) \in A} x_{ji}^k = \begin{cases} b^k & \text{if } i = O(k), \\ -b^k & \text{if } i = D(k), \; i \in N, k \in \mathscr{K}, \\ 0 & \text{otherwise}, \end{cases} \tag{2}$$

$$\sum_{r \in R^f} \beta_i^r y_r^f = 0, \quad i \in N, f \in F, \tag{3}$$

$$\sum_{r \in R^f} y_r^f \leq n_f, \quad f \in F, \tag{4}$$

$$\sum_{f \in F} \sum_{r \in R^f} \delta_h^r y_r^f \leq a_h, \quad h \in H, \tag{5}$$

$$x_{ij}^k \geq 0, \quad (i, j) \in A, k \in \mathscr{K}, \tag{6}$$

$$y_r^f \in \mathbb{Z}_+, \quad r \in R^f, f \in F. \tag{7}$$

Forcing constraints (1) restrict the amount of flow on any given arc to the capacity assigned to the arc. *Flow balance constraints* (2) enforce conservation of flow for each origin-destination commodity. Constraints (3) are the *aircraft balance* constraints, which force the number of aircraft of a given fleet type departing a location on pickup routes to be offset by the same number of aircraft of that fleet type landing at that location on delivery routes. Additional constraints enforce the *number of available aircraft* of each fleet type (4), and *landing capacities* at the hubs (5). One additional constraint is the sorting capacity of each hub. It is assumed that the given gateway-to-hub commodity assignments do not exceed the hubs' sorting capacities.

1.3. Literature Review

Conventional network design formulations can be applied to instances of limited size. In such formulations, we include two types of decision variables: those for the aircraft routing (i.e., design) decisions and those for the package flow decisions. These types of problems have been well studied in the network design and network loading literature. Magnanti and Wong (1984), Minoux (1989), Kim et al. (1999), Gendron et al. (1999), and Crainic (2000) provide surveys of network design models and applications. Magnanti and Wong (1984) provide a unified framework for describing network design problems and deriving network design algorithms.

Recent research on solving network design problems has focused primarily on strengthening the LP relaxation and on the development of approximation algorithms. Characterizing polyhedra and deriving valid inequalities for network design can be traced to the development of valid inequalities for 0-1 programming (see Wolsey 1975 and Crowder et al. 1983) and the development of valid inequalities for fixed-charge network problems (see Van Roy and Wolsey 1985 and Padberg et al. 1985). Applications to network loading problems include Magnanti et al. (1993), Magnanti and Mirchandani (1993), Pochet and Wolsey (1995), Magnanti et al. (1995), Bienstock and Günlük (1996), Bienstock et al. (1996), and Chopra et al. (1998). Development of algorithms that embed polyhedral elements include Bienstock and Günlük (1995), Barahona (1995), Günlük (1999), and Stallaert (2000). Approximation algorithms for network design include those developed in Goemans and Bertsimas (1993), Agrawal et al. (1995), Goemans and Williamson (1995), Williamson et al. (1995), Jain (1998), Gabow et al. (1998), Hochbaum and Naor (1996), Berstimas and Teo (1998), and Karger (1999).

For express shipment service network design, Kuby and Gray (1993) develop network design models for the case of Federal Express. Barnhart and Schneur (1996) address the problem of designing a single-hub overnight delivery network using column generation techniques to obtain near-optimal solutions. Kim et al. (1999) apply branch-and-price-and-cut methods to the multihub express shipment problem using a heuristic solution strategy. Grünert and Sebastian (2000) identify planning tasks faced by postal and express shipment companies and define corresponding optimization models, and Büdenbender et al. (2000) develop a hybrid tabu search/branch-and-bound solution methodology for direct flight postal delivery.

2. Composite Variables and Reformulations

In conventional express shipment service network design (*ESSND*) formulations, both aircraft routes and package flows are modeled explicitly. Initial computational tests using models of the conventional form could not routinely solve realistic instances of the *ESSND* planning problem. The formulation's linear programming relaxation gives poor bounds on the optimal integer solution. Two primary factors contribute to these poor bounds. First, the forcing constraints (1) induce fractionality in the aircraft decision variables. This stems from the fact that rarely does the package volume assigned to a plane utilize all of the plane's capacity. As such, the LP relaxation chooses fractional planes rather than incur the cost of unused capacity. Second, the aircraft balance constraints (3) amplify this problem. A fractional plane that might otherwise be isolated to one route is connected to the rest of the network via these constraints, which propagate fractionality throughout the network.

To overcome this, we introduce a new formulation approach that we show strengthens conventional network design formulations. The formulation relies on two key ideas. First, we capture multiple aircraft routes with a single variable. Second, we build package flows *implicitly* into the new variables. The resulting *composite variable* represents a combination of aircraft routes such that there exists a feasible flow for all packages between some set of origins and destinations.

We present the transformation using an intermediate formulation that (1) removes the explicit representation of package flows, (2) provides an intuitive method for understanding composite variable formulations, and (3) allows us to establish the strength of the composite variable formulation relative to the original formulation.

2.1. A Restricted Version of *ESSND*

We initially assume that *no ramp transfers* are permitted. This means that packages, once they are loaded on a plane, stay on the plane until they reach the hub for sorting (if a pickup route) or until they reach their final destination (if a delivery route). No package exchanges between planes can occur at intermediate locations.

Under this assumption, we redefine the *ESSND* package flow variables accordingly: We assign demands to *routes* (versus arcs) and ensure that the fleet types assigned to fly the routes have sufficient capacity to carry the demands. We represent each *commodity* as a gateway-hub pair $(g, h) \in \mathcal{K}$. The commodity set, $\mathcal{K}$, is split into two disjoint sets, $\mathcal{K}_P$ and $\mathcal{K}_D$, corresponding to the pickup side and delivery side, respectively. We define x_r^{gh} to be the amount of (g, h) demand assigned to route r. The set $R(g, h)$ includes all routes that connect gateway g with hub h. Subscripts divide $R(g, h)$ into $R_P(g, h)$ and $R_D(g, h)$, the set of pickup and delivery routes, respectively. The parameter b_P^{gh} is the demand to be moved from g to h on the pickup side and b_D^{gh} is the demand to be moved from h to g on the delivery side.

With package flows now assigned based on path (route) flows, *ESSND* is rewritten as the following (*ESSND-R*):

$$\min \mathbf{d}'\mathbf{y} \tag{8}$$

subject to

$$\sum_{(g,h)\in\mathcal{K}} x_r^{gh} - \sum_{f\in F} u_r^f y_r^f \le 0, \quad r \in R, \tag{9}$$

$$\sum_{r\in R_P(g,h)} x_r^{gh} = b_P^{gh}, \quad (g,h) \in \mathcal{K}_P, \tag{10}$$

$$\sum_{r\in R_D(g,h)} x_r^{gh} = b_D^{gh}, \quad (g,h) \in \mathcal{K}_D, \tag{11}$$

$$\mathbf{B}\mathbf{y} \le \beta, \tag{12}$$

$$y_r^f \in \mathbb{Z}_+, r \in R^f, f \in F, \tag{13}$$

$$x_r^{gh} \ge 0, r \in R, (g,h) \in \mathcal{K}. \tag{14}$$

Constraints (9) are the *forcing* constraints. There is one forcing constraint for each *route*, compared to one forcing constraint for each flight arc in the time-space network of the *ESSND* formulation. Constraints (10) ensure that demand for each pickup commodity is fully assigned to pickup routes. Similarly, constraints (11) ensure all delivery demands are fully assigned to delivery routes. Constraints (12) are a concise representation of the *balance, landing*, and *plane count* constraints described in §1. Constraints (13) and (14) represent *aircraft route integrality* and *nonnegativity*, respectively.

ESSND-R exploits the assumption that transferring packages between aircraft is allowed only at the hubs. The new formulation is equivalent to *ESSND*, as are their linear programming relaxations. This new formulation provides the starting point from which we create a model consisting only of aircraft route variables.

2.2. The Extreme Route Formulation

The "no-ramp-transfer" assumption implies that an aircraft route will carry only the commodities corresponding to the gateways it visits and the hub at which it terminates (for pickup routes) or originates (for delivery routes). We now focus on characterizing the *available capacity* an aircraft route may use to carry these demands. For double-leg routes, there might be an infinite number of ways to divide the available capacity between the two gateway-hub commodity demands. For single-leg routes, the entire capacity of the aircraft route is available to move the demand between the single gateway and hub. We will show that we can characterize the use of a plane's capacity as a convex combination of the extreme uses of this capacity.

For each aircraft route, we associate a set of *extreme routes*, each of which specifies an extreme allocation of the aircraft's available capacity on that route. The *actual* flow might be less than the *available* capacity. For double-leg routes, we can give preference to the route's first location, loading the aircraft with as much of the first location's demand as possible and using the excess capacity (if any) for the second location's demand. At the other extreme, we may give capacity preference to the second location. Once again, we are specifying an extreme allocation of the *capacity* and are not specifying the actual flows.

To formalize the definition, consider an aircraft route denoted by (f, r) visiting gateways i and j and hub h (we'll work strictly with the pickup routes, but the same results apply to the delivery routes). The *available capacities* for the first extreme route corresponding to (f, r) are $\hat{u}_{ih}^1$ for commodity (i, h) and $\hat{u}_{jh}^1$ for commodity (j, h) and are defined as follows:

$$\hat{u}_{ih}^1 = \min\{b_P^{ih}, u_r^f\},$$
$$\hat{u}_{jh}^1 = \min\{u_r^f - \hat{u}_{ih}^1, b_P^{jh}\}.$$

We characterize the second extreme route using the available capacities $\hat{u}_{jh}^2$ for commodity (j, h) and $\hat{u}_{ih}^2$ for commodity (i, h), which are defined as follows:

$$\hat{u}_{jh}^2 = \min\{b_P^{jh}, u_r^f\},$$
$$\hat{u}_{ih}^2 = \min\{u_r^f - \hat{u}_{jh}^2, b_P^{ih}\}.$$

It is possible that both extreme routes are characterized by the same available capacities. This occurs when the total capacity of the aircraft route exceeds the *total* demand to be moved, that is $u_r^f \ge b_P^{ih} + b_P^{jh}$. Then $\hat{u}_{ih}^1 = \hat{u}_{ih}^2 = b_P^{ih}$ and $\hat{u}_{jh}^1 = \hat{u}_{jh}^2 = b_P^{jh}$.

For single-leg routes, available capacity is allocated to a single gateway-hub commodity, (i, h), and we have only a single extreme route. The available capacity for this extreme route is defined as:

$$\hat{u}_{ih}^1 = \min\{b_P^{ih}, u_r^f\}. \tag{15}$$

Example 1. Consider the aircraft route shown in Figure 2. The plane has a 10,000 package capacity. The first gateway-hub commodity has a volume of 7,000 packages and the second gateway-hub

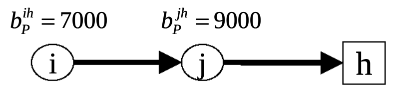

Figure 2 Two-Leg Aircraft Route with 10,000-Package Capacity (Example 1)

commodity has a volume of 9,000 packages. The first extreme route picks up as much of the (i, h) demand as possible and uses the remaining capacity for the (j, h) demand. The available capacities associated with extreme route 1 are $\hat{u}^1_{ih} = 7{,}000$ and $\hat{u}^1_{jh} = 3{,}000$. The second extreme route gives preference to the (j, h) demand. The available capacities associated with extreme route 2 are $\hat{u}^2_{ih} = 1{,}000$ and $\hat{u}^2_{jh} = 9{,}000$.

Using the same set of extreme routes, any feasible package flow can be transported using the capacity specified by a convex combination of the extreme routes. For instance, if the aircraft picks up a partial load, say 5,000 packages from each location, we weight the first extreme point by $\lambda_1 = \frac{2}{3}$ and the second extreme point by $\lambda_2 = \frac{1}{3}$ to yield an available capacity of 5,000 packages for each gateway-hub commodity.

Finally, consider the case in which, using the same extreme routes, the flow on the aircraft route does not use the entire capacity of the aircraft. Say the aircraft carries 2,000 packages from each location. Then there is a range of multipliers that provide adequate capacity: $\lambda_1 \in [\frac{1}{6}, 1]$ and $\lambda_2 = 1 - \lambda_1$.

In the extreme route formulation we are concerned only with extreme points corresponding to maximum flows because we use extreme routes to specify *available capacity*, not actual flows. The following result relates extreme routes to actual flows on a given aircraft route and is central to the development of the extreme route formulation.

Proposition 1. *A package flow is feasible on an aircraft route (f, r) with capacity u^f_r **if and only if** it is feasible on some convex combination of the extreme routes of (f, r).*

Proof. (All arguments are presented in terms of *pickup* routes and the results apply similarly to delivery routes.) We consider three cases. The first is for single-leg routes, the second is for a double-leg route with capacity that exceeds its gateway-hub demands, and the third is for a double-leg route with gateway-hub demands that exceed its capacity. For each case, we show that for a given flow on an aircraft route there is a convex combination of extreme routes for which that flow is also feasible, and vice versa.

Consider the *first case*, in which (f, r) is a single-leg route from i to h with flow $\hat{x}^{fr}_{ih}$. We have a single extreme route with $\hat{u}^1_{ih} = \min\{b^{ih}_P, u^f_r\}$ and the total flow on (f, r) cannot exceed demand b^{ih}_P. Any flow less than u^f_r cannot be greater than $\hat{u}^1_{ih}$ and vice versa.

Consider the *second case*, in which (f, r) is a double-leg route from i to j to h with $b^{ih}_P + b^{jh}_P \le u^f_r$. There is a single extreme route with capacities $\hat{u}^1_{ih} = b^{ih}_P$ and $\hat{u}^1_{jh} = b^{jh}_P$. Given any feasible flow $(\hat{x}^{fr}_{ih}, \hat{x}^{fr}_{jh})$ on this double-leg route, we have $\hat{x}^{fr}_{ih} \le b^{ih}_P = \hat{u}^1_{ih}$ and $\hat{x}^{fr}_{jh} \le b^{jh}_P = \hat{u}^1_{jh}$ and the flow is feasible with respect to the extreme route. Given a flow $(\bar{x}^{fr}_{ih}, \bar{x}^{fr}_{jh})$ that is feasible with respect to the extreme route, we have $\bar{x}^{fr}_{ih} \le \hat{u}^1_{ih}$ and $\bar{x}^{fr}_{jh} \le \hat{u}^1_{jh}$. Summing, we get $\bar{x}^{fr}_{ih} + \bar{x}^{fr}_{jh} \le \hat{u}^1_{ih} + \hat{u}^1_{jh} \le u^f_r$ and the flow is feasible with respect to the aircraft route, (f, r).

The *third case* is when (f, r) is a double-leg route from i to j to h when $b^{ih}_P + b^{jh}_P > u^f_r$. Given a feasible flow on (f, r), we have $\hat{x}^{fr}_{ih} + \hat{x}^{fr}_{jh} \le u^f_r$. The flow of commodity (i, h) satisfies both $\hat{x}^{fr}_{ih} \le u^f_r$ and $\hat{x}^{fr}_{ih} \le b^{ih}$ and it follows that $\hat{x}^{fr}_{ih} \le \hat{u}^1_{ih}$. If $\hat{u}^2_{ih} < \hat{x}^{fr}_{ih} \le \hat{u}^1_{ih}$, we can find λ_1 and λ_2 such that

$$\begin{aligned} \lambda_1 + \lambda_2 &= 1, \\ \hat{u}^1_{ih}\lambda_1 + \hat{u}^2_{ih}\lambda_2 &= \hat{x}^{fr}_{ih}. \end{aligned} \tag{16}$$

These multipliers also provide sufficient capacity to cover the demand from j to h:

$$\begin{aligned} \hat{x}^{fr}_{jh} &\le u^f_r - \hat{x}^{fr}_{ih} \\ &= u^f_r - (\hat{u}^1_{ih}\lambda_1 + \hat{u}^2_{ih}\lambda_2) \\ &= \lambda_1(u^f_r - \hat{u}^1_{ih}) + \lambda_2(u^f_r - \hat{u}^2_{ih}) \\ &= \lambda_1\hat{u}^1_{jh} + \lambda_2\hat{u}^2_{jh}. \end{aligned}$$

When $\hat{x}^{fr}_{ih} \le \hat{u}^2_{ih}$, the second extreme route covers both demands and we let $\lambda_1 = 0$ and $\lambda_2 = 1$.

Conversely (for the third case), assume we have a flow, $(\bar{x}^{fr}_{ih}, \bar{x}^{fr}_{jh})$, that is feasible with respect to a convex combination of its extreme routes. That is, $\bar{x}^{fr}_{ih} \le \hat{u}^1_{ih}\lambda_1 + \hat{u}^2_{ih}\lambda_2$ and $\bar{x}^{fr}_{jh} \le \hat{u}^1_{jh}\lambda_1 + \hat{u}^2_{jh}\lambda_2$. Summing the two inequalities, we obtain

$$\bar{x}_{ih}^{fr} + \bar{x}_{jh}^{fr} \le (\hat{u}_{ih}^1 + \hat{u}_{jh}^1)\lambda_1 + (\hat{u}_{ih}^2 + \hat{u}_{jh}^2)\lambda_2 = u_r^f(\lambda_1 + \lambda_2) = u_r^f,$$

which yields the desired result. □

2.2.1. Formulation. We introduce additional notation to create a formulation based on extreme routes. We define E to be the set of extreme routes as constructed above and we let w_e, $e \in E$, be the decision variables corresponding to the selection of each extreme route. The set E consists of two disjoint sets, E_P and E_D, corresponding to the pickup and delivery extreme routes, respectively. We let δ_e^{fr} indicate the extreme routes associated with aircraft route (f, r): It takes a value of one for extreme routes that correspond to the aircraft route and zero otherwise. For any aircraft route (f, r), the number of indicators with nonzero value is at most two (because each route visits at most two gateway locations and, therefore, the number of extreme routes is at most two).

We require each aircraft route, constructed from its extreme routes, to be integral. That is, $\sum_{e\in E} \delta_e^{fr} w_e$ must be integral for each aircraft route (f, r), while the decision variables, w_e, may have fractional values. The number of decision variables in the new formulation is $|E|$, which is at most twice the number of aircraft route decision variables in *ESSND-R*.

For the aircraft balance, landing, and plane count constraints, the column associated with each extreme route is identical to the corresponding aircraft route column in *ESSND-R*. For the new formulation, we denote these constraints by the matrix $\widehat{\mathbf{B}}$ and the vector β. Similarly, we introduce a new cost vector $\hat{\mathbf{d}}$, with each component identical to the cost of its corresponding aircraft route (taken from the corresponding component of the *ESSND* cost vector, $\mathbf{d}$).

The formulation for the extreme route model (*ER*) is given by:

$$\min \hat{\mathbf{d}}'\mathbf{w} \tag{17}$$

subject to

$$\sum_{e\in E_P} \hat{u}_{gh}^e w_e \ge b_P^{gh}, \quad (g, h) \in \mathcal{K}_P, \tag{18}$$

$$\sum_{e\in E_D} \hat{u}_{gh}^e w_e \ge b_D^{gh}, \quad (g, h) \in \mathcal{K}_D, \tag{19}$$

$$\widehat{\mathbf{B}}\mathbf{w} \le \beta, \tag{20}$$

$$\sum_{e\in E} \delta_e^{fr} w_e \in \mathbb{Z}_+, \; r \in R^f, \; f \in F. \tag{21}$$

Constraints (18) ensure that the total pickup capacity made available for pickup commodity (g, h) exceeds the demand for that commodity (referred to as the *pickup capacity-demand* constraints). Similarly, constraints (19) ensure that the total capacity made available for delivery commodity (g, h) exceeds the demand for that commodity (referred to as the *delivery capacity-demand* constraints). Constraints (20) are the concise representation of the *balance, landing*, and *plane count* constraints described earlier. Finally, constraints (21) ensure that each aircraft route constructed from its extreme routes is selected in integer multiples. Note that the decision variables w_e need not be integral; only the resulting aircraft routes are integral.

2.2.2. Bounds and Strength. Consider any column of *ER* corresponding to an extreme route. The coefficients for the cost (17) and the aircraft side constraints (20) are *the same* as entries for the corresponding aircraft route column in *ESSND-R*. The differences in the two columns are the coefficients for the pickup capacity-demand constraints ((9) and (10) in *ESSND-R* and (18) in *ER*) and delivery capacity-demand constraints ((9) and (11) in *ESSND-R* and (19) in *ER*). Moreover, two extreme routes corresponding to the same aircraft route differ only in their coefficients for the capacity-demand constraints (18) and (19).

We establish a two-way mapping between solutions of *ER* and solutions of *ESSND-R* as follows. Given a solution to *ER*, we construct a set of aircraft routes via the mapping:

$$y_r^f = \sum_{e\in E} \delta_e^{fr} w_e. \tag{22}$$

Given a solution to *ESSND-R*, Proposition 1 guarantees the existence of a convex combination of the (f, r) extreme routes to cover the flow specified by the *ESSND-R* solution, and so the mapping (22) is two way. It is straightforward to show that the two-way mapping (22) preserves feasibility because $\mathbf{B}\mathbf{y} = \widehat{\mathbf{B}}\mathbf{w}$. The mapping also preserves the cost of the solution, that is, $\mathbf{d}'\mathbf{y} = \hat{\mathbf{d}}'\mathbf{w}$. Finally, the mapping also preserves integrality of the solution.

We use these observations to establish the following relationship between the mixed integer program *ESSND-R* and the integer program *ER*.

PROPOSITION 2. *The* ESSND-R *and* ER *are equivalent (mixed) integer programming formulations.*

PROOF. (We argue using pickup routes and the result follows similarly for delivery routes.) Given that the mapping (22) preserves the feasibility, integrality, and cost of the aircraft routes, establishing equivalence requires proving that the mapping ensures feasibility of package flows. Assume we are given an *ESSND-R* solution $(\hat{\mathbf{x}}, \hat{\mathbf{y}})$. All package flows are assigned to aircraft routes. For *each* aircraft route, there exists a convex combination of extreme routes that provides available capacity to cover the flow assigned to that aircraft route (see Proposition 1). Then for each gateway-hub demand, summing the available capacities over all aircraft routes gives $\sum_{e\in E_P} \hat{u}^e_{gh} w_e$, which exceeds the total flow, b^{gh}_P (from constraint (10)). Thus, any feasible integer solution to *ESSND-R* has a corresponding integer solution in *ER* with the same cost.

Conversely, assume we are given a feasible *ER* solution, $\bar{\mathbf{w}}$, and we construct an aircraft route solution, $\bar{\mathbf{y}}$, to *ESSND-R* via (22). Proposition 1 implies the existence of a feasible flow on each aircraft route and, therefore, a feasible flow on the set of routes specified by $\bar{\mathbf{y}}$. The forcing constraints (9) and the demand constraints (10) and (11) are, therefore, satisfied. Moreover, the mapping yields an *ESSND-R* solution with the same cost as the *ER* solution. □

The mapping (22) can be applied to a feasible solution to *ER*'s linear programming relaxation to generate a feasible solution to *ESSND-R*'s LP relaxation with the same cost. For the LP relaxation, the mapping is one way. Hence, it follows that:

PROPOSITION 3. *The linear programming relaxation of* ER *is at least as strong as that of* ESSND-R.

There are cases when the improvement in strength is *strict* and the bound provided by the *ER* LP relaxation is tighter (strictly greater) than the bound provided by the *ESSND-R* LP relaxation. We present an example of such a case in §2.5.

2.3. The Composite Variable Formulation

The *ER* formulation explicitly models aircraft routes and ensures *feasible* package flows by providing sufficient capacity through weighted combinations of extreme routes. We take this one step further by combining routes into *composite variables*, each of which has sufficient capacity to carry some set of commodities. The combined routes might have excess capacity in the same way that *ER* allocates excess capacity. It is likely that further strengthening can occur by reducing coefficients in the composite variables for which excess capacity exists. We motivate this reformulation strategy via the mechanics demonstrated in the following example.

EXAMPLE 2 (COMPOSITE VARIABLE EXAMPLE). Consider the pickup network shown in Figure 3. We must satisfy two gateway-hub demands. The first has demand of $b^{ih}_P = 4{,}000$ packages and the second has $b^{jh}_P = 6{,}000$. We have one double-leg route (i.e., Route 1) from i to j to h and one single-leg route (i.e., Route 2) from j to h. We have two aircraft types, Type 1 with low capacity (5,000 packages) and Type 2 with high capacity (8,000 packages).

For each double-leg route we have two extreme routes (because their capacity does not exceed the sum of the gateway-hub demands). For each of the single-leg routes, we have one extreme route. The capacity-demand constraints of *ER* are given by:

$$\left[\begin{array}{cc|cc|c|c} 4{,}000 & 0 & 4{,}000 & 2{,}000 & 0 & 0 \\ 1{,}000 & 5{,}000 & 4{,}000 & 6{,}000 & 5{,}000 & 6{,}000 \end{array}\right]\mathbf{w} \geq \begin{bmatrix} 4{,}000 \\ 6{,}000 \end{bmatrix},$$

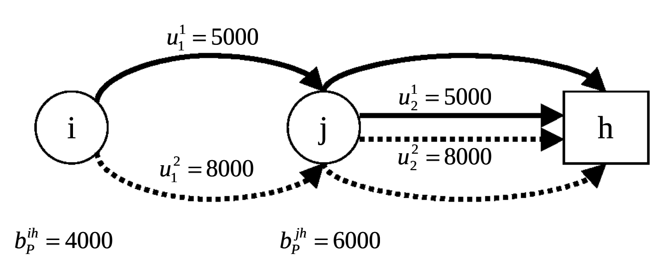

Figure 3 **Two-Gateway, One-Hub Network for the Composite Variable Example**

where the first row corresponds to commodity (i, h) and the second corresponds to commodity (j, h). Dividing each row by its right-hand side yields:

$$\left[\begin{array}{cc|cc|c|c} 1 & 0 & 1 & \frac{1}{2} & 0 & 0 \\ \frac{1}{6} & \frac{5}{6} & \frac{2}{3} & 1 & \frac{5}{6} & 1 \end{array}\right] \mathbf{w} \geq \begin{bmatrix} 1 \\ 1 \end{bmatrix}.$$

While Columns 2 and 5 appear to be identical, the two routes originate at different gateways and their aircraft balance coefficients differ. For extreme routes corresponding to the same aircraft route (e.g., Columns 1 and 2), the entries for constraints other than the capacity-demand constraints are identical. The only difference in the two extreme routes is how they allocate their capacity.

Next, we add columns 3 and 5 to obtain a column with capacity-demand coefficients:

$$\begin{bmatrix} 1 \\ \frac{9}{6} \end{bmatrix}.$$

This indicates that by selecting the composite consisting of routes y_1^2 and y_2^1 we have the available capacity to cover the entire demand of both commodities. Finally, because the right-hand side of each row is 1, we can reduce the second coefficient from $\frac{9}{6}$ to 1 without affecting the set of feasible integer solutions.

Adding a new composite column to the existing set of decision variables does not change the optimal integer solution. Any feasible solution (including the optimal solution) to *ER* remains feasible (optimal) after we add composite variables. We obtain a stronger formulation by removing the extreme routes and including only composites.

2.3.1. Formulation. Before describing the composite variable formulation, we offer the following definitions.

Definition 1. A composite, denoted by c, is a collection of distinct aircraft routes (f, r), $f \in F$, $r \in R^f$. Associated with c are the parameters γ_c^{fr}, which indicate the (integral) number of planes of fleet type f that fly route r in composite c.

Definition 2. A composite, c, is a composite cover of a set of gateway-hub commodities if there exists a feasible flow in c for the entire demand of these commodities. We denote the covered set of commodities as $\mathscr{K}_c \subset \mathscr{K}$.

Let $\mathscr{C}$ be the set of all composite covers. The separation of aircraft routes into pickup and delivery routes allows us to divide the set of composites into two distinct sets, $\mathscr{C}_P$ and $\mathscr{C}_D$, for pickup and delivery, respectively. We let $\delta_c^{gh} = 1$ for each commodity (g, h) covered by composite $c \in \mathscr{C}$ (that is, $(g, h) \in \mathscr{K}_c$). We denote the aircraft side constraint matrix as $\overline{\mathbf{B}}$, with each column of $\overline{\mathbf{B}}$ corresponding to a composite. Letting $\overline{\mathbf{B}}_{(f,r)}$ denote the column of $\mathbf{B}$ corresponding to aircraft route (f, r), then $\overline{\mathbf{B}}_c = \sum_{(f,r)} \gamma_c^{fr} \mathbf{B}_{(f,r)}$, and the elements of $\overline{\mathbf{B}}$ are integral. The composite variable cost vector is denoted by $\bar{\mathbf{d}}$; each component of $\bar{\mathbf{d}}$ is the sum of the costs of each aircraft route contained in the composite.

The composite variable formulation (*CVF*), is defined as follows:

$$\min \bar{\mathbf{d}}' \mathbf{v} \tag{23}$$

subject to

$$\sum_{c \in \mathscr{C}_P} \delta_c^{gh} v_c \geq 1, \quad (g, h) \in \mathscr{K}_P, \tag{24}$$

$$\sum_{c \in \mathscr{C}_D} \delta_c^{gh} v_c \geq 1, \quad (g, h) \in \mathscr{K}_D, \tag{25}$$

$$\overline{\mathbf{B}} \mathbf{v} \leq \beta, \tag{26}$$

$$\sum_{c \in \mathscr{C}} \gamma_c^{fr} v_c \in \mathbb{Z}_+, \; r \in R^f, \; f \in F. \tag{27}$$

Constraints (24) and (25) are the *covering* constraints associated with the pickup demands and delivery demands, respectively. Constraints (26) represent the combined *aircraft balance, landing*, and *plane count* constraints described earlier. Finally, constraints (27) ensure that the selection of each aircraft route is integral.

2.3.2. Bounds and Strength. To relate *CVF* to *ER*, the coefficients δ_c^{gh} arise from the *combination* of *ER* columns and *coefficient rounding* in the resulting column. Specifically, we take a linear combination of *ER* columns to yield integral aircraft routes with known available capacities. The resulting aircraft route columns are summed to give a new column with available capacity to cover some number of gateway-hub demands. The indicator δ_c^{gh} equals 1 if, among the extreme routes that comprise c, the available capacity for commodity (g, h) exceeds the total demand for the commodity. It is set to zero otherwise. In Example 2, we built a composite in which each aircraft route was represented by a single extreme route. In general, we allow each aircraft route to be specified by a combination of its extreme routes and combined with other such aircraft routes to form a composite.

The parameter γ_c^{fr} (an integer) specifies the number of times an aircraft route (f, r) is utilized in composite $c \in \mathscr{C}$. The composites selected by *CVF* must ensure the integrality of the *aircraft routes*, which is why we specify $\sum_{c\in\mathscr{C}} \gamma_c^{fr} v_c$ to be integral. The more restrictive integrality requirement, $v_c \in \mathbb{Z}_+$, is not necessary because integral aircraft routes may be generated by taking fractions of composites.

To map a composite to its extreme routes, we let γ_c^e denote the usage of extreme route e in composite c, where γ_c^e can be fractional. We construct γ_c^{fr} from the extreme route usage by the relation $\gamma_c^{fr} = \sum_{e\in E} \delta_e^{fr} \gamma_c^e$. Any solution to *CVF* is mapped back to the *ER* solution by the relation:

$$w_e = \sum_{c\in\mathscr{C}} \gamma_c^e v_c. \tag{28}$$

We can also use γ_c^e to link the composite's available capacity to a particular gateway-hub demand via the relation:

$$\sum_{e\in E_P} \hat{u}_{gh}^e \gamma_c^e \geq \delta_c^{gh} b_P^{gh}. \tag{29}$$

This says that if the extreme routes selected for a composite provide enough capacity, we can treat the demand as *covered* (i.e., $\delta_c^{gh} = 1$). We can also map any *CVF* solution to an *ESSND-R* solution with the relation:

$$y_r^f = \sum_{c\in\mathscr{C}} \gamma_c^{fr} v_c. \tag{30}$$

Proposition 4. CVF *and* ER *are equivalent integer programming formulations.*

Proof. (Arguments are presented in terms of *pickup* routes and apply similarly to the delivery side.) Given a feasible *ER* solution, $\hat{\mathbf{w}}$, we construct a *CVF* solution, consisting of two columns, as follows. We construct one column in *CVF*, denoted by c^*, such that $\overline{\mathbf{B}}_{c^*} = \sum_{e\in E_P} \widehat{\mathbf{B}}_e \hat{w}_e$, and where $\bar{d}_{c^*}$ is found similarly. For the capacity-demand constraints (18), the composite variable for c^* is found by summing the extreme route columns in *ER*, giving $\bar{u}_{gh}^{c^*} = \sum_{e\in E_P} \hat{u}_{gh}^e \hat{w}_e \geq b_P^{gh}$ for the pickup capacity-demand constraints. So $\delta_{c^*}^{gh} = 1$ for all $(g, h) \in \mathscr{H}_P$. Aircraft usage in c^* is given by $\gamma_{c^*}^{fr} = \sum_{e\in E_P} \delta_e^{fr} \hat{w}_e$, which is integral for all (f, r). We similarly define a second composite, c^{**}, for the delivery side. Let $\hat{v}_{c^*} = 1$ and $\hat{v}_{c^{**}} = 1$, thus satisfying all constraints in *CVF* with the same cost as the *ER* solution. Finally, $\gamma_{c^*}^{fr} \hat{v}_{c^*} + \gamma_{c^{**}}^{fr} \hat{v}_{c^{**}}$ is integral due to the integrality of $\gamma_{c^*}^{fr}$ and $\gamma_{c^{**}}^{fr}$, for all (f, r).

Conversely, assume we are given a *CVF* solution, $\bar{\mathbf{v}}$. We map $\bar{\mathbf{v}}$ to an *ER* solution, $\bar{\mathbf{w}}$, as in (28). Using the capacity relation (29), the capacity assigned to commodity $(g, h) \in \mathscr{H}_P$ in *ER* is:

$$\begin{aligned}\sum_{e\in E_P} \hat{u}_{gh}^e \bar{w}_e &= \sum_{e\in E_P} \sum_{c\in\mathscr{C}_P} \hat{u}_{gh}^e \gamma_c^e \bar{v}_c \\ &\geq \sum_{c\in\mathscr{C}_P} \bar{v}_c \delta_c^{gh} b_P^{gh} \\ &\geq b_P^{gh},\end{aligned}$$

and capacity-demand constraints (18) are satisfied. The delivery capacity-demand constraints (19) are similarly satisfied. The side constraints (20) in *ER* are satisfied because $\widehat{\mathbf{B}}\bar{\mathbf{w}} = \sum_{e\in E} \sum_{c\in\mathscr{C}} \widehat{\mathbf{B}}_e \gamma_c^e \bar{v}_c = \sum_{c\in\mathscr{C}} \overline{\mathbf{B}}_c \bar{v}_c = \overline{\mathbf{B}}\bar{\mathbf{v}} = 0$. We similarly establish that the cost of the *ER* solution is the same as the cost of the *CVF* solution. Integrality of the aircraft routes follows directly from mapping (30), completing the proof. □

For a feasible solution to the LP relaxation of *CVF*, we establish the following result directly from the arguments used in proving the converse of Proposition 4 by showing a feasible solution to *CVF*'s LP relaxation has a corresponding feasible solution to the *ER* LP relaxation with the same cost.

Proposition 5. *The* CVF *LP relaxation is at least as strong as that of* ER.

Just as *ER's* LP relaxation can be strictly greater than that of *ESSND-R*, *CVF*'s LP relaxation can be strictly greater than that of *ER*; we demonstrate this later with two examples. The main result of this section follows directly from Propositions 2 and 4 for equivalence and Propositions 3 and 5 for strength.

Proposition 6. CVF *is equivalent to* ESSND-R, *and its LP relaxation is at least as strong as that of* ESSND-R.

In addition to the improved bounds, composites provide a means for capturing difficult constraints that would, when using conventional network design methods, yield intractable models. One example of such a constraint is the requirement to model commodities as integers, which is the case when we plan container, rather than package, movements (a container holds hundreds of packages). For the conventional *ESSND* formulation, we would need to add the requirement for all flow variables to be integral. With the composite variable formulation, enumerative procedures for constructing composites can ensure integrality of flows; the formulation itself remains unchanged. A second complicating constraint exists when we relate aircraft capacity to the distance flown. This type of range-payload trade-off is easily modeled using composite variables—we simply use the reduced capacity in the enumerative construction of our composites. Such operating constraints affect the composite construction procedures, but they do not interfere with the structure of the composite variable formulation. Moreover, they often reduce the size of the composite variable set, which generally makes the model easier to solve.

2.4. Two-Node Example

To illustrate how the formulation is strengthened as we move from *ESSND-R* to *CVF*, we present a simple two-node problem shown in Figure 4. The route from gateway j to hub h can be flown by two aircraft types. The first has a capacity of 5,000 packages and a cost of 3. The second has a capacity of 8,000 packages and a cost of 4. Our objective is to move all 6,000 packages at minimum cost from the gateway to the hub.

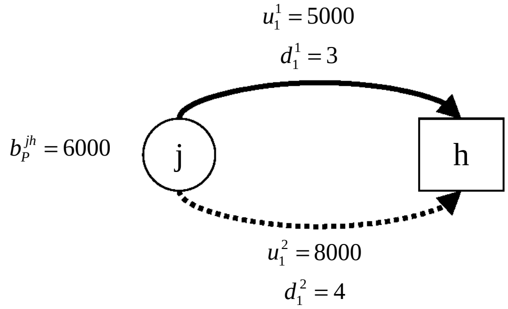

Figure 4 Simple Two-Node Network Demonstrating Formulation Strength

The *ESSND-R* formulation (excluding the landing, plane count, and aircraft balance constraints)

$$\min 3y_1^1 + 4y_1^2$$

subject to

$$x_1^{jh} - 5{,}000y_1^1 - 8{,}000y_1^2 \le 0,$$
$$x_1^{jh} = 6{,}000,$$
$$y_1^1, y_1^2 \in \mathbb{Z}_+.$$

We could simplify this formulation but will keep it in the *ESSND* form for the purpose of exposition. The optimal solution to the LP relaxation can be found by flowing all packages on the aircraft with the lowest cost per unit of capacity. The solution to the LP relaxation is $x_1^{jh} = 6{,}000$, $y_1^1 = 0$, and $y_1^2 = 0.75$, with a total cost of 3. Note that the optimal integer solution is to fly the Type 2 aircraft, or $y_1^2 = 1$, with a cost of 4.

Next, we construct the *ER* formulation. With two *single-leg* routes, the *ER* formulation has two extreme routes, one for each aircraft route. The available capacities of our two extreme routes are $\hat{u}_{jh}^1 = \min\{b_P^{jh}, u_1^1\} = 5{,}000$ and $\hat{u}_{jh}^2 = \min\{b_P^{jh}, u_1^2\} = 6{,}000$. *ER* is given by:

$$\min 3w_1 + 4w_2$$

subject to

$$5{,}000w_1 + 6{,}000w_2 \ge 6{,}000,$$
$$w_1, w_2 \in \mathbb{Z}_+.$$

Using the same reasoning to solve *ER*'s LP relaxation we find the optimal solution for the LP relaxation to be $w_1 = 1.2$ with a cost of 3.6, which is a better (higher) bound on the optimal solution. The essence of bound improvement is as follows. *ESSND-R* flew a fractional aircraft to avoid being charged for the unused capacity. By using extreme routes with capacities that are lower than the capacities of the original routes, *ER* reclaims part of the empty portion of aircraft routes. That is, *ER* places total aircraft route cost on the reduced capacity rather than on the actual capacity. The *ESSND-R* fractional solution of $\frac{3}{4}$ Type 2 aircraft *cannot* be represented in *ER*. The available capacity for the Type 2 aircraft is $\hat{u}_{jh}^2 = 6{,}000$. By selecting $\frac{3}{4}$ of the Type 2 aircraft in *ER*, we would only be able to flow 4,500 packages, which is infeasible. Note that the optimal integer solution to *ER* is $w_2 = 1$ with a total cost of 4.

We reformulate *ER* by scaling the capacity-demand constraints:

$$\min 3w_1 + 4w_2$$

subject to

$$\tfrac{5}{6}w_1 + 1w_2 \geq 1,$$
$$w_1, w_2 \in \mathbb{Z}_+.$$

The second extreme route is itself a composite because it covers the entire demand. We call this Composite 1. We build a second composite variable by doubling the first column. We'll call this Composite 2. The mapping of each composite to aircraft routes (γ_c^{fr}) is given by $\gamma_1^{1,1} = 0$ and $\gamma_1^{2,1} = 1$ for Composite 1 and $\gamma_2^{1,1} = 2$ and $\gamma_2^{2,1} = 0$ for Composite 2. *CVF* is then:

$$\min 4v_1 + 6v_2$$

subject to

$$v_1 + v_2 \geq 1,$$
$$v_1\gamma_1^{1,1} + v_2\gamma_2^{1,1} \in \mathbb{Z}_+,$$
$$v_1\gamma_1^{1,2} + v_2\gamma_2^{1,2} \in \mathbb{Z}_+.$$

The first integrality requirement reduces to $2v_2 \in \mathbb{Z}_+$ and the second reduces to $v_1 \in \mathbb{Z}_+$. The optimal solution to the LP relaxation of *CVF* is $v_1 = 1$ with a cost of 4.0. The bound provided by *CVF* is tighter than the bound provided by *ER*. In fact, the LP relaxation gives the optimal integer solution. Recall that the fractional solution to *ER* used $\frac{6}{5}$ of Type 1 aircraft and *no* Type 2 aircraft. There is no corresponding feasible solution in *CVF*, as any *CVF* solution that uses *no* Type 2 aircraft is forced to use at least *two* Type 1 aircraft. This accounts for *CVF* being stronger than *ER*.

To summarize this phenomenon, the presence of excess capacity in aircraft routes increases the opportunity for fractionality in the solution to the *ESSND-R* LP relaxation. By defining a model that uses extreme routes, we are able to absorb part of the excess capacity through coefficient reduction, thereby removing some of this fractionality. In *ER*, however, this absorption is accomplished only on individual aircraft routes with capacity exceeding one of its gateway-hub demands. When individual aircraft routes do not cover their gateway-hub demands, we create composite variables to cover these demands. The process of combining routes may, however, result in excess capacity, which we can absorb by reducing the capacity-demand coefficients, which then leads to the covering constraints seen in *CVF*.

2.5. Single-Hub Example

To provide a sense of how the increased strength affects the computational workload, we present a small single-hub example. Without package flow costs, there is no benefit in switching packages between planes at an intermediate gateway location when all packages are bound for the same hub. The result is a case that satisfies the no-ramp-transfer assumption and that provides a clear example of the improvement in formulation strength as we move from *ESSND* to *CVF*.

We consider a network with a single hub and five gateway locations. Timing restrictions exist for pickup and delivery at each of five gateways and for the hub sort. We use three fleet types, ranging in capacity from 8,000 packages to 10,000 packages. One fleet type is restricted to flying only single-leg routes. The solutions for the three formulations were generated using XPRESS-MP v.10 on a 300 Mhz Pentium PC.

The results are summarized in Table 1. These are consistent with what we have established with respect to the strength of the three models. As we move from *ESSND-R* to *CVF*, we see the LP relaxations give better approximations to the optimal integer solution. In fact, for this example, the composite variable formulation's LP relaxation returns the optimal integer solution. In general, the LP relaxation is not guaranteed

Table 1 Solution Summary for *ESSND-R*, *ER*, and *CVF* Applied to Single-Hub Example

	ESSND-R	ER	CVF
LP relaxation objective value	10,663.037	23,154.683	28,474.014
IP objective value	28,474.014	28,474.014	28,474.014
Nodes in branch-and-bound tree	781	111	1
Gap $\left(\frac{z_{IP}^* - z_{LP}^*}{z_{IP}^*}\right)$	0.6255	0.1868	0.0000

to generate an integer solution at the root node. Yet, the tighter bounds provided by the composite variable formulation allow faster generation of very good integer solutions via branch and bound and, in some cases, the possibility of establishing the optimality of a feasible integer solution.

2.6. Ramp Transfers

To this point, we have worked under the assumption that ramp transfers cannot occur, yet the general definition of composites includes the case when ramp transfers are allowed. The presence of ramp transfers has no effect on the strength results of *CVF* and *ESSND*, but does cause an increase in the size of the composite set. Under several key assumptions relating to the carrier's operation, the number of composite variables does not explode and enumerative procedures can be used to construct the entire set of composites.

Flow Assumption 1. *Double-leg routes shall only be used if they cover at least one of their gateway-hub demands.*

Under this assumption, a double-leg aircraft route can only be represented by its actual extreme routes, not convex combinations of its extreme routes. If the total demand to be picked up on a double-leg route exceeds the capacity, any (nontrivial) convex combination of the extreme routes will allocate capacity that covers neither of the gateway-hub demands. If a double-leg route has only one extreme route (i.e., the capacity of the aircraft route exceeds the gateway-hub demand to be moved), there is no convex combination to take. The same is true for single-leg routes. Thus, we represent an aircraft route using its extreme routes, not convex combinations of the extreme routes.

The implication to *ER* is that we can replace the requirement for integral aircraft routes (which are constructed from extreme routes) with the requirement for integral extreme routes. That is, replace $\sum_{e\in E} \delta_e^{fr} w_e \in \mathbb{Z}_+$, for all (f, r), with the requirement $w_e \in \mathbb{Z}_+$, for all $e \in E^*$. E^* denotes the subset of extreme routes that satisfy the operational assumption. The only extreme routes not included in E^* are double-leg extreme routes where none of the allocated capacities covers its gateway-hub demands. This happens when one of the gateway-hub demands exceeds the entire capacity of the aircraft route. The implication to *CVF* is that composites can be built simply using the extreme routes in E^*.

Flow Assumption 2. *A gateway-hub demand that is ramp transferred must be transferred in its entirety.*

This implies that the gateway-hub commodity must fit entirely on the first leg of the first aircraft involved in the transfer and on the second leg of the second aircraft involved in the transfer. When we build extreme routes, the excess capacity on each leg of the route is known. For each gateway location, we search the legs with excess capacity to see if a given gateway-hub demand can be transferred. Furthermore, we check all flight legs to ensure that the proper timing exists to make the transfer.

Flow Assumption 3. *A gateway is only included on routes to/from hubs in whose territory that gateway lies.*

A hub's territory is simply defined as the set of locations which have demand going to or from the hub. It follows that the number of planes involved in a ramp transfer at any given gateway is limited to the number of hubs to which that gateway is connected.

As a result of these assumptions, we construct the complete set of composites as follows: First, we identify the possible aircraft routes, considering issues such as level-of-service requirements, speed and range of aircraft, and hub sort hours of operation. We then create the extreme routes for each aircraft route. Those that cover their gateway-hub demands are single-route composites. We use the remaining extreme routes to construct multiple-route composites without considering ramp transfers. Finally, we search the set of composites to look for opportunities to ramp transfer additional commodities using excess capacity on the legs of those routes. The result is a "composite of composites" tied together by the ramp transfer of one or more commodities. This enumerative process allows us to create an oracle of decision variables and incorporate them into the formulation either a priori or via explicit column generation.

Under the given operational assumptions, this is the fully enumerated set of combined aircraft routes. Because we include all possible combinations in the composite set, we can change the integrality requirement in *CVF* from enforcing the selection of integer aircraft routes to enforcing integer composites. That is, we change the requirement $\sum_{c\in\mathscr{C}} \gamma_c^{fr} v_c \in \mathbb{Z}_+$ for all $r \in R^f$, $f \in F$ to the requirement $v_c \in \mathbb{Z}_+$ for all $c \in \mathscr{C}$. In addition, the composite variable formulation actually benefits from the complexity of this planning problem, specifically from the aircraft capacity limitations, the timing of aircraft routes, and the operational restrictions on package flows. While these additional restrictions make the enumerative procedures more complicated, they make the composite variable formulation easier to solve because the number of composite variables is reduced.

The results pertaining to formulation strength derived under the "no-ramp-transfer" assumption still apply. Any solution to *CVF* uses aircraft routes that satisfy the landing, plane count, and aircraft balance constraints. They cover all demands with their capacity because each composite covers some set of commodities and because the covering constraints ensure that all commodities are covered. Thus, there exists a corresponding solution in *ESSND* using the same aircraft routes and having the same cost. Likewise, a fractional solution to *CVF* satisfies all constraints and there exists a corresponding fractional *ESSND* solution. Finally, any integral solution to *ESSND* that satisfies the constraints and the flow assumptions is easily mapped into a set of composites that cover all commodity demands. As a result, *CVF* is at least as strong as *ESSND*. In practice, this is typically a strict improvement in strength.

3. The Case of the United Parcel Service

The *CVF* model was designed to support planning function at the United Parcel Service (UPS), with three groups of planners and analysts that will use the model. First, the *long-range planners* develop network plans for periods of 2–10 years in the future to determine what assets will be needed to operate such a system. *Network planners* work on current year plans and adjust existing plans to accommodate actual or anticipated changes in the system. *Peak planners* focus on the network plan to enable operations during the peak retail season of November and December.

We apply *CVF* to the Next Day Air network of UPS. The air network consists of 101 locations, seven of which are hubs. The carrier's aircraft inventory includes seven aircraft types and 160 total aircraft. Demand consists of a nightly volume of 2,250 containers carrying 926,268 packages on the pickup side and 2,288 containers carrying 967,172 packages on the delivery side (planners typically work with containers as they provide a more realistic characterization of the demand's footprint on an aircraft). The objective is to minimize operating cost, which depends upon fleet type and consists of the cycle cost incurred for each leg flown and the variable operating cost that depends on the duration of a route. All computations were performed on an HP9000 Model D370 using HP's ANSI C/C++ compiler with calls to the ILOG CPLEX 6.5 Callable Library (see ILOG 1999).

We first examine the effect of the complexity of ramp transfer composites on solution quality and run time. Recall that for each gateway-hub pair (g, h) involved in a ramp transfer, the entire (g, h) demand is assigned to a single aircraft on the first leg and a single aircraft on the second leg. This follows from the planning consideration that if gateway-hub volume is to be ramp transferred it cannot be split, neither at its origin nor at its ramp transfer gateway.

Planners at UPS place an upper limit on the amount of an inbound aircraft's capacity that can be used for ramp transfer volume. Standard planning procedures place this limit at 50% because of the time associated with unloading the aircraft. Increasing this limit provides more opportunities for these package transfers, thus increasing the number of composite variables in the formulation.

We begin by setting the ramp transfer volume limit to 15% of the inbound aircraft route's capacity. The reduction in the number of composite variables enables us to find the optimal integer solutions. This allows us to examine the effect of allowing a greater number of aircraft to transfer packages at a given location, thereby increasing the complexity of interactions between aircraft routes. We are primarily interested in the effect of the complexity of composite variables on the bounds provided by *CVF*'s linear

Table 2 *CVF* **Solution Varying Maximum Number of Aircraft Routes Allowed in Ramp Transfer Composites**

		Maximum Number of Aircraft Routes in Ramp Transfer			
		0	2	3	4
Composites	Single route	7,237	7,237	7,237	7,237
	Multiroute	24,078	24,078	24,078	24,078
	Ramp transfer	0	12,287	28,462	36,092
	Ferry routes	9,700	9,700	9,700	9,700
Problem size	Columns	38,838	53,198	69,373	77,003
	Rows	1,064	1,064	1,064	1,064
	Nonzeroes	244,090	397,484	667,533	827,064
Objective value ($M)	LP relaxation	1.65048	1.62357	1.62158	1.62158
	First integer	1.65518	1.64008	1.63468	1.63601
	Optimal integer	1.65351	1.62766	1.62586	1.62586
	LP-IP gap	0.0018	0.0025	0.0026	0.0026
Run time (sec)	LP relaxation	44.66	49.38	102.29	105.94
	Optimal integer	2,320	4,928	11,817	6,753

programming relaxation, the quality of the solution, and the run time of the model.

Table 2 shows the effect of varying the maximum number of aircraft routes allowed in each ramp transfer composite from no ramp transfers to those involving four aircraft. The objective function value of the best integer solution is reduced by a total of 1.56% when we introduce the two-aircraft route ramp transfers and by a total of 1.67% when we use three-aircraft route ramp transfer composites. Note that the addition of four-route ramp transfers improves neither the LP relaxation nor the best lower bound. Finally, notice the behavior of the LP-IP gap. When ramp transfers are included, the gap jumps from 0.18% to 0.25%, but building more complex ramp transfer variables has little additional effect on the gap.

We next examine the effect of the aircraft balance constraints. As noted in §2, these constraints propagate fractionality throughout the network. This fractionality would otherwise be localized to a single aircraft route. Thus, the presence of these balance constraints will likely worsen the bounds provided by the linear programming relaxation. This type of constraint is common to many transportation network design problems where both the commodities and the units of capacity flow through the network. Using the composite set without ramp transfers, we run *CVF* with and without the aircraft balance requirement at gateway locations. We continue to enforce balance at the hubs; that is, the number of planes of a given aircraft type arriving to a hub's sort must be offset by the same number of planes of that type departing the hub following the sort.

Comparing the solutions in Table 3 leads to three key observations. First, balancing the design variables is detrimental to run time. Second, without balance, the LP relaxation provides a much tighter bound on the optimal integer solution, though the gap found with balance is still very tight. Third, without balance, *CVF* produces a solution with significantly lower cost (almost $76 thousand per day, or a reduction of 4.59%). If aircraft used in the Next Day Air (i.e., overnight) network can be repositioned through clever construction of aircraft routes during daytime operations, we might be able to realize a por-

Table 3 *CVF* **Solution with and Without Gateway Balance**

		With Balance	Without Balance
Composites	Single route	7,237	7,237
	Multiroute	24,078	24,078
Problem size	Columns	26,525	29,216
	Rows	909	434
	Nonzeroes	205,170	174,208
Objective value ($M)	LP relaxation	1.65120	1.57725
	First integer	1.65906	1.58839
	Optimal integer	1.65434	1.57843
	LP-IP gap	0.0019	0.00075
Run time (sec.)	LP relaxation	27.23	18.46
	Optimal integer	1,188	121.09

tion of these NDA operating cost savings. Thus, a natural, and likely profitable, extension to *CVF* is its application to the combined overnight/daytime network design problem.

Finally, we apply *CVF* with the full set of composites (see Table 4), which includes ramp transfers with a maximum transfer of 50% of the inbound plane's capacity, as UPS planners specify. With the nightly operating cost of the Next Day Air network in the millions of dollars, each percentage point translates to significant savings. The 6.96% reduction in operating cost translates to more than $20 million annually. The more significant reduction comes in the number of aircraft and, therefore, the total ownership cost. The 10.74% reduction corresponds to *CVF* using 16 fewer aircraft than the plan generated by the UPS planners. With the cost of a single airframe in the neighborhood of one hundred million dollars—take UPS's recent purchase of 60 Airbus A300-600 aircraft for $6 billion—avoiding or deferring the cost of a single aircraft yields significant savings.

The gaps shown in the table are defined as follows: "Best Bound" is the difference between the best integer solution and the best lower bound, divided by the best lower bound, and "LP-IP" is the difference between the best integer solution and the root node LP relaxation, divided by the root node LP relaxation. These gaps are similar because the best lower bound does not improve significantly in the branch-and-bound tree. The running time for *CVF* to obtain the fifth integer solution, which is the solution we report in the table, was just over 100 minutes. Thus, our new formulation strategy allows us to generate high quality solutions quickly. By allowing the run to continue, we can search for a better integer solution, the presence of which is likely, given the tight gaps found in the earlier computational tests. Conventional network design approaches were unable to yield an integer solution for this scenario.

Table 4 *CVF* **Versus Planners' Solution, with Objective to Minimize Operating Cost**

		CVF
Problem size	Columns	124,572
	Rows	1,117
	Nonzeroes	1,492,014
Solution	Operating cost	6.96%
(% improvement from	Cycle	4.74%
carrier's solution)	Hourly	8.22%
	Number of aircraft	10.74%
	Aircraft ownership cost	29.24%
	Total cost	24.45%
Run time	LP relaxation	317.10
(sec.)	Best IP	6,324
Optimality gaps	Best bound	2.14%
	LP-IP	2.14%

We contrast routes generated manually by the UPS planners and those generated by *CVF*. In Figure 5, we have extracted the aircraft routes of a single fleet type from the complete solution for all seven fleet types. The timing requirements, capacities, landing restrictions, and aircraft range result in a planning problem with so many dimensions that a natural way to manually plan routes is to have the delivery routes mirror the pickup routes. This is clearly seen in the planners' solution. *CVF*, however, can handle all modeling dimensions and generate a solution that better allocates aircraft capacity. This is demonstrated in the asymmetry we see in the *CVF* routes. In spite of this asymmetry, *CVF* ensures that all planes starting from a given location are balanced by a plane of the same fleet type landing at that location.

For short-term planning with a *fixed aircraft inventory*, *CVF* can be used to redesign the air network for the peak retail season in November and December to reduce or eliminate the need to lease temporary aircraft. For long-range planning, using fewer aircraft for future service allows the carrier to purchase fewer new aircraft or to avoid accelerating production under existing contracts. The savings are potentially enormous, given the cost of new aircraft. Implementation and use of this model within the UPS Airline began recently with tremendous success in the first months of operation. Details of that work are highlighted in a companion paper (see Armacost et al. 2002).

4. Conclusions

The inherent difficulty of the Express Shipment Service Network Design problem and the massive size of the instances we consider cause conventional network

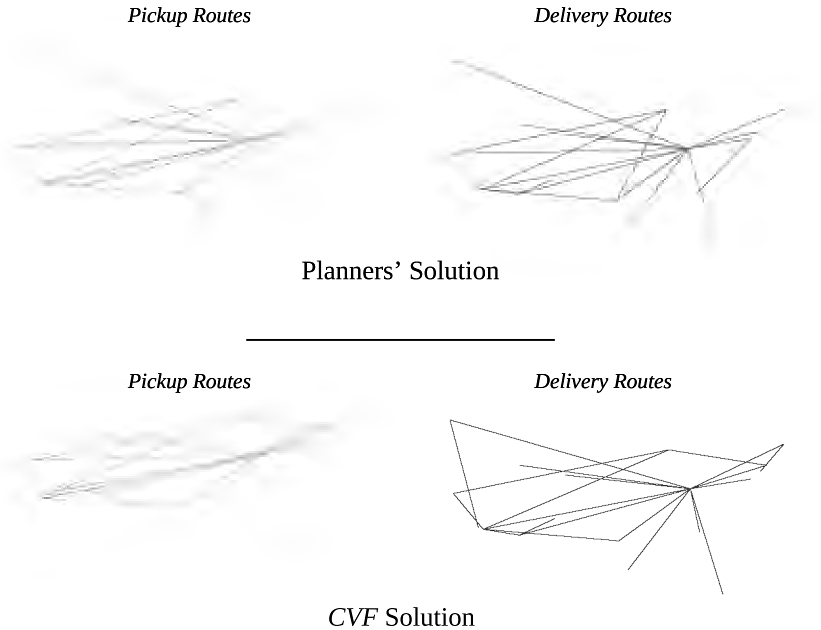

Figure 5 Comparing Planners' Routes Versus *CVF's* Routes for a Single Fleet Type

design formulations to fail. By removing package flows as explicit decisions, we construct an equivalent formulation that provides linear programming relaxations with stronger lower bounds. We further strengthen our formulation by considering sets of aircraft routes, called *composites,* which implicitly represent package flows and cover the demand of some set of commodities.

With *composite variables,* we have tractable formulations for this large-scale planning problem. The potential impact on a company such as the United Parcel Service is tremendous. The planning cycle can be reduced from months to days, the operating cost of the resulting network reduced by almost 7%, and the number of required aircraft reduced significantly. This translates directly to savings in the hundreds of millions of dollars.

The flexibility of the composite variable approach allows the model to easily adjust to additional constraints. Such constraints hurt conventional formulations, as they simply add complexity to the formulation. However, they actually help the composite variable formulation, as these constraints further reduce the size of the set of composite variables in the formulation. A very important example of this is when we want to ensure that commodity flows are integral, as we did when demand projections are specified by containers. We can handle this easily in our construction of the composite variable set, but this integrality requirement would make the already intractable conventional network design formulation, *ESSND,* even more difficult to solve.

While the general applicability of composite variable formulation is not completely known, the theoretical basis and utility to the large-scale planning problem faced by UPS is clear. The *potential* of composite variable formulations can be seen in a broader set of problems. Exploring a combined Next Day Air and Second Day Air problem will allow the UPS planners to consider both overnight and daytime operations simultaneously. Further, the resulting dual information from the composite variable formulation may provide a basis for adjusting a key input to this planning problem: the gateway-to-hub assignments for commodities. In the passenger airline industry, composite variable formulations can be applied to the problem of determining fleet assignments and origin-destination passenger flows simultaneously (see Barnhart et al. 2000).

In the more general setting of network design and fixed-charge problems, composite variable formulations can be applied to core problems such as the pure fixed-charge transportation problem (PFCTP). This problem provides a starting point for determining the classes of problems to which composite variable formulations apply, for developing generalized composite variable approaches, and for efficiently handling large numbers of decision variables through either implicit or explicit column generation.

Acknowledgments

The authors appreciate the contributions of the three anonymous referees whose insightful comments strengthened this paper.

References

Agrawal, A., P. Klein, R. Ravi. 1995. When trees collide: An approximation algorithm for the generalized Steiner problem on networks. *SIAM J. Comput.* **24** 440–456.

Armacost, A. P., C. Barnhart, K. A. Ware, A. M. Wilson, W. C. DuPuy. 2002. Planning the United Parcel Serivce air network. Forthcoming, *Interfaces*.

Barahona, F. 1996. Network design using cut inequalities. *SIAM J. Optim.* **6** 823–837.

Barnhart, C., R. R. Schneur. 1996. Air network design for express shipment service. *Oper. Res.* **44** 852–863.

——, A. Farahat, M. Lohatepanotot. 2000. Extending fleet assignment models and algorithms. Working paper, Operations Research Center, Massachusetts Institute of Technology, Cambridge, MA.

Bertsimas D., C. P. Teo. 1998. From valid inequalities to heuristics: A unified view of primal-dual approximation algorithms in covering problems. *Oper. Res.* **46** 503–514.

Bienstock, D., O. Günlük. 1995. Computational experience with a difficult mixed-integer multicommodity flow problem. *Math. Programming* **68** 213–237.

——, ——. 1996. Capacitated network design—Polyhedral structure and computation. *INFORMS J. Comput.* **8** 243–259.

——, S. Chopra, O. Günlük, C. Y. Tsai. 1998. Minimum cost capacity installation for multicommodity networks. *Math. Programming* **81** 177–199.

Büdenbender, K., T. Grünert, H-J. Sebastian. 2000. A hybrid tabu search/branch-and-bound algorithm for the direct flight network design problem. *Transportation Sci.* **34** 364–380.

Chopra, S., L. Gilboa, S. T. Sastry. 1998. Source sink flows with capacity installation in batches. *Discrete Appl. Math.* **85** 165–192.

Crainic, T. G. 2000. Service network design in freight transportation. *Eur. J. Oper. Res.* **122** 272–288.

Crowder, H., E. L. Johnson, M. W. Padberg. 1983. Soving large-scale zero-one linear programming problems. *Oper. Res.* **31** 803–834.

Gabow, H. N., M. X. Goemans, D. P. Williamson. 1998. An Efficient approximation algorithm for the survivable network design problem. *Math. Programming* **82** 13–40.

Gendron, B., T. G. Crainic, A. Frangioni. 1999. Multicommodity capacitated network design. B. Sanso, P. Soriano, eds. *Telecommunications Network Planning.* Kluwer Academic Publishers, Dordrecht, The Netherlands. 1–19.

Goemans, M. X., D. J. Bertsimas. 1993. Survivable networks, linear programming relaxations and the parsimonious property. *Math. Programming* **60** 145–166.

——, D. P. Williamson. 1995. A general approximation technique for constrained forest problems. *SIAM J. Comput.* **24** 296–317.

Grünert, T., H. J. Sebastian. 2000. Planning models for long-haul operations of postal and express shipment companies. *Eur. J. Oper. Res.* **122** 289–309.

Günlük, O. 1999. A branch-and-cut algorithm for capacitated network design problems. *Math. Programming* **86** 17–39.

Hochbaum, D. S., J. S. Naor. 1996. Approximation algorithms for network design problems on bounded subsets. *J. Algorithms* **21** 403–414.

ILOG. 1999. *CPLEX 6.5 User's Manual.* ILOG, Inc., Incline Village, NV.

Jain, K. 1998. A factor 2 approximation algorithm for the generalized Steiner network problem. *Proc. 39th Annual Sympos. on Foundations of Computer Science (FOCS '98).*

Karger, D. R. 1999. Random sampling in cut, flow, and network design problems. *Math. Oper. Res.* **24** 383–413.

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

Kuby, M., R. Gray. 1993. The hub network design problem with stopovers and feeders: The case of Federal Express. *Transportation Res. A: Policy and Practice* **27A** 1–12.

Magnanti, T. L., P. Mirchandani. 1993. Shortest paths, single origin-destination network design and associated polyhedra. *Networks* **23** 103–121.

——, R. T. Wong. 1984. Network design and transporation planning: Models and algorithms. *Transportation Sci.* **18** 1–55.

——, P. Mirchandani, R. Vachani. 1993. The convex hull of two core capacitated network design problems. *Math. Programming* **26** 233–250.

——, ——, ——. 1995. Modeling and solving the two-facility capacitated network loading problem. *Oper. Res.* **43** 142–157.

Minoux, M. 1989. Network synthesis and optimum network design problems: Models, solution methods and applications. *Networks* **19** 313–360.

Padberg, M. W., T. J. Van Roy, L. A. Wolsey. 1985. Valid linear inequalities for fixed charge problems. *Oper. Res.* **33** 842–861.

Pochet, Y., L. A. Wolsey. 1995. Integer knapsack and flow covers with divisible coefficients: Polyhedra, optimization, and separation. *Discrete Appl. Math.* **59** 57–74.

Stallaert, J. 2000. Valid inequalities and separation for capacitated fixed charge flow problems. *Discrete Appl. Math.* **98** 265–274.

Standard and Poor's. 2000. Commercial transportation industry survey. February.

Van Roy, T. J., L. A. Wolsey. 1985. Valid inequalities and separation for uncapacitated fixed charge networks. *Oper. Res. Letters* **4** 105–112.

Williamson, D. P., M. X. Goemans, M. Mihail, V. V. Vazirani. 1995. A primal-dual approximation algorithm for generalized Steiner network problems. *Combinatorica* **15** 435–454.

Wolsey, L. A. 1975. Faces of linear inequalities in 0-1 variables. *Math. Programming* **8** 165–178.

Received: November 2000; revision received: August 2001; accepted: September 2001.