Cost Modeling and Design Techniques for Integrated Package Distribution Systems

Karen R. Smilowitz*and Carlos F. Daganzo†

June 27, 2002

Abstract

Complex package distribution systems are designed using idealizations of network geometries, operating costs, demand and customer distributions, and routing patterns. The goal is to find simple, yet realistic, guidelines to design and operate a network that is integrated both by transportation mode and service level; i.e., overnight (express) and longer (deferred) deadlines. The decision variables and parameters that define the problem are presented along with the models that approximate the total cost of operation. The design problem is then reduced to a series of optimization subproblems that can be solved easily. The proposed approach provides valuable insight for the design and operation of integrated package distribution systems. Qualitative conclusions suggest that benefits of integration are greater when deferred demand exceeds express demand. This insight helps to explain the different business strategies of package delivery firms in industry today.

This paper introduces design strategies and cost modeling techniques for multiple mode, multiple service level package delivery networks where service levels are defined by guaranteed delivery times (i.e., overnight, two-day delivery). The design and operation of large-scale transportation networks is difficult due to the large number of decision variables and constraints, and their intricate interdependencies. This is particularly true of the complex hierarchical networks adopted in the package delivery industry. Two principal approaches have been employed in the literature to address components of this problem: mixed-integer programming with detailed discrete data and continuous approximations. While the former provide a much higher level of detail, the latter are more revealing of "the big picture".

Numerical optimization approaches to network modeling have been studied extensively, see Magnanti and Wong (1984); Ahuja *et al.* (1993); Ball *et al.* (1995). As discussed in these and other more general references; e.g., Nemhauser and Wolsey (1999), optimal solutions can be found numerically for small network problems. In some special cases, it is possible to solve large problems. In general, however, as the network size increases, problems become more difficult to solve, and heuristic approaches are often necessary. Numerical optimization models have been successful in solving tactical and operational problems for transportation networks, offering detailed, cost-minimizing operating plans; see review in Crainic (2000), as well as Powell and Sheffi (1983); Barnhart and Schneur (1996); Armacost (2000). However, collecting demand and cost data for these models can be time-sensitive and, at times, impossible.

On the other hand, continuous approximation models use smooth functions to describe the data, such as a demand density function that varies with location, see for example Daganzo and Newell (1986). Smooth functions are also used to describe decisions (in place of decision variables), e.g. as in the case of spatially varying terminal densities. Knowledge of these decision functions gives enough information to develop a network configuration and an operating plan with a predictable

*Department of Industrial Engineering and Management Sciences, Northwestern University
†Department of Civil and Environmental Engineering, University of California, Berkeley

cost, see Daganzo (1999). Early work on approximation methods (Eilon *et al.* (1971); Newell (1973); Geoffrion (1976)) recognized that approximations provide near optimal solutions and offer valuable insight into operating strategies and network design.

Whereas numerical optimization models perform better on smaller problems, the opposite is true with continuous approximations. The larger the problem, the more accurate the approximations become, see Daskin (1985); Campbell (1993); Daganzo (1999). Although continuous approximation methods are well suited for the design of large-scale transportation networks, the topic has not been explored thoroughly. A recent review paper (Langevin *et al.* (1996)) identifies gaps in the state of the art. This paper fills key gaps described below.

Current multiple origin/multiple destination distribution models do not adequately incorporate multiple transshipments and multistop (peddling) tours. Such activities should be included to realistically model complex systems. Additional operating costs beyond transportation and inventory are also missing. For the most part, continuous approximation models have not considered the cost of repositioning empty vehicles, except perhaps Jordan and Burns (1984); Hall (1991).

Several studies have extended earlier work on analytic models to consider distribution of time sensitive items, see Han (1984); Daganzo (1987a,b); Kiesling (1995); however, multiple service levels have not been addressed in these papers. Multiple transportation modes have been included in a limited number of continuous models, see Hall (1989).

There is also a need for further integration of continuous approximation and discrete models. Previous work on hybrid continuous approximation and numerical optimization models can be found in Hall (1986); Robuste *et al.* (1990); Campbell (1993). Using the two approaches together can provide a complete design and operating methodology for complex logistics systems.

This paper is part of a larger research project (Smilowitz (2001)) that combines continuous approximation and discrete methodologies to study complex multimodal package delivery systems. It was motivated by the need to study the various degrees of mode and service level integration within the package delivery industry. Tactical and operational decisions involving the routing of vehicles and items are covered in Smilowitz *et al.* (2002). More qualitative findings can be found in Smilowitz (2001). This paper presents a complete modeling framework for strategic design problems for large-scale integrated distribution networks with some results and performance tests. While the package delivery network design problem is quite complex, we demonstrate the ability to estimate costs and obtain designs for such systems using continuous approximations.

Section 1 introduces the package delivery network design problem, and discusses network configurations and routing principles. Section 2 presents the cost modeling formulation and section 3 the solution method. Highlights of results and validation tests are presented in section 4.

1 Problem formulation

Package delivery firms operate very complex networks. A typical Federal Express or UPS package passes through a hierarchy of terminals en route from origin to destination, transported by several modes. Here we present a stylized version of the package delivery network design problem.

Given are expected demand data by origin and destination, and by service level. Two service levels are assumed: $\mathcal{S} = \{E, D\}$ for express and deferred demand. Express products are highly time sensitive; deferred packages are not. Items are assumed to be the same size. Given are operating parameters for vehicle and facilities. Two transportation modes are assumed: air and ground. All local and access (regional) transportation is conducted by ground vehicles (delivery vans, trucks, etc.), but long haul transportation can be performed by ground (tractor-trailers) and air.

An integrated distribution network, as shown in Figure 1, operates in the following manner. Items first travel via local pick-up tours to the nearest regional consolidation terminal where cargo within the region is consolidated for efficient long haul transportation. Items are then transported along access routes from consolidation terminals to either breakbulk terminals or airports, depending on the long haul mode.

Items traveling by air are delivered to the nearest airport for an evening flight to the main hub. Aircraft may stop more than once to/from the hub to increase aircraft loads and maintain daily frequencies. Items typically arrive at the main hub between 10 pm and 2 am, where they are sorted by destination airport and loaded onto aircraft for a morning departure. After arriving at the destination airport, the ground process is reversed: items travel to a consolidation terminal and then to their final destination.

The long haul ground system includes several breakbulk terminals, which, like airports, act as gateways to the long haul network. However, unlike the air network, there is no single main hub. All breakbulk terminals serve as hubs, albeit for smaller percentages of the total network volume. Items are routed from originating consolidation terminal to destination consolidation terminal through at most two breakbulk terminals.

In non-integrated delivery networks, express items are transported by air for long haul trips due to restrictive time constraints.[1] Deferred items are sent over ground long haul networks. The labels A and G are used to identify network type. The air network is the set of links and terminals used by items traveling by air, including the ground portion of their travel. The ground network is the set of links and terminals used by items not traveling by air. The package network design problem determines ground and air network configurations (number and location of terminals for all terminal types) and routing guidelines for both items and vehicles.

The set of distribution levels is: $\mathcal{L} = \{0, 1, 2\}$: local (0), access (1) and long haul (2). The set of route directions is $\mathcal{B} = \{i, o\}$ for trips inbound to and outbound from a terminal. The set of terminal types is $\mathcal{T} = \{C, B, P, H\}$ for consolidation terminals (C), breakbulk terminals (B), airports (P), and main air hub (H). In the discussion here, two vehicle types are assumed for simplicity, $\mathcal{V} = \{a, t\}$, for air and truck. However, numerical results are generated assuming three truck types: the largest trucks serve long haul routes, smaller trucks serve access routes, and delivery vans serve local routes. The following cost parameters are used:

c_d^u costs of overcoming distance, for vehicle of type $u \in \mathcal{V}$ (\$/*distance*)

$c_d^{\prime u}$ marginal transportation cost per item, for vehicle of type $u \in \mathcal{V}$ (\$/*item*trip*)

c_q^u cost of stopping a vehicle of type $u \in \mathcal{V}$ at a terminal or customer (\$/*stop*)

c_f^y annualized fixed terminal cost of terminals of type $y \in \mathcal{T}$ (\$/*time*)

$c_f^{\prime y}$ annualized variable terminal cost of terminals of type $y \in \mathcal{T}$ (\$/*item*time*)

c_k sorting cost (\$/*item*bit*); i.e, number of bits required to identify a sorting class (2^n classes= n bits)

c_h storage (rent) cost for items (\$/*item*time*)

For simplicity of illustration, facilities are assumed to have the same costs and the superscript $y \in \mathcal{T}$ for c_f and c_f' is dropped. However, these costs do vary by facility type in implementation, and may even vary by location. The expressions for sorting cost assume that the sorting operation has been optimized, and therefore that the unit sorting cost is proportional to the logarithm of the number of classes (the bits).[2] The cost parameter c_k is the proportionality constant.

Three types of distribution systems are studied:

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

[2]It is well known in computer science (for the sorting of information) that optimum schemes are hierarchical, and that their complexity increases logarithmically with the number of classes; i.e., that the unit cost is proportional to the number of bits. Because the manipulation of physical items is qualitatively similar to the manipulation of data, we assume that the unit sorting costs of well run terminals is proportional to the logarithm of the number of classes. This result can be proven formally under fairly mild assumptions, but this is beyond the scope of this paper.

Non-integrated networks: segregated facilities, segregated routing

Non-integrated networks are simply the superposition of two separate networks offering express and deferred service independently.

Semi-integrated networks: integrated facilities, segregated routing

In this form of operation, all consolidation terminals may be used by all service types, but routing is performed separately for the two service types at all levels. The main advantage of this form of integration is that fewer consolidation terminals are needed to provide the same coverage level.

Fully integrated networks: integrated facilities and routing

In this form of operation, routing too is integrated. Express items are transported by air, but deferred items can travel either way. Flexible routing allows excess aircraft capacity to be filled with deferred items, and this reduces the ground transportation needs. In addition, significant savings in local transportation costs are possible through economies of density.

In this paper, formulae are developed for semi-integrated and fully integrated networks. Non-integrated networks are not explicitly considered since these networks can be modeled as two semi-integrated networks, each of only one demand type. Next this combined location and routing problem with multiple terminal and vehicle types is solved using a continuous approximation.

1.1 Continuous approximation

This section introduces the notation necessary to describe a generic problem and derive its solution using continuous approximations. Because much notation is needed, a summary is provided in the appendix. The structure of the distribution network over a service area $\mathcal{A}$ is defined by the network topology, and by demand, level of service, and cost data functions (parameters) that may vary with the coordinates, x, of points on the plane. The solution is also described in terms of decision functions of location (variables).

Exact customer location data are averaged over a region and replaced with spatial customer densities for service level, $\delta^s(x)$ *(customers/unit area)*. The origin-destination demand is averaged over a region as well and replaced with a temporal demand rate $\lambda^s(x^o, x^i)$ from a region of unit area about x^o to a region of unit area about x^i for service level $s \in \mathcal{S}$ *(items/area*^{2}*time). The following auxiliary functions are defined:

$\lambda_i^s(x)$ trip attraction rate in a region of unit area about x *(items/unit area*time)*; $\lambda_i^s(x) = \int_{x^o \in \mathcal{A}} \lambda^s(x^o, x)dx^o$, for $s \in \mathcal{S}$

$\lambda_o^s(x)$ trip generation rate about x *(items/unit area*time)*; $\lambda_o^s(x) = \int_{x^i \in \mathcal{A}} \lambda^s(x, x^i)dx^i$, for $s \in \mathcal{S}$

$\lambda_b^A(x)$ directional air network demand, $\lambda_b^A(x) = \lambda_b^E(x) + \omega_b(x)\lambda_b^D(x)$, for $b \in \mathcal{B}$

$\lambda_b^G(x)$ directional ground network demand, $\lambda_b^G(x) = (1 - \omega_b(x))\lambda_b^D(x)$, for $b \in \mathcal{B}$

$\lambda_T^m(x)$ bidirectional network-specific demand, $\lambda_T^m(x) = \sum_{b \in \mathcal{B}} \lambda_b^m(x)$, for $m = A, G$

$\lambda_b(x)$ directional demand for combined networks, $\lambda_b(x) = \sum_{m=A,G} \lambda_b^m(x)$, for $b \in \mathcal{B}$

$\lambda_T(x)$ bidirectional demand for combined networks, $\lambda_T(x) = \sum_{b \in \mathcal{B}} \lambda_b(x)$

$\delta(x)$ total customer density for combined networks, $\delta(x) = \sum_{s \in \mathcal{S}} \delta^s(x)$

Discrete location and routing decision variables are replaced with continuous decision functions. These decision functions are sufficient to develop a network with a well-defined cost, as explained in Daganzo and Newell (1986). Rather than specifying the exact number of terminals of type $y \in \mathcal{T}$ within a region, a density of terminals, $\Delta_y(x)$ (*terminals/unit area*), is used. The network design is obtained by partitioning the service region into "round" service regions of approximate size $\Delta_y(x)^{-1}$ and locating terminals at their centroids.

Item and vehicle routing decisions are approximated as well. The headway of a route of type $l \in \mathcal{L}$ for network $m = A, G$ in direction $b \in \mathcal{B}$ is $h_l^{m,b}(x)$. Headways are time intervals between consecutive dispatches. As such they indicate how often a route is run. The number of stops on the route is $n_l^{m,b}(x)$. At each stop, a vehicle picks up a shipment of size $v_l^{m,b}(x)$. The average linehaul distance of the route is $r_l^m(x)$, representing the length of the route minus detours for customer visits. The terminal densities dictate the number of customers served from one terminal. The hierarchical package delivery network structure ensures that a customer is served by only one terminal of each type. This information, along with route headways, number of stops and shipments sizes, provides sufficient detail to determine fleet size. The fraction of deferred items sent by air for long haul transportation in direction $b \in \mathcal{B}$ is $\omega_b(x)$.

A series of service level parameters enforces express and deferred deadlines. The maximum headway length for a route of type $l \in \mathcal{L}$ is H_l^m for $m = A, G$. Tight restrictions on express item delivery force $H_l^A \leq 1$ day. Headways in the ground network may be longer, yet item storage costs will prevent headways from becoming excessive. The maximum number of stops on a route of type $l \in \mathcal{L}$ is N_l^m may also vary by network type. To meet air time restrictions, the term ρ represents the maximum airport service radius.

2 Logistic cost functions

The path of a typical item from origin to destination is illustrated in Figure 2. As explained in Section 1, all items travel from an origin, to the closest consolidation terminal, to the long haul network via the nearest airport or breakbulk terminal, and then the process is reversed on the way to the final destination. Since no step is skipped in this hierarchical scheme, operating costs can be neatly separated by distribution level and terminals visited. For a specific pair of origin-destination regions, the average cost per item, z, is comprised of the following components:

$$z = z_{local} + z_{access} + z_{longhaul} + z_{reposition} + z_{CT} + z_{airport} + z_{BBT} + z_{hub} \tag{1}$$

Equation (1) contains transportation costs for each distribution level: z_{local}, z_{access}, and $z_{longhaul}$; terminal costs: z_{CT}, $z_{airport}$, z_{BBT}, and z_{hub}; and vehicle repositioning costs: $z_{reposition}$. Formulae for these components are developed in the subsections that follow (2.1 - 2.5). The sum (integral) of z across all items is an approximation for the total system cost, which is given in Section 2.6.

2.1 Local transportation costs

Local transportation costs cover pickup and delivery costs between origins/destinations and consolidation terminals. In the morning, delivery vehicles depart from a consolidation terminal and complete their deliveries. Vehicles that will be used for pick-up tours in the afternoon are then repositioned without returning to the terminal, and the rest return. In the afternoon, pick-up tours are conducted and then vehicles return to the consolidation terminal. This section only accounts for the delivery and pick-up costs, assuming that pickup and delivery routes are designed independently. The repositioning costs are covered in Section 2.4. Figure 3 illustrates the local distribution process for (a) semi-integrated networks and (b) fully integrated networks.

The models presented in this and other subsections are based on earlier work on cost estimates for the vehicle routing problem (VRP), see Daganzo (1999). It is assumed that every customer is visited on a tour. The VRP cost per item of one tour for items that are delivered in batches of size v on routes making n stops to customers of density δ and r distance units away from a depot is approximated by the following function:

$$f(r, v, n, \delta) = c'_d + \frac{rc_d + c_q}{nv} + \left(\frac{n-1}{n}\right)\frac{c_d k(\delta)^{-\frac{1}{2}} + c_q}{v} \tag{2}$$

where k is a constant dependent on the distance metric; $k \approx 0.8$ for grids.

The first component represents a per-item cost of delivery. The second term represents the linehaul cost component: travel from the depot to the customer region and a stop at the depot. The last term represents the local detour cost component: travel between customers in the region and stops at each customer.

Using expression (2), the local transportation costs is expressed on a cost per item basis as a function of decision variables $r_0^m(x)$, $v_0^{m,b}(x)$, and $n_0^{m,b}(x)$ and parameter $\delta^s(x)$ for each routing direction b and network type m. The optimization problem for local transportation is then:

$$\min z_{local}^{m,b}(x) = f(r_0^m(x), v_0^{m,b}(x), n_0^{m,b}(x), \delta^s(x)) \tag{3a}$$

subject to:

$$n_0^{m,b}(x)v_0^{m,b}(x) \le V_0 \qquad \textit{(vehicle capacity)} \tag{3b}$$

$$1 \le n_0^{m,b}(x) \le N_0^m \qquad \textit{(restriction on number of stops)} \tag{3c}$$

$$h_0^{m,b}(x) \le H_0^m \qquad \textit{(restriction on headways)} \tag{3d}$$

$$v_0^{m,b}(x) = \frac{\lambda_b^s(x)}{\delta^s(x)} h_0^{m,b}(x) \qquad \textit{(definitional: Little's formula)} \tag{3e}$$

$$r_0^m(x) = \frac{2}{3}(\pi\Delta_C(x))^{-\frac{1}{2}} \qquad \textit{(definitional: geometry)} \tag{3f}$$

$$r_0^m(x), v_0^{m,b}(x), h_0^{m,b}(x) > 0 \qquad \textit{(strictly positive)} \tag{3g}$$

Equation (3b) ensures vehicle capacity. Equation (3c) ensures that routes have at least one stop, and prohibits long routes. Its upper bound is used instead of a time constraint on the length of a shift to avoid the introduction of more notation.[3] Equation (3d) ensures that customers are visited with a minimum frequency. Equation (3e) expresses the dependence between shipment size and headway for constant demand. Equations (3d) and (3e), combined, constrain the shipment sizes in the objective function. Equation (3f) expresses the dependence between linehaul distance and terminal density, assuming that each terminal serves a region that is approximately circular and the terminal is located at the center of that region. This assumption then allows one to estimate the average distance from a customer to the depot as 2/3 of the radius of the circular region. This estimate is on the low side, but quite accurate if the terminals are arranged on a lattice.

With semi-integrated networks, four copies of $z_{local}^{m,b}(x)$ appear in expression (1). With fully integrated networks, only two copies of $z_{local}^{b}(x)$ (one for each direction) appear. The same is true of constraints.

2.2 Access transportation costs

Access tours between consolidation terminals and breakbulk terminals or airports are similar to local tours. Again, the VRP approximation is used. Access costs are expressed on an average cost per item basis, as a function of $r_1^m(x)$, $v_1^{m,b}(x)$, $n_1^{m,b}(x)$, and $\Delta_C(x)$.

$$\min z_{access}^{m,b}(x) = f(r_1^m(x), v_1^{m,b}(x), n_1^{m,b}(x), \Delta_C(x)) \tag{4a}$$

[3] Since it is often the number of stops that limits the number of items in a vehicle, a rough approximation of the average size of an item is sufficient.

subject to:

$$n_1^{m,b}(x)v_1^{m,b}(x) \leq V_1 \tag{4b}$$

$$1 \leq n_1^{m,b}(x) \leq N_1^m \tag{4c}$$

$$h_1^{m,b}(x) \leq H_1^m \tag{4d}$$

$$v_1^{m,b}(x) = \frac{\lambda_b^m(x)}{\Delta_C(x)} h_1^{m,b}(x) \tag{4e}$$

$$r_1^G(x) = \frac{2}{3}(\pi\Delta_B(x))^{-\frac{1}{2}} \qquad r_1^A(x) = \frac{2}{3}(\pi\Delta_P(x))^{-\frac{1}{2}} \tag{4f}$$

$$r_1^m(x), v_1^{m,b}(x), h_1^{m,b}(x) > 0 \tag{4g}$$

For fully integrated networks, one must specify the network demand rates that appear in (4e) with another equation since the network demand rates no longer equal the service level demand rates. Recall that $\lambda_b^A(x) = \lambda_b^E(x) + \omega_b(x)\lambda_b^D(x)$ and $\lambda_b^G(x) = (1-\omega_b(x))\lambda_b^D(x)$ for $b \in \mathcal{B}$, where $\omega_b(x) \geq 0$. Excess aircraft capacity determines the values of $\omega_b(x)$ and this is discussed next.

2.3 Long haul transportation costs

2.3.1 Air network

Since all items traveling by air are served through one main hub, the problem decomposes into a many-to-one distribution problem inbound to the hub, and a similar one-to-many problem in the outbound direction. In both inbound and outbound directions, we use the VRP approximation of one depot (the air hub) serving several customers (the airports). Operating headways are restricted to one day ($h_2^A = \tilde{h} = 1$ day), and are not decision variables here. The average linehaul distance, $r_2^A(x)$ is simply the distance from x to the hub which depends on the location of the main hub. As shown in Smilowitz (2001), it may be inefficient to operate a symmetric air network (inbound trips to a region mirror outbound trips from that region). Thus, inbound and outbound long haul trips are modeled separately.

$$\min z_{longhaul}^{A,b}(x) = f(r_2^A(x), v_2^{A,b}(x), n_2^{A,b}(x), \Delta_P(x)) \tag{5a}$$

subject to:

$$n_2^{A,b}(x)v_2^{A,b}(x) \leq V_2^A \tag{5b}$$

$$1 \leq n_2^{A,b}(x) \leq N_2^A \tag{5c}$$

$$v_2^{A,b}(x) = \frac{\lambda_b^A(x)}{\Delta_P(x)}\tilde{h} \tag{5d}$$

$$v_2^{A,b}(x) > 0 \tag{5e}$$

$$\Delta_P(x) \geq \frac{1}{\rho^2\pi} \tag{5f}$$

A new constraint on the density of airports (5f) is added to ensure that the service radius from an airport does not exceed a maximum distance ρ that guarantees the timely completion of access and local tours.

For semi-integrated networks, expressions (5) are used only for express items, $\lambda_b^A(x) \equiv \lambda_b^E(x)$. With integrated routing, a fraction of deferred items may travel by air, provided excess capacity exists. Constraint (5g) is added to restrict the fractional amount shifted $\omega_b(x)$ by the available capacity.

$$\frac{n_2^{A,b}(x)\tilde{h}}{\Delta_P(x)}\left(\nu\omega_b(x)\lambda_b^D(x) + \lambda_b^E(x)\right) \leq V_2^A, \qquad \text{for } \nu \geq 1, \omega_b(x) \leq 1 \tag{5g}$$

The constant ν is introduced to recognize that it is not economical to fill aircraft with deferred items to the same capacity level as with more profitable express items since operating costs increase with the load carried.

2.3.2 Ground network

The ground network contains multiple breakbulk terminals; therefore, the problem cannot be decomposed in the same manner. Fortunately, continuous approximation models for many-to-many non-integrated systems with breakbulk terminals have been developed. It has been shown (see Daganzo (1999)) that the vehicle distance traveled between breakbulk terminals can be easily estimated, without specifying the exact routing of items, when it can be assumed that vehicles travel full. As the number of terminals in the service region ($\int_{x\in\mathcal{A}} \Delta_B(x)dx$) increases, the linehaul component of this distance rapidly approaches the ratio of the total item-miles demanded and the vehicle capacity. Therefore, the detour component associated with multiple stops is estimated by:

$$\left(\frac{n_2^G(x)-1}{n_2^G(x)}\right)\frac{c_d k(\Delta_B(x))^{-\frac{1}{2}}+c_q}{v_2^G(x)} \tag{6a}$$

Definitional constraints are approximated as follows. The average shipment size collected from a breakbulk terminal at x for destination x^i is $\frac{\lambda^G(x,x^i)h_2^G(x)}{\Delta_B(x)\Delta_B(x^i)}$. Averaged across destinations, this may be approximated by

$$v_2^G(x) \approx \frac{\bar{\lambda}^G h_2^G(x)}{\bar{\Delta}_B^2} \tag{6b}$$

The average number of collecting stops per vehicle trip at or around x is

$$n_2^G(x) \approx \frac{V_2^G}{v_2^G(x)} \approx \frac{V_2^G\bar{\Delta}_B^2}{\bar{\lambda}^G h_2^G(x)} \tag{6c}$$

where

$\bar{\lambda}^G = \int_{x^o\in\mathcal{A}}\int_{x^i\in\mathcal{A}} \frac{\lambda^G(x^o,x^i)}{|\mathcal{A}|^2}dx^i dx^o$, the average ground network demand rate for the entire service region across all origins and destinations

$\bar{\Delta}_B = \int_{x\in\mathcal{A}} \frac{\Delta_B(x)}{|\mathcal{A}|}dx$, the average breakbulk terminal density

2.4 Vehicle repositioning costs

2.4.1 Local and access levels

On local tours, the number of vehicles dispatched for morning deliveries may be insufficient to cover afternoon pick-up; extra empty vehicles must be deployed for collection. Conversely, vehicles may return empty to the consolidation terminal after morning distribution if inbound demand exceeds outbound demand. The same is true for access trips. It is assumed that all local and access tours operate from one terminal. Thus, the number of repositioning trips is simply the number of vehicles needed to serve the demand imbalance, $|\lambda_o^m(x)-\lambda_i^m(x)|$. For vehicles with capacity V operating in a region of terminals with density $\Delta(x)$, the number of trips is $\frac{|\lambda_o^m(x)-\lambda_i^m(x)|}{\Delta(x)V}$.

The total repositioning cost in that region is therefore equal to the number of trips multiplied by the distance cost and the average distance from the terminal to the customers, $r(x)$. Prorating this cost the total $\frac{\lambda_T^m(x)}{\Delta(x)}$ items served by the terminal, and replacing $r(x)$ with $\frac{2}{3}(\pi\Delta(x))^{-\frac{1}{2}}$, we obtain the average cost per item:

$$\frac{\frac{2}{3}c_d|\lambda_o^m(x)-\lambda_i^m(x)|}{\lambda_T^m(x)V}(\pi\Delta(x))^{-\frac{1}{2}} \tag{7}$$

2.4.2 Long haul level

The repositioning of empty tractor trailers between breakbulk terminals is more difficult to model since demand imbalances between breakbulk terminals require the repositioning of vehicles between terminals. If demand is perfectly homogenous, this repositioning term between breakbulk terminals is zero. If demand is homogeneous with random variations, this term is modeled as a transportation problem as shown in Daganzo and Smilowitz (2000). In this case the total cost only depends on one decision variable (the number of terminals), and the dependence is very weak. This reference also shows that if there are systematic imbalances, the added repositioning costs due to the imbalances can bound tightly from above by a function of the O-D table, again independent of all our decision variables. Since repositioning costs are a very small part of the total cost of our problem (less then 1%), and since the effect of our decision variables on this cost is very small, a long haul repositioning costs term is not included in the optimization phase.

2.5 Terminal costs

Terminal costs consist of handling costs, facility charges, sorting expenses, and storage fees. The value of decision variables and parameters vary across terminal types, but the functional form of the terminal cost is the same. This section introduces the generic cost model; specific costs for each terminal type are included in Section 2.6. For further detail see Smilowitz (2001).

Consider a terminal serving a total inbound and outbound flow of $Q(x) = Q^i(x) + Q^o(\text{x})$ items per unit time where two sorts are performed daily, one sort for items traveling outbound from the region served by the terminal and one sort for items inbound to that region. The number of classifications in a sort, $K_b(x), b \in i, o$, is assumed to be the number of direct destinations connected with the terminal in a given direction. Thus, sorting costs are assumed to increase logarithmically with $K_b(x)$. The cost per item for a generic terminal with both inbound and outbound sorts is

$$g\big(Q(x), K_i(x), K_o(x), h_o(x), h_i(x)\big) = c_f' + \frac{c_f}{Q(x)} + \sum_{b \in \mathcal{B}} c_k \frac{Q^b(x)}{Q(x)} \log(K_b(x)) + \sum_{b=i,o} c_h h_b(x) \quad (8)$$

This expression includes a marginal cost per item through the terminal, c_f', representing the handling costs, and a fixed cost per terminal per unit time c_f, which is prorated to the items served in a time unit, $Q(x)$. This flow is given by the trip attraction and trip generation, $Q(x) = \frac{\lambda_i^m(x)}{\Delta_y(x)} + \frac{\lambda_o^m(x)}{\Delta_y(x)}$. The expression also includes a sorting complexity term that increases with the direction-dependent flow ($Q^b(x)$), and a storage cost dependent on the length of time an item is held at a terminal. Equation (8) assumes that this length of time is proportional to the routing headways, $h_b(x)$.

2.6 Complete model

A complete logistic cost function, containing all transportation and terminal costs per unit time, is used to obtain optimal designs and to compare the different scenarios. To obtain this function, the cost components described in the previous sections should be integrated over all items in the service area. The expression below is the result for a fully integrated network. By setting $\omega_b(x) = 0, \forall x \in \mathcal{A}, b \in \mathcal{B}$ and separating local costs by network, the expression represents a semi-integrated network. While the complete model may appear difficult to optimize, the problem is reduced in Section 3 to a series of subproblems that can be easily programmed into a spreadsheet. The logistic cost function is:

$$
\begin{aligned}
\min z = \int_{x\in\mathcal{A}} \Bigg\{ & \sum_{b\in\mathcal{B}} \lambda_b(x)\left(c_d^{'t} + \frac{r_0(x)c_d^t + c_q^t}{n_0^b(x)v_0^b(x)} + \left(\frac{n_0^b(x)-1}{n_0^b(x)}\right)\frac{c_d^t k(\delta(x))^{-\frac{1}{2}} + c_q^t}{v_0^b(x)}\right) \\
& + \sum_{b\in\mathcal{B}}\sum_{m=A,G} \lambda_b^m(x)\left(c_d^{'t} + \frac{r_1^m(x)c_d^t + c_q^t}{n_1^{m,b}(x)v_1^{m,b}(x)} + \left(\frac{n_1^{m,b}(x)-1}{n_1^{m,b}(x)}\right)\frac{c_d^t k(\Delta_C(x))^{-\frac{1}{2}} + c_q^t}{v_1^{m,b}(x)}\right) \\
& + \sum_{b\in\mathcal{B}} \lambda_b^A(x)\left(c_d^{'a} + \frac{r_2^A(x)c_d^a + c_q^a}{n_2^{A,b}(x)v_2^{A,b}(x)} + \left(\frac{n_2^{A,b}(x)-1}{n_2^{A,b}(x)}\right)\frac{c_d^a k(\Delta_P(x))^{-\frac{1}{2}} + c_q^a}{v_2^{A,b}(x)}\right) \\
& + \lambda_o^G(x)\left(c_d^{'t} + \left(\frac{n_2^G(x)-1}{n_2^G(x)}\right)\frac{c_d^t k(\Delta_B(x))^{-\frac{1}{2}} + c_q^t}{v_2^G(x)}\right) + \frac{c_d\bar{d}}{V_2^G}(\bar{\lambda}^G) \\
& + \frac{\frac{2}{3}|\lambda_o(x)-\lambda_i(x)|}{V_0\sqrt{\pi\Delta_C(x)}}c_d + \frac{\frac{2}{3}|\lambda_o^A(x)-\lambda_i^A(x)|}{V_1\sqrt{\pi\Delta_P(x)}}c_d + \frac{\frac{2}{3}|\lambda_o^G(x)-\lambda_i^G(x)|}{V_1\sqrt{\pi\Delta_B(x)}}c_d \\
& + \lambda_T(x)c_f' + \Delta_C(x)c_f + \lambda_i(x)c_k\log\left(\frac{\delta(x)}{\Delta_C(x)V_0}\right) + \lambda_o(x)c_k\log(2) + \sum_{b\in\mathcal{B}} c_h\lambda_b(x)h_0^b(x) \\
& + \lambda_T^G(x)c_f' + \Delta_B(x)c_f + \lambda_i^G(x)c_k\log\left(\frac{\Delta_C(x)}{\Delta_B(x)}\right) + \lambda_o^G(x)c_k\log(\bar{\Delta}_B|\mathcal{A}|) \\
& + \lambda_T^A(x)c_f' + \Delta_P(x)c_f + \lambda_i^A(x)c_k\log\left(\frac{\Delta_C(x)}{\Delta_P(x)}\right) + \lambda_i^A(x)c_k\log(\bar{\Delta}_P|\mathcal{A}|) \\
& + \sum_{b\in\mathcal{B}} c_h\lambda_b^G(x)h_1^{G,b}(x) + \lambda_o^G(x)c_h h_2^G(x) + \sum_{b\in\mathcal{B}} c_h\lambda_b^A(x)(\tilde{h} + h_1^{A,b}(x))\Bigg\}dx \qquad (9)
\end{aligned}
$$

The integrand of (9) begins with local transportation costs, summing both collection and delivery costs. The next line represents access costs for trips to and from airports and breakbulk terminals. The following two lines represent long haul costs for air and ground transportation, respectively. The next line includes the repositioning costs for local and access vehicles. The final four lines include terminal costs. The goal is to choose the decision functions that minimize (9) subject to constraints defined in the previous subsections. This continuous optimization problem (COP) is examined in the next section.

3 Optimization

This section describes the reduction of problem COP to a series of subproblems that can be solved in closed form. Section 3.1 shows how the decision functions for the number of stops, the amount of deferred items shifted to air, the shipment sizes, and the linehaul distances can be eliminated from the model. Consequently, the entire model can be written as a function of only the terminal densities and operating headways. Section 3.2 presents the subproblem decomposition.

3.1 Elimination of variables

Since in-vehicle inventory costs are not considered, ground vehicles should make as many stops as possible; see the full vehicle theorem in Daganzo (1999). Therefore ground vehicles will either reach their capacity or their maximum number of stops; i.e., either (3b) or (3c) should be binding (on the upper side) and we set $n_l^{m,b}(x) = \min\left\{N_l, \frac{V_l^m}{v_l^{m,b}(x)}\right\}$, removing $n_l^{m,b}(x)$ from the formulation. For the long haul air routes nothing needs to be eliminated since the number of stops is assumed to be exogenously given. As explained in Smilowitz (2001), this number is usually one or two, depending on location (including time zone change relative to the hub) and demand level.

In all scenarios, it is assumed that the long haul air network is optimally configured for express items only (i.e., the location of airports is determined by express demand only and this, in turn, specifies the excess capacity available). It is further assumed that in fully integrated scenarios the maximum amount of deferred items is shifted to the air network. This is reasonable for large ground networks, as confirmed in Smilowitz *et al.* (2002). The amount of deferred items shifted $\omega_b(x)$ is then eliminated as a decision variable, and $\lambda_b^G(x)$ and $\lambda_b^A(x)$ are treated as input data. These variables are dependent on the level of integration, i.e., $\omega_b(x)$ should be as large as allowed by (5g).

The shipment size $v(x)$ is replaced using definitional constraints (3e, 4e, and 5d). Likewise, linehaul distance $r_l^m(x)$ is replaced using constraints (3f) and (4f). Thus, the only remaining variables in the complete cost model are the terminal densities and operating headways, as claimed.

3.2 Subproblem decomposition

Note from (9) the logarithmic sorting terms involving $\Delta_C(x)$ cancel each other out if terminals have the same sorting cost c_k. The logarithmic sorting terms involving $\Delta_B(x)$ and $\Delta_P(x)$ approximately cancel out as well. Since the number of breakbulk terminals is not expected to change significantly between large geographic regions as a result of the optimization and the log function mutes small relative differences, $\bar{\Delta}_B$ and $\bar{\Delta}_P$ can be accurately approximated by $\Delta_B(x)$ and $\Delta_P(x)$. This approximation will systematically underestimate costs by Jensen's inequality. Fortunately, as the area of the service region increases, this underestimation decreases. It should be quite small for a system spanning the United States. Further, if $\lambda_i^G(x) \approx \lambda_o^G(x)$ then the log terms involving $\Delta_B(x)$ should cancel out; e.g., $\lambda_i^G(x) c_k \log\left(\frac{1}{\Delta_B(x)}\right) + \lambda_o^G(x) c_k \log(\Delta_B(x)) \approx 0$. Unlike the first approximation, the direction of the error caused by this approximation is not known a priori. However, since $\bar{\lambda}_i^G(x) = \bar{\lambda}_o^G(x)$, it is likely quite small. As shown in Smilowitz (2001), the maximum error from these approximations for large problems with freight imbalances is under 2% of sorting costs, which are a small part of the total cost.

With the changes of sections 3.1 and 3.2, the complete model can be written in a compact form that highlights the decision variables. Expressions for the variable coefficients and independent constant Π are given in the appendix. For further economy of notation, their dependence on x is not explicitly included . The complete model is:

$$\min z = \int_{x\in\mathcal{A}} \Bigg\{ \sum_{b=i,o} \Big(\alpha_1^b h_0^b(x) + \alpha_2 (h_0^b(x))^{-1} \Big) + \beta_1 \Delta_C(x)^{-\frac{1}{2}}$$

$$+ \sum_{b=i,o} \sum_{m=A,G} \left(\beta_2 \frac{\sqrt{\Delta_C(x)}}{h_1^{m,b}(x)} + \beta_3 \frac{\Delta_C(x)}{h_1^{m,b}(x)} + \beta_4^b h_1^{m,b}(x) \right) + \beta_6 \Delta_C(x) + \chi_1 \Delta_P^{-\frac{1}{2}}(x) + \chi_2 \Delta_P(x) + \chi_3 \Delta_P^{\frac{1}{2}}(x)$$

$$+ \kappa_1 \Delta_B^{-\frac{1}{2}}(x) + \kappa_2 \Delta_B(x) + \kappa_3 \frac{\Delta_B(x)^{\frac{3}{2}}}{h_2^G(x)} + \kappa_4 \frac{\Delta_B(x)^2}{h_2^G(x)} + \kappa_5 h_2^G(x) + \Pi \Bigg\} dx \quad (10a)$$

subject to:

$$\frac{\lambda_b(x)\delta(x)}{N_0 V_0} \le h_0^b(x) \le \frac{\lambda_b(x)\delta(x)}{V^0} \qquad \forall b \in \mathcal{B} \quad (10b)$$

$$\frac{\lambda_b^m(x)}{V_1} \le \frac{\Delta_C(x)}{h_1^{m,b}(x)} \le \frac{N_1^m \lambda_b^m(x)}{V_1} \qquad \forall b \in \mathcal{B}; m = A, G \quad (10c)$$

$$\frac{\lambda_b^A(x)}{V_2^A} \le \Delta_P(x) \le \frac{N_2^A \lambda_b^A(x)}{V_2^A} \qquad \forall b \in \mathcal{B} \quad (10d)$$

$$\frac{\bar{\lambda}^G}{V_2^G} \le \frac{\Delta_B(x)^2}{h_2^G(x)} \le \frac{N_2^G \bar{\lambda}^G}{V_2^G} \quad (10e)$$

$$0 < h_l^{m,b}(x) \le H_l^m \qquad \forall b \in \mathcal{B}; m = A, G; l \in \mathcal{L} \quad (10f)$$

$$\Delta_P(x) \ge \frac{1}{\rho^2 \pi} \quad (10g)$$

This formulation contains a significantly smaller set of decision variables. With the changes, there are fewer constraints as well. It is clear from (10) that the problem decomposes by sets of decision functions. There are five distinct groups of functions that are not linked to each other either in the objective function or the constraints. The total cost model is separated into five subproblems that determine the following variables:

- 1^o local outbound headways, $h_0^o(x)$
- 1^i local inbound headways, $h_0^i(x)$
- 2 consolidation terminal densities and access headways, $\Delta_C(x), h_1^{m,b}(x)$
- 3 airport densities, $\Delta_P(x)$
- 4 breakbulk terminal densities and long haul ground headways, $\Delta_B(x), h_2^G(x)$

3.2.1 Subproblems 1^o and 1^i: local headways

We start with the first two subproblems which are the easiest to solve and help to introduce more complicated subproblems later. For $b = i, o$, the subproblems are:

$$\min z_{1^b} = \int_{x\in\mathcal{A}} \left(\alpha_1^b h_0^b(x) + \frac{\alpha_2}{h_0^b(x)} \right) dx \quad (11a)$$

subject to:

$$\frac{\lambda_b(x)\delta(x)}{N_0 V_0} \le h_0^b(x) \le \frac{\lambda_b(x)\delta(x)}{V_0} \quad (11b)$$

$$0 < h_0^b(x) \le H_0 \quad (11c)$$

Note that the subproblems can be further decomposed by x because the integrand and the constraints are local in nature. Hence, one can simply minimize the integrand for every x and then sum across all such subdivisions of the total area. The results are easy to obtain because the integrand is a simple economic order quantity (EOQ) problem. The optimal headway is $h_0^b(x)^* = \sqrt{\frac{\alpha_2}{\alpha_1^b}}$, provided all constraints are met. Otherwise, the optimal solution will exist at one of the extreme points defined by the constraints. The solution should be intuitive. With higher transportation costs, headways should be lengthened, and with higher rent costs, headways should be shortened. We find that for reasonable values of the parameters, (11c) is binding at its upper bound.

3.2.2 Subproblem 2: consolidation terminal densities and access headways

Subproblem 2 is defined as:

$$\min z_2 = \int_{x\in\mathcal{A}} \Bigg(\beta_1 \Delta_C(x)^{-\frac{1}{2}} + \sum_{b=i,o}\sum_{m=A,G}\left(\beta_2 \frac{\sqrt{\Delta_C(x)}}{h_1^{m,b}(x)} + \beta_3 \frac{\Delta_C(x)}{h_1^{m,b}(x)} + \beta_4^b h_1^{m,b}(x)\right) + \beta_6 \Delta_C(x)\Bigg) dx \quad (12a)$$

subject to:

$$\frac{\lambda_b^m(x)}{V_1} \le \frac{\Delta_C(x)}{h_1^{m,b}(x)} \le \frac{N_1^m \lambda_b^m(x)}{V_1} \qquad b\in\mathcal{B}; m=A,G \quad (12b)$$

$$0 < h_1^{m,b}(x) \le H_1^m \qquad b\in\mathcal{B}; m=A,G \quad (12c)$$

As shown by (12), subproblem 2 can also be decomposed by location. The resulting problem is more complicated than subproblem 1 because it contains a non-convex objective function and non-linear constraints. However, the following changes of variable transform (12) into a convex problem with linear constraints: $w_C = \ln(\Delta_C(x))$, $w_{m,b} = \ln(h_1^{m,b}(x)), b \in \mathcal{B}; m = A, G$. The problem is then:

$$\min z_{2'} = \int_{x\in\mathcal{A}} \left(\beta_1 e^{-\frac{w_C}{2}} + \sum_{b=i,o}\sum_{m=A,G}\left(\beta_2 e^{\frac{w_C}{2}-w_{m,b}} + \beta_3 e^{w_C - w_{m,b}} + \beta_4^b e^{w_{m,b}}\right) + \beta_6 e^{w_C}\right) dx \quad (13a)$$

subject to:

$$\ln\left(\frac{\lambda_b^m(x)}{V_1}\right) \le w_C - w_{m,b} \le \ln\left(\frac{N_1^m \lambda_b^m(x)}{V_1}\right) \qquad b\in\mathcal{B}; m=A,G \quad (13b)$$

$$w_{m,b} \le \ln(H_1^m) \qquad b\in\mathcal{B}; m=A,G \quad (13c)$$

Since the transformed subproblem is convex, it can be solved with gradient search techniques. The same spatial decomposition and logarithmic transformation techniques reduce subproblems 3 and 4 to simple convex programs.

3.2.3 Subproblem 3: airport densities

Here the following change of variable is introduced: $w_P = \ln(\Delta_P(x))$. Subproblem 3 is then:

$$\min z_{3'} = \int_{x\in\mathcal{A}} \left(\chi_1 e^{-\frac{w_P}{2}} + \chi_2 e^{w_P} + \chi_3 e^{\frac{w_P}{2}}\right) dx \quad (14a)$$

subject to:

$$\ln\left(\frac{\lambda_b^A(x)}{V_2^A}\right) \le w_P \le \ln\left(\frac{N_2^A \lambda_b^A(x)}{V_2^A}\right) \qquad b \in \mathcal{B} \tag{14b}$$

$$w_2 \ge \ln(\frac{1}{\rho^2 \pi}) \tag{14c}$$

3.2.4 Subproblem 4: breakbulk terminal densities and long haul ground headways

Here the terminal density and headway variables are transformed as follows: $w_B = \ln(\Delta_B(x))$, $w_2 = \ln(h_2^G(x))$. This results in:

$$\min z_{4'} = \int_{x \in \mathcal{A}} \left(\kappa_1 e^{-\frac{w_B}{2}} + \kappa_2 e^{w_B} + \kappa_3 e^{\frac{3w_B}{2} - w_2} + \kappa_4 e^{2w_B - w_2} + \kappa_5 e^{w_2} \right) dx \tag{15a}$$

subject to:

$$\ln\left(\frac{\bar{\lambda}^G}{V_2^G}\right) \le 2w_B - w_2 \le \ln\left(\frac{N_2 \bar{\lambda}^G}{V_2^G}\right) \tag{15b}$$

$$w_2 \le \ln(H_2^G) \tag{15c}$$

This problem can again be decomposed by location and solved easily. The entire COP can be solved as a series of convex subproblems.

4 Results and concluding remarks

The design methodology introduced in this paper is employed in Smilowitz (2001), where near optimal network configurations are obtained for large package delivery networks roughly the size of the contiguous United States. Operating costs and statistics are derived from Kiesling (1995) and company literature. Population density is used as a proxy for package demand level and housing density as a proxy for customer density. The 1990 U.S. census includes population, housing counts, land area and geographic coordinates for all Metropolitan Statistical Areas (MSA) in the United States; see U.S. Census Bureau (1990). The largest MSA's are aggregated into groups of common demand and geographic features to form twenty subregions. The subregions are large enough to contain multiple terminals, yet small enough such that average network characteristics (demand levels, distances to main air hub, etc.) are representative of the entire subregion. The entire service region is 2,500,000 mi^2. The areas of the twenty subregions range from 16,383 mi^2 to 225,974 mi^2.

Importantly, the methodology is extended in Smilowitz (2001) to account for uncertainty in demand and for seasonal variations. The express air network should be designed for extreme cases of high demand. Hence, temporal demand variations for air services increase the availability of excess capacity at all levels of the express network. As a result, the benefits of integration have the potential to increase significantly in these cases. Lack of space prevents discussion here.

Also included in Smilowitz (2001) is more detailed model validation. That reference shows that cost models presented here are accurate enough for design purposes. The cost approximations for vehicle routing tours used here to model local and access tour costs have been successfully compared with costs obtained with simulated annealing techniques in Robuste *et al.* (1990). Recent work by Erera (2000) has shown that the costs of advanced local distribution strategies can be approximated within 5% of cost results from simulation. Costs from hub location decisions obtained with continuous approximations have also been validated against discrete cost models in the literature; see Campbell (1993). Numerical optimization techniques are used to validate long haul operating costs in Smilowitz (2001) and empty vehicle repositioning costs in Daganzo and Smilowitz (2000).

4.1 Test case results

Here we focus on the following question: given a pair of near optimal non-integrated networks for deferred and express demand, when does it make sense to integrate their operations to reduce costs? Both deterministic and random demands are considered. A series of test cases is studied varying the levels of deferred demand with the goal of determining the level of deferred demand required to justify integration. In Figure 4, the total savings achieved through integration is plotted as a percent of the total pre-integration cost of the original air network. The x-axis represents the average daily deferred demand and the figure indicates the point at which deferred demand exceeds express demand. An average demand of 1.8 million packages per day are assumed for express items. Deferred demand ranges from 0 (no integration) to 5.2 million packages per day. On the y-axis, the total (air and ground) network cost savings divided by the total pre-integration air network costs are plotted. Both deterministic and random demand scenarios are tested.

As the figure indicates, the benefits of integration are greater when deferred demand exceeds express demand. Express carriers may be reluctant to integrate operations with deferred carriers when the level of deferred demand is significantly less than that of express. However, savings grow quickly as deferred demand increases, even before deferred demand equals express demand. The growth rate of savings decreases and savings reach an asymptote since excess air capacity is filled and the maximum benefits of local transportation integration are realized. In addition, the figure reveals that savings increase significantly when demands are uncertain. Additional savings may be achieved when it is necessary to overcapacitate the air network for seasonal demand fluctuations.

This insight helps to explain the different business strategies of United Parcel Services (UPS) and Federal Express. UPS has adopted a more integrated strategy than Federal Express. A large deferred carrier such as UPS should realize greater cost savings from integration.

Another striking feature of the test cases is the dominance of local transportation costs. This should not be surprising since local transportation consists of many trips made in small vehicles operating on short headways. In turn, changes in local costs have a large impact on total cost. As expected, large savings in local transportation costs are realized with integrated routing as a result of higher customer density, see Figure 3(b) versus Figure 3(a). The total savings are greatest in regions of the service network where local transportation costs account for over 45% of total ground and air network operating costs. These regions typically have low demand levels and low customer densities. With the rise of e-commerce, the importance of local distribution to individual customers should increase and the incentive for integrating local distribution between service levels should increase too.

Additional analysis of test cases suggests that merging infrastructure from existing networks to form an integrated network yields cost savings comparable with designing an entirely new integrated network. Across all test cases, the largest difference in total cost between integration strategies with existing infrastructure and redesigning a network is only 0.5% which hardly justifies the cost of relocating or building facilities.

Of course, there are other costs and benefits to integration not considered here that could impact decisions. Integration gives carriers the ability to move deferred items quickly in response to routing problems in the ground network (weather, surge in demand, etc.). Further, overhead costs including administrative costs, sales costs, etc. can be reduced through integration when a delivery firm can use one office to multiple service levels. However, there may be additional costs of complexity involved with integration.

4.2 Conclusions

This research addresses significant gaps identified in the continuous approximation literature. The methodology introduced here is capable of modeling complex integrated distribution systems. The systems modeled in this research include multiple service levels and multiple transportation modes.

The continuous approximation cost functions used are capable of realistically modeling complex distribution systems. Distribution activities include multiple transshipments, peddling tours, and shipment choice. All key distribution costs (sorting, facility charges, and vehicle repositioning, as well as transportation and inventory) are included in these functions. This research demonstrates that complex problems can be reduced to a series of simple subproblems.

This research asks general systematic questions with the goal of formalizing the process of design, operation, and evaluation of complex integrated logistics systems. With the methodology developed, package delivery companies can be more proactive and explore a wider range of "what if" scenarios. The results presented in Section 4.1 provide valuable insight into real world applications.

This work is one step in exploring integrated logistics systems; there are many other scenarios to explore. One key advantage of continuous approximation models is the ability to model large systems. The modeling approach can be extended to examine the performance of multiple hub air networks, including multimodal hubs. Multimodal hubs would enable inbound and outbound transportation of deferred items to be modally decoupled at the hubs. Items traveling between a given origin/destination pair could then be served by a combination of modes with the mode transfer occuring at the main hub. This variation would provide greater flexibility to balance loads on aircraft into and out of the hub. Furthermore, other supply-chain models for e-commerce applications can be studied. Models can be used to consider how the performance of integrated networks affects the optimal configuration of on-line commerce firms such as web-based grocers and other types of retailers that sell and purchase multiple products with service delivery windows.

Results presented in Smilowitz (2001) highlight the importance of local distribution and the potential opportunities for savings from integration. The study of local distribution has been, and should continue to be, a rich area for research. Specific local distribution research for integrated networks could focus on more intelligent strategies that exploit the advantages of integration. For example, delivery vehicles could be loaded with both express and deferred items at consolidation terminals each morning for delivery. Express items with tight deadlines could be delivered first before deferred items. Afternoon item pickup could be performed in the reverse order. This would allow vehicles to make better use of capacity and more evenly distribute driver work shifts throughout the day.

References

Ahuja, R., Magnanti, T., and Orlin, J. (1993). *Network Flows: Theory, Algorithms and Applications.* Prentice-Hall, Inc., Englewood Cliffs, N.J.

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

Ball, M., Magnanti, T., Monma, C., and Nemhauser, G., editors (1995). *Network Models*, volume 7 of *Handbooks in Operations Research and Management Science.* Elsevier Science Publishing, New York.

Barnhart, C. and Schneur, R. R. (1996). Air network design for express shipment service. *Operations Research*, **44**(6), 852–863.

Campbell, J. F. (1993). Continuous and discrete demand hub location problems. *Transportation research B*, **27B**(6), 473–482.

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

Daganzo, C. F. (1987a). Modeling distribution problems with time windows: Part i. *Transportation Science*, **21**(3), 171–179.

Daganzo, C. F. (1987b). Modeling distribution problems with time windows: Part ii. *Transportation Science*, **21**(3), 180–187.

Daganzo, C. F. (1999). *Logistics Systems Analysis*. Springer, New York.

Daganzo, C. F. and Newell, G. F. (1986). Configuration of physical distribution networks. *Networks*, **16**, 113–132.

Daganzo, C. F. and Smilowitz, K. (2000). Asymptotic solutions to the balanced transportation problem of linear programming. Working paper, University of California at Berkeley.

Daskin, M. S. (1985). Logistics: An overview of the state of the art and perspectives on future research. *Transportation Research*, **19A**(5/6), 383–398.

Eilon, S., Watson-Gandy, C. T., and Christofides, N. (1971). *Distribution Management*. Griffin, London.

Erera, A. L. (2000). *Design of Logistics Systems for Uncertain Environments*. PhD dissertation, University of California, Berkeley, Institute of Transportation Studies.

Geoffrion, A. (1976). The purpose of mathematical programming is insight, not numbers. *Interfaces*, **7**(1), 81–92.

Hall, R. (1986). Discrete models / continuous models. *Omega, International Journal of Management Science*, **14**, 213–220.

Hall, R. (1989). Dispatching regular and express shipments between a supplier and manufacturer. *Transportation Research B*, **23B**(3), 195–211.

Hall, R. (1991). Characteristics of multi-stop / multi-terminal delivery routes with backhauls and unique items. *Transportation Research B*, **25B**(6), 391–403.

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

Jordan, W. and Burns, L. (1984). Truck backhauling on two terminal networks. *Transportation Research B*, **18B**(6), 487–503.

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

Langevin, A., Mbaraga, P., and Campbell, J. F. (1996). Continuous approximation models in freight distribution: An overview. *Transportation Research B*, **30B**(3), 163–188.

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

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

Newell, G. (1973). Scheduling, location, transportation, and continuum mechanics: some simple approximations to optimization problems. *SIAM, Journal of Applied Mathematics*, **25**, 346–360.

Powell, W. and Sheffi, Y. (1983). The load planning problem of motor carriers: Problem description and a proposed solution approach. *Transportation Research A*, **17A**(6), 471–480.

Robuste, F., Daganzo, C. F., and II, R. R. S. (1990). Implementing vehicle routing models. *Transportation Research B*, **24**(4), 263–286.

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

Smilowitz, K., Atamtürk, A., and Daganzo, C. F. (2002). Deferred item and vehicle routing within integrated networks. Submitted for publication in *Transportation Research. Part E, Logistics and Transportation Review.* Available at http://www.ce.berkeley.edu/Programs/Transportation/Daganzo/publications.html.

U.S. Census Bureau (1990). Tiger/geographic identification code scheme.

Appendix: Notation

Network sets

$\mathcal{S}$ Set of service levels, $\mathcal{S} = \{E, D\}$ for express and deferred items.

$\mathcal{L}$ Set of distribution levels, $\mathcal{L} = \{0, 1, 2\}$: local (0), access (1) and long haul (2).

$\mathcal{B}$ Set of route directions, $\mathcal{B} = \{i, o\}$ for trips inbound to and outbound from a terminal.

$\mathcal{V}$ Set of vehicle types, for simplicity $\mathcal{V} = \{a, t\}$, for air and truck.

$\mathcal{T}$ Set of terminal (node) types, $\mathcal{T} = \{C, B, P, H\}$ for consolidation terminals (C), breakbulk terminals (B), airports (P), and main air hub (H)

Demand Parameters

$\delta^s(x)$ spatial customer densities for service level $s \in \mathcal{S}$ *(customers/unit area)*

$\lambda^s(x^o, x^i)$ temporal demand rate from a region of unit area about x^o to a region of unit area about x^i for service level $s \in \mathcal{S}$ *(items/area*^{2}*time)

$\lambda^s_i(x)$ trip attraction rate in a region of unit area about x *(items/unit area*time)*; $\lambda^s_i(x) = \int_{x \in \mathcal{A}} \lambda^s(x, x^i)dx$

$\lambda^s_o(x)$ trip generation rate about x *(items/unit area*time)*; $\lambda^s_o(x) = \int_{x \in \mathcal{A}} \lambda^s(x^o, x)dx$

Level of Service Parameters

H_l maximum headway length for a route of type $l \in \mathcal{L}$ *(time)*

N_l maximum number of stops on a route of type $l \in \mathcal{L}$

V^u vehicle capacity for vehicle of type $u \in \mathcal{V}$ (*items*)

ρ maximum airport service radius *(distance)*

Cost Parameters

c^u_d costs of overcoming distance, for vehicle of type $u \in \mathcal{V}$ ($/*distance*)

c'^u_d marginal transportation cost per item, for vehicle of type $u \in \mathcal{V}$ ($/*item*trip*)

c^u_q cost of stopping a vehicle of type $u \in \mathcal{V}$ at a terminal or customer ($/*stop*)

c^y_f annualized fixed terminal cost of terminals of type $y \in \mathcal{T}$ ($/*time*)

$c_f^{'y}$ annualized variable terminal cost of terminals of type $y \in \mathcal{T}$ (\$/*item*time*)

c_k sorting cost (\$/*item*bit*); i.e, number of bits required to identify a sorting class, (2^n classes = n bits)

c_h storage (rent) cost for items (\$/*item*time*)

For simplicity of illustration, facilities are assumed to have the same costs and the superscript $y \in \mathcal{T}$ for c_f and c_f' is dropped.

Decision functions

$\Delta_y(x)$ density of terminals of type $y \in \mathcal{T}$ (*terminals/unit area*)

$h_l^{m,b}(x)$ headway of a route of type $l \in \mathcal{L}$ for network $m = A, G$ in direction $b \in \mathcal{B}$ (*time*)

$n_l^{m,b}(x)$ number of stops on a route of type $l \in \mathcal{L}$ for network $m = A, G$ in direction $b \in \mathcal{B}$

$v_l^{m,b}(x)$ shipment size per terminal on a route of type $l \in \mathcal{L}$ for network $m = A, G$ in direction $b \in \mathcal{B}$ *(items/terminal)*

$r_l^m(x)$ average linehaul distance on a route of type $l \in \mathcal{L}$ for network $m = A, G$ *(distance)*

$\omega_b(x)$ fraction of deferred items sent by air for long haul transportation in direction $b \in \mathcal{B}$

Coefficients and constants

Constant Π,

$$\begin{aligned}\Pi =& \lambda_T(x)\left(c_d^{t'} - \frac{c_d^t k(\delta(x))^{-\frac{1}{2}}}{V_0}\right) + \lambda_T(x)c_d^{t'} + \lambda_o^G(x)c_d^{t'} + \lambda_T^A(x)\left(c_d^{a'} + 2c_h\tilde{h}\right) \\ &+ \lambda_i(x)c_k \log\left(\frac{\delta(x)}{V_0}\right) + \lambda_o(x)c_k\log(2) + 2\lambda_T(x)c_f' + \lambda_i(x)c_k\log(|\mathcal{A}|)\end{aligned}$$

where it is understood that Π is a function of x. Coefficients for local operating headways:

$$\alpha_1^b = \lambda_b c_h;\ b = i, o \qquad\qquad \alpha_2 = c_d^t k(\delta)^{\frac{1}{2}} + c_q^t \delta$$

Coefficients for consolidation terminal densities and access operating headways:

$$\beta_1 = \lambda_T\left(\frac{\frac{2}{3}c_d^t}{\sqrt{\pi}V_0} - \frac{c_d^t k}{V_1}\right) + \frac{\frac{2}{3}|\lambda_o - \lambda_i|}{\sqrt{\pi}V_0}c_d$$

$$\beta_2 = c_d^t k \qquad \beta_3 = c_q \qquad \beta_4^b = \lambda_b^A \frac{c_h}{2} \qquad \beta_5^b = \lambda_b^G c_h;\ b = i, o \qquad \beta_6 = c_f$$

Coefficients for airport densities:

$$\chi_1 = \lambda_T^A\left(\frac{\frac{2}{3}c_d^t}{\sqrt{\pi}V_1}\right) + \frac{\frac{2}{3}|\lambda_o^A - \lambda_i^A|}{\sqrt{\pi}V_1}c_d \qquad \chi_2 = c_f + \sum_{b\in\mathcal{B}} \frac{r_2^A c_d + n_2^{A,b}c_q}{n_2^{A,b}} \qquad \chi_3 = \sum_{b\in\mathcal{B}} \frac{n_2^{A,b} - 1}{n_2^{A,b}} c_d k$$

Coefficients for breakbulk terminal densities and long haul operating headways:

$$\kappa_1 = \lambda_o^G\left(\frac{\frac{2}{3}c_d^t}{\sqrt{\pi}V_1}\right) + \frac{\frac{2}{3}|\lambda_o^G - \lambda_i^G|}{\sqrt{\pi}V_1}c_d - \bar{\lambda}^G c_d \frac{k}{V_2^G}$$

$$\kappa_2 = c_f \qquad \kappa_3 = c_d k \qquad \kappa_4 = c_q \qquad \kappa_5 = \lambda_o^G c_h$$

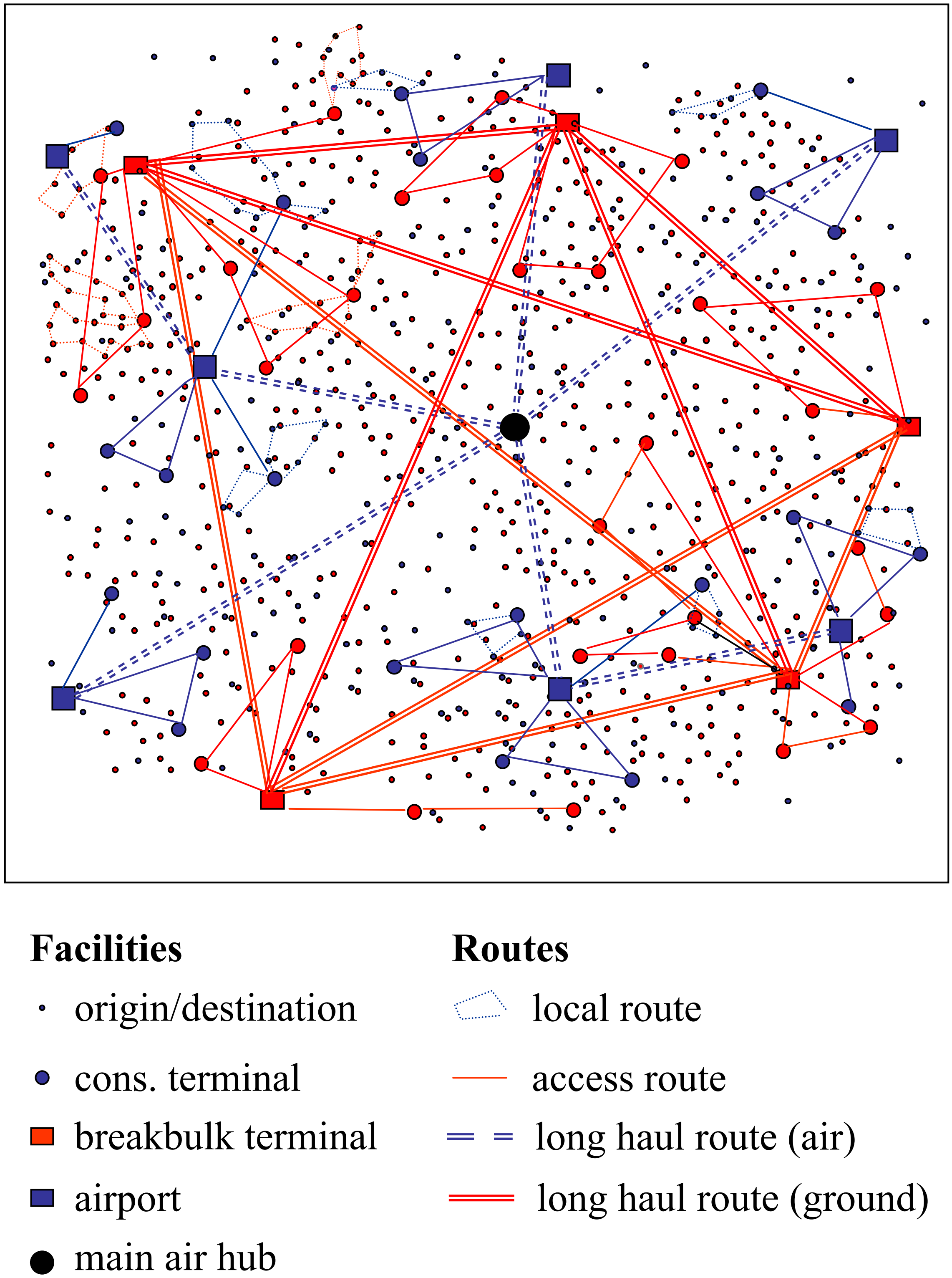

Figure 1: Integrated network

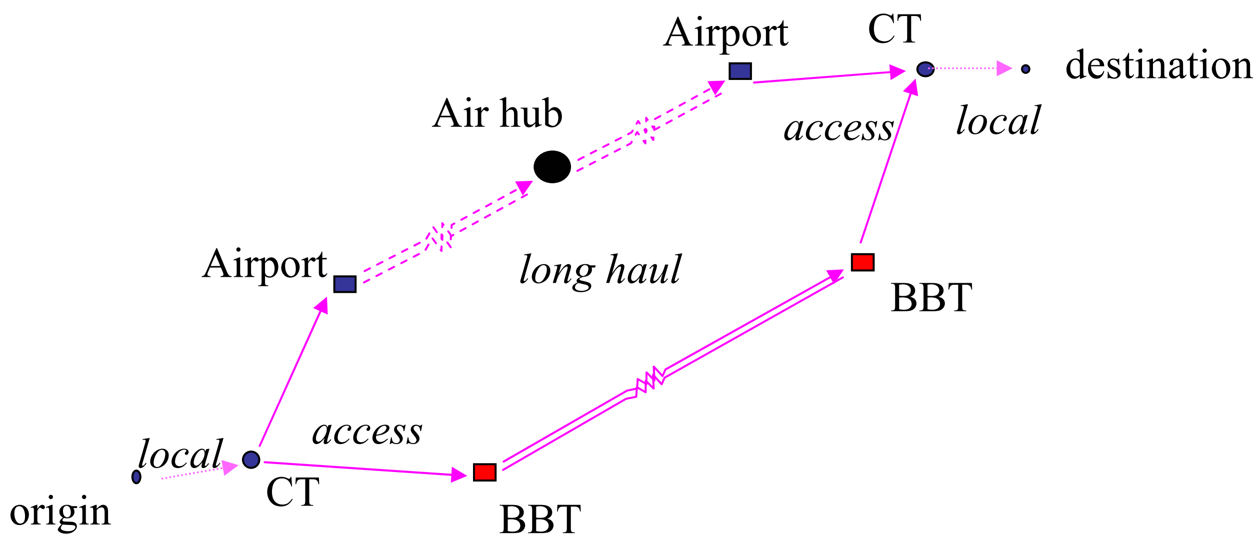

Figure 2: Distribution from origin to destination

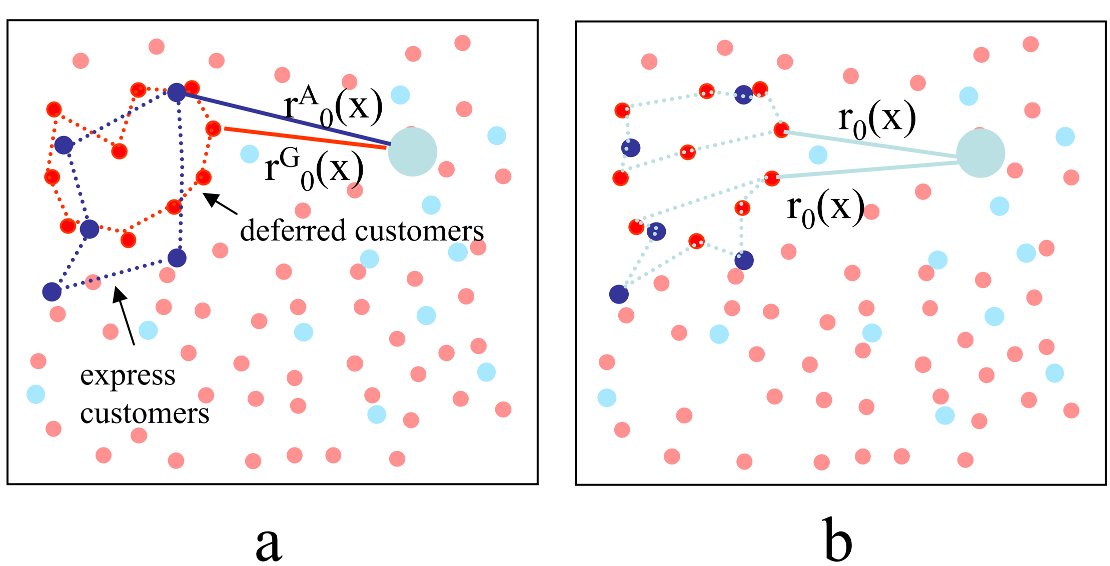

Figure 3: Local distribution

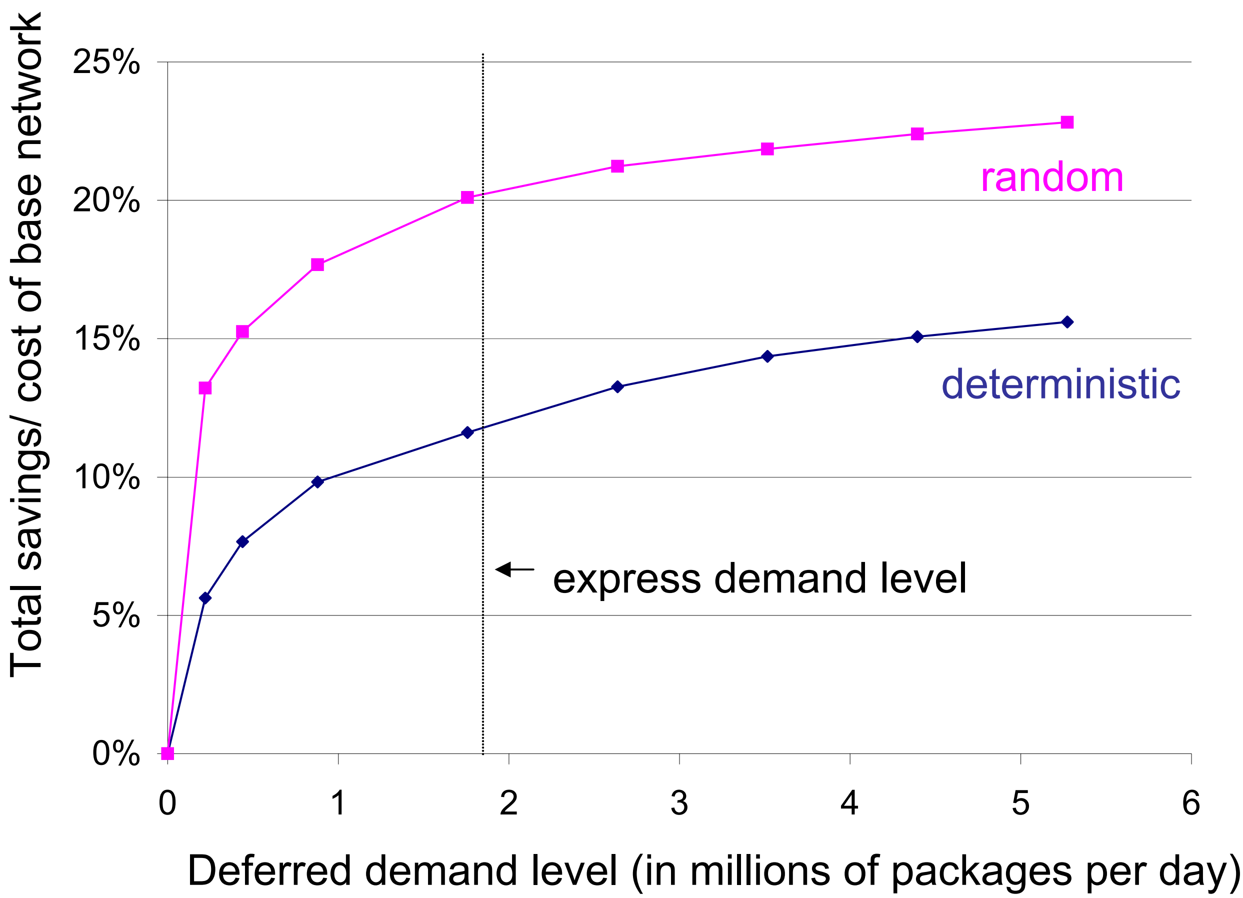

Figure 4: Savings comparison