

An integral constrained generalized hub-and-spoke network design problem

Cheng-Chang Lin^{a,*}, Sheu-Hua Chen^b

^a Department of Transportation and Communication Management Science, National Cheng Kung University,
1 University Road, Tainan 701, Taiwan, ROC

^b Department of Distribution Management, National Chin-Yi University of Technology, 35 Lane 215, Section 1, Chung-Shan Road,
Taiping City, Taichung County 411, Taiwan, ROC

Received 3 August 2007; received in revised form 19 November 2007; accepted 7 February 2008

Abstract

The hub-and-spoke network design problem determines the smallest fleet size with their routes and freight paths to minimize operating costs. We proposed a *generalized* hub-and-spoke network in a capacitated and directed network configuration that integrates the operations of three common hub-and-spoke networks: *pure*, *stopover* and *center directs*. An implicit enumeration algorithm with embedded integrally constrained multi-commodity min-cost flow was developed. Tested using the FedEx AsiaOne air network, the solution showed that the generalized network structure provided an operational flexibility, using center directs, topoffs/drop-offs, transits and rehandling via hubs, to determine the most cost-effective operations plan.

© 2008 Elsevier Ltd. All rights reserved.

Keywords: Freight transportation; Network design; Hub-and-spoke networks; Implicit enumeration algorithm

1. Introduction

The supply chain integrates from the suppliers to retailers in order to meet the end consumer needs. While key players (material providers, semi/final product manufacturers or assemblers) on the chain focus on their own core competent function, they outsource non-core logistics activities to the third-party logistics providers (3PLs). Time-definite freight delivery common carriers (as examples UPS, FedEx, DHL), under the less-than-truckload industry of transportation and warehousing sector include ground delivery carriers and integrated global express couriers. They are one of the third-party service providers who offer on-time transport service. They pick up, consolidate and deliver small shipments within committed service levels. They offer economies of scale by consolidating small shipments, allowing shippers to ship-on-order.

* Corresponding author. Tel.: +886 6 275 7575x53240; fax: +886 6 275 3882.
E-mail address: cclin@mail.ncku.edu.tw (C.-C. Lin).

The hub-and-spoke network with variations is widely used by the time-definite freight delivery carriers for the *line-haul* operations (Fig. 1). *Centers* are points of pickup and delivery while *hubs* are points of consolidation in a hub-and-spoke network. *Spoke routes* connect centers and hubs, while *inter-hub routes* connect a pair of hubs. Inter-hub routes are sometimes referred as *primary (trunk) routes*, while spoke routes are *secondary routes* (Current, 1988; Lin and Chen, 2004). The configuration substantially reduces the number of partial loads in the line-haul operations. The decrease in transportation cost outweighs a small increase in the handling cost, and the overall line-haul operating cost to the carriers is reduced.

In the line-haul operations, pickup and delivery centers are, respectively, *origin* and *destination* for the shipments. They together are called an *OD pair*, with an associated committed service level. To deliver all the OD shipments while meeting the service levels and operational restrictions, the carriers must make three interrelated sets of decisions, *freight paths*, *loads/empties*, and *vehicle routes*. A single *freight path* for each OD must be determined that starts at its origin (pickup) while proceeding along its path toward its destination (delivery) through a sequence of hubs or centers, unloading, consolidating/transiting and reloading, as required while meeting service level and operational restriction. A fleet of heterogeneous trailers/aircraft, carrying equipment is then assigned to haul freight constrained by their individual carrying capacities. The units of hauled freight are called *loads*. The existence of unproductive *empty* moves within the system ensures carrying equipment of all types is balanced for all facilities. *Vehicles*, a generalized term for means of transportation are tractor–trailers (in practice, called *feeders*) for ground freight delivery carriers (Lin, 2001). But they are aircraft for air express couriers. A tractor may couple with a number of sizes of trailers, as an example, a double (triple) is a tractor pulls two (three) 28-ft trailers. Subsequently, vehicle routes, a sequence of *hubs and/or centers* (stops) together with a timetable for the whole fleet of heterogeneous vehicles with departure and arrival times are scheduled to transport loads as well as reposition *empties*. As the result, there are two types of flows in the hub-and-spoke network, *freight* (package or parcel), and *vehicle* (tractor–trailer or aircraft). Thus, the network design problem in the line-haul operations for the carriers is to simultaneously determine a smallest possible fleet of vehicles with their routes and freight paths so that the sum of fixed fleet, handling and transportation operating costs is minimized while satisfying the demand and meeting the service levels and operational constraints.

There are in practice, three types of hub-and-spoke (H/S) networks, pure, stopovers with feeders, and stopovers with feeders and center directs.

1.1. Pure H/S network

The network shown in Fig. 1 requires that all the vehicle movements (on directed links) must either start or end at a hub. As the result, freight will be flown from their origin centers, via a/some hub(s), before traveling

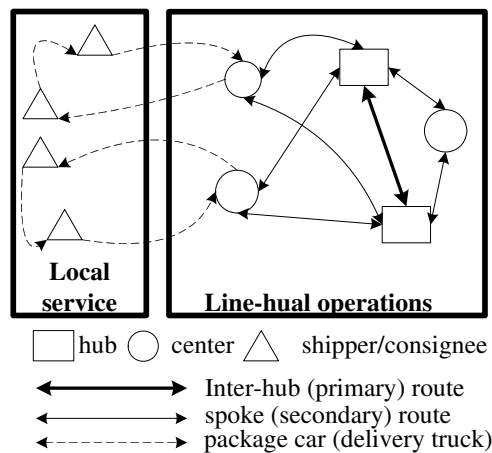


Fig. 1. An illustrative pure hub-and-spoke network.

to and arriving at their destination centers. At hubs, freight on the inbound vehicles is completely unloaded, sorted and reloaded unto outbound vehicles for destination centers or other hubs if further handling is economical. Depending on the number of hubs a center is allowed to connect with, the pure H/S network is further classified (Bryan and O’Kelly, 1999). A *single-assignment* H/S requires any center to be uniquely assigned, meaning vehicle and freight to an exclusive hub (Ernst and Krishnamoorthy, 1998; Sohn and Park, 2000; Ebery, 2001; O’Kelly and Bryan, 2002). On the other hand, when all centers are free to connect to one, several or all of the hubs, it becomes a *multiple-assignment* network design problem (Lin, 2001; O’Kelly and Bryan, 2002; Groothedde et al., 2005).

Lin (2001) studied the integral constrained freight routing planning, while Leung et al. (1990) and Lin and Lin (2007) respectively studied the single- and multiple-time-period load planning that simultaneously determines the freight paths and a balanced carrying equipment network. Their studies were in a single ground mode network. On the other hand, it may result in an intermodal network design problem when aircraft or trucks are the transport equipment on the spoke routes (O’Kelly and Lao, 1991); or barges and trucks are the respective transport equipment on the trunk and spoke routes (Groothedde et al., 2005).

1.2. H/S network with stopovers and feeders

The network is shown in Fig. 2a. In addition to the primary (trunk) and secondary (spoke) routes, smaller capacity vehicles are used to haul commodity from (to) light pickup (delivery) centers to (from) *transit* centers on the secondary routes. Kuby and Gray (1993) called this *feeder* operations (*feeder*, used as an adjective in this case, is distinct from a tractor–trailer as described in third paragraph). At transit centers, freight is unloaded from the inbound feeders and reloaded onto (higher carrying capacity) vehicles on the secondary route and vice versa, called *feeder–secondary transit*. Rarely, but if necessary, transit centers may rehandle feeder’s inbound pickups (outbound delivery). Center 5 is a transit city for feeder route center 7 → center 5 in Fig. 2a. Furthermore, the network allows vehicles on the primary as well as secondary routes to stop over at intermediate center(s) (hubs) which is rarely the case for feeders. As the result, an inbound (outbound) vehicle departing from an origin center (hub) terminated at a hub (destination center) may stop over at other centers loading (or unloading) their respective pickup (or delivery) freight, called *topoffs* (*drop-offs*) as center 1 → center 2 → hub A (hub B → center 3 → center 4). Similarly, the analogous operations may be performed equally well on hub-to-hub trunk routes, as hub C → center 8 → hub A. This detail of routing is called *stopovers*. Even though detours may increase the transportation cost, the saving through reduction in otherwise required vehicle movement may in some cases decrease the overall operating cost.

The stopover and feeder operations resulted in two patterns of freight paths. The first pattern, origin center → hub(s) → destination center, is similar to the exclusive pattern in the pure H/S network. However, there is an additional freight path for centers on feeder routes, that is, origin center (on feeder) → *transit* center (on secondary) → hub(s) → *transit* center (on secondary) → destination center (on feeder). Armacost et al. (2002) and Kuby and Gray (1993), respectively, studied the network design problem with only stopovers and with stopovers and feeders in a single air-mode network; while Kim et al. (1999) studied the multimodal network design problem with stopovers for the multimodal express delivery.

A *hierarchical* H/S network, as shown in Fig. 2b is a special type of H/S with stopovers and feeders network. The network restricts the assignment of each center to one and only one hub. In addition, no primary routes can stop over any other centers. When the primary routes, connecting hubs, are determined, centers are connected to their respective hubs by means of minimum spanning trees, assuming no side constraints attached on the secondary routes. Thus, Current (1988) enumerated all possible primary routes to determine the optimal hierarchical H/S network. Lin and Chen (2004), on the other hand studied the degree and time constrained hierarchical H/S for time-definite common carriers.

1.3. H/S network with stopovers, feeders and center directs

The network is similar to the H/S network with stopovers and feeders, but with additional operational flexibility. That is, whenever a vehicle on secondary routes stops over a center loading additional pickups, carriers may simultaneously unload the delivery freight. This is called *center directs*, no hub rehandling is performed,

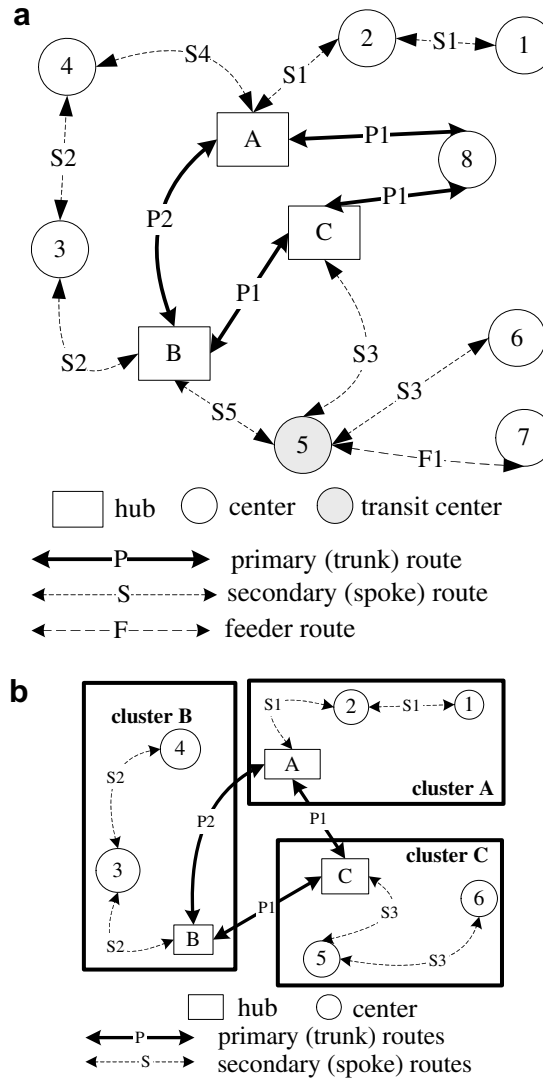


Fig. 2. An illustrative H/S with stopovers and feeders and variations: (a) a H/S with stopovers and feeders; (b) a hierarchical H/S network.

as shown in Fig. 3b. The advantage of unloading the arrival freight, which may require stacking together during transports, is in freeing up vehicle’s carrying capacity. The increase in space utilization may result in a smaller fleet of vehicles for line-haul operations. The operations also allow freight being carried on secondary routes to be unloaded at (*secondary–secondary*) transit cities and reloaded onto other secondary routes, as shown in Fig. 3b. Thus, in addition to the freight path patterns in the pure H/S and H/S with stopover and feeder, comparing Fig. 3a and b, it has several additional unique freight patterns. They are (1) origin center (on feeder) → (*feeder–secondary*) transit (destination) center (on secondary), (2) origin center (on feeder) → (*feeder–secondary*) transit center (on secondary) → destination center (on secondary), (3) origin center (on secondary) → destination center (on secondary) and (4) origin center (on secondary) → (*secondary–secondary*) transit center (on secondary) → destination center (on secondary). Respectively, these are the OD shipments of center 5–center 4, center 5–center 3, center 4–center 3 and center 4–center 6 in Fig. 3b. Assuming a *symmetric* demand, Lin et al. (2003) enumerate all possible freight paths and vehicle routes to show that allowing center directs can provide additional carrying capacity and allow carriers to dispatch not only a smaller vehicle but also a smaller vehicle fleet than is usually for stopover operations.

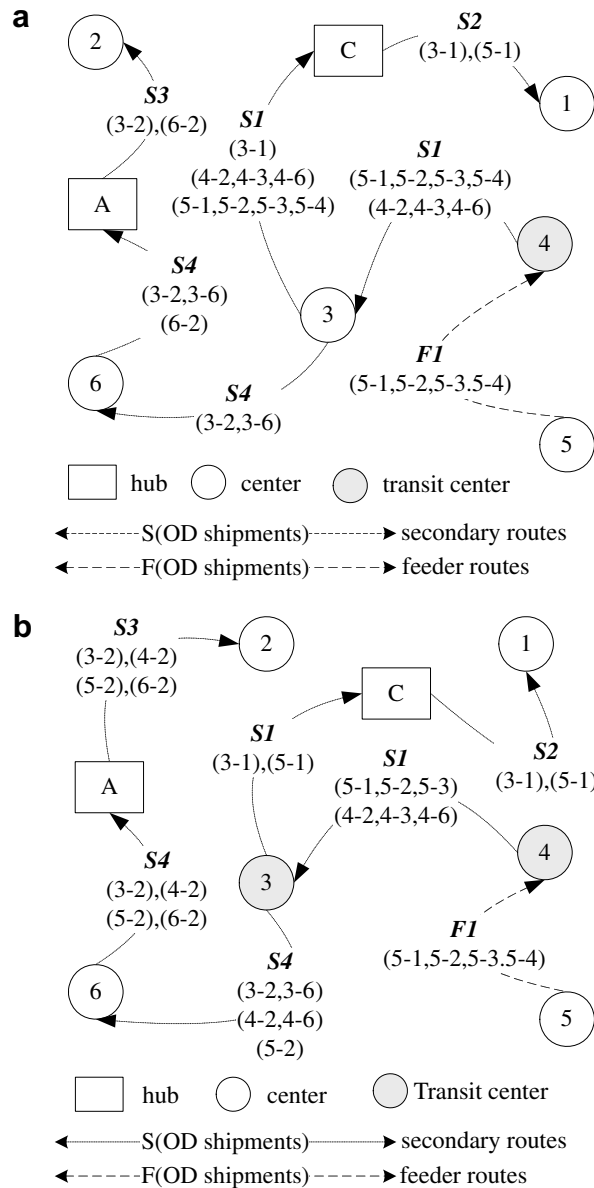


Fig. 3. An illustrative one-directional H/S with stopovers, feeders and center directs: (a) freight flows without center directs; (b) freight flows with center directs.

The main purpose of this research in the context of past researches is to study the network design problem in the *generalized* hub-and-spoke network, that is capable to integrate all three above-mentioned H/S networks and each respectively is a special case. Such a generalized network allows an evaluation of the trade-offs between different possible freight operations, hub rehandling, toffoffs/drop-offs, transits and center directs, and also vehicle route operations, stopovers and feeders. As the result, the carriers may rehandle freight via hubs to reduce empties, pick up (deliver) freight on vehicle routes to reduce loads, transit freight at transit center to use small vehicle fleet, or ship directly from pickup to delivery centers to increase space utilization, whichever is economical. In this research, the network makes no symmetric assumption, which is a rigid assumption used in some previous studies (Kuby and Gray, 1993; Lin et al., 2003). The symmetric assumption excluding asymmetric vehicle routes resulted in a smaller feasible set and therefore excludes some of the superior solutions.

Several complexities of network design problem in the generalized H/S network include: (1) the integration of *candidate* primary, secondary and feeder routes, and then the determination of a minimal but balanced vehicle route plan for a fleet of heterogeneous vehicles; (2) the integration of various freight operations, hub rehandling, topoffs/drop-offs, transits and center directs, and then the determination of minimal integral and time constrained freight paths; (3) the mutual interaction between the vehicle route plan and freight paths coupled by the carrying capacities. The complexity on integral- and time-constrained freight paths is particularly applicable to the time-definite freight delivery industry (Lin, 2001), while the complexity of the freight operational flexibility merely in the generalized H/S network.

Our methodology consists of (a) the space–time network configuration design, (b) the mathematical model in the space–time network, (c) the algorithmic scheme, a separable approach, and (d) the computation and analysis. We propose a two dimensional *space–time* network which is represented as a capacitated and directed network configuration. It integrates the timetable and carrying capacities of all *candidate* primary, secondary and feeder routes for the fleet of heterogeneous vehicles, that is the complexity (1), and the time windows and handling capacities of freight handling activities at all facilities. Furthermore, vehicle arrivals (departures) are physically connected to the start (from the end) times of handling activities when time permits, the complexity (3). As the result, it allows via hub, topoff/drop-off, transit and center directs operations to be performed on the primary and secondary routes and transit on feeder routes. Thus, the freight may be unloaded for direct movements (H/S with center directs), unload and reload for feeder transit or stopover operations (H/S with stopovers and feeders), and shipped directly to hub (pure H/S), the complexity (2). Therefore, no explicit enumerations of all possible paths for all OD pairs are necessary (Kuby and Gray, 1993; Lin et al., 2003) which could be enormous and intractable.

Furthermore, such a configuration is also fundamental to the model formulation and algorithmic scheme. It is capable of separately solving the respective vehicle routes and freight paths subproblems. The implicit enumeration is an effective approach for solving models with binary variables. We may branch on the vehicle route binary variable, whether or not to dispatch. As the result, the carrying capacities can be computed and constrained on the freight paths subproblem, if operationally feasible. We may solve the separable freight routing subproblem which is an integral and time constrained multi-commodity min-cost flow problem (IMCP). This allows some of computationally efficient network flow algorithms, such as shortest path (Ahuja et al., 1993) to be applied and/or modified for practical applications.

The structure of this paper is as follows. In Section 2, we describe the capacitated and directed space–time network configuration for the generalized H/S network. The network design problem is to simultaneously determine a smallest possible fleet of vehicles with their routes and freight paths so that the sum of fixed fleet, handling and transportation operating costs is minimized while satisfying the demand and meeting the operational constraints. In Section 3, we model the network design problem as an integer program. In Section 4, we develop an implicit enumeration algorithm with embedded integral constrained multi-commodity min-cost flow subproblem to solve the network design problem. Federal Express is an integrated air express common carrier that provides time-definite services for small shipments. In Section 5, we use Federal Express AsiaOne network for numerical testing. It allows us to verify the model development, and also assess the efficiency and effectiveness of the algorithm. In addition, we conduct the sensitivity analysis on demands to realize the behavior of the optimal solutions. We conclude this paper with a research summary and future research.

2. Network structure

The generalized H/S network is represented as a capacitated and directed space–time network shown in Fig. 4. The horizontal dimension represents candidate primary, secondary and feeder routes with departure and arrival times, while the vertical represents pickup, delivery and stopover operations of centers and stopover and consolidation operations of hubs with handling durations. They are integrated by connecting arrival to and/or departure from the freight handling activities in the facilities, if time permits. Thus, the network consists of nodes and directed links. The node and link classifications and their associated attributes are organized in Table 1. A node represents an *activity* in a facility. There are six types of nodes. Each node has at most three associated attributes, the unit freight handling cost, the elapsed time (with associated start and end times) and handling capacity:

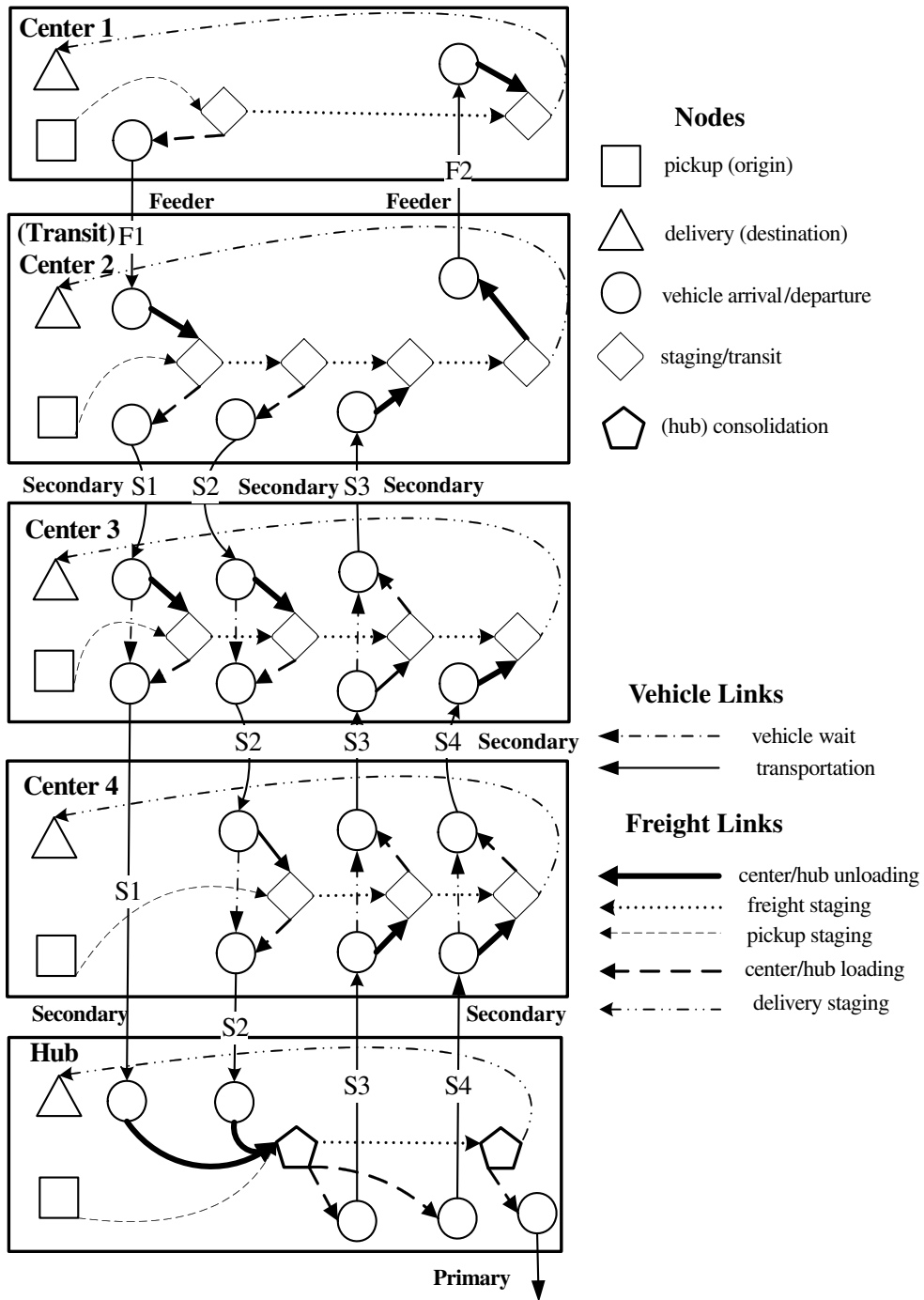


Fig. 4. The configuration for an illustrative generalized hub-and-spoke network.

a. The *pickup* node is where the freight of an OD pair originates, while the *delivery* node is where it terminates. The cutoff time and to be delivered time are the respective start and end times for pickup and delivery nodes. There are no operational time restrictions (or limits) associated with either node. However, their costs and capacities are respective handling and design capacities.

Table 1
Node and link types and attributes

Type	Function	Cost	Capacity	Time duration
<i>Nodes</i>				
Pickup	Freight originates	Handling ^a	Handling	None
Delivery	Freight destines	Handling ^a	Handling	None
Arrival	Vehicle arrives	None	None	None
Departure	Vehicle departs	None	None	None
Staging/transit	Stack pickup, delivery and/or transit	None	Handling	None
Consolidation	Consolidate	Consolidation ^a	Handling	Consolidation
<i>Links</i>				
Center unloading	Unload inbound	Unloading ^a	None	Unloading
Center loading	Load outbound	Loading ^a	None	End of the staging to vehicle departure
Hub unloading	Unload inbound	Unloading ^a	None	Vehicle arrival to consolidation start
Hub loading	Load outbound	Loading ^a	None	End of the consolidation to vehicle departure
Vehicle wait	Stop over	Crew idle time ^b	Carrying	Vehicle arrival to its departure
Transportation	Travel	Crew ^b ; unit carrying ^a	Carrying	Travel time between two facilities
Pickup staging	Stage center pickups	None	None	Pickup cutoff to the start of staging
Delivery staging	Stage center delivery	None	None	End of the staging to the start of delivery
Mixed staging	Stack pickup, delivery and/or transit	None	None	The elapsed time between the start time of two staging nodes

^a Unit cost associate with freight.

^b Unit cost associate with vehicle.

- b. A stop at any center by a vehicle resulted in three nodes, *arrival*, *departure* and *staging/transit*. The arrival and departure nodes are associated with the operations of the vehicle, namely, the arrival at and departure from that center. Their start times are the respective arrival and departure times of the vehicle with no duration for either node. Of course, the origin (or conversely, destination) of a vehicle does not have arrival (or conversely, departure) node. The staging/transit node represents the freight inventory status and also the transit operations at the center. It stocks (1) the pickup, and/or (2) feeder or secondary transit waiting to be loaded onto outgoing vehicles, and/or (3) delivery unloaded from the incoming vehicles for the respective center. The start as well as the end time for the staging/transit node is the vehicle arrival (or conversely, departure, if the origin of a vehicle) with an additional (deductive) unloading (loading) time. Thus, each staging node has an associated design capacity with no duration.
- c. At hub, in addition to arrival and departure nodes, the *consolidation* node is a sort of that facility. At the consolidation nodes, freight must be completely unloaded from the incoming vehicles and sorted before reloading unto outgoing vehicles. Typically, there are several consolidation operations with a minimum time interval in between for operator rotation at each hub. The larger the number of sorts, the lower the average fixed hub cost. Each has an associated consolidation cost, predetermined start and end times, duration and design capacity.

A directed link connects two nodes with no intermediate nodes in between. There are five types of links associated with at most three attributes, the unit freight carrying cost, the elapsed time and carrying capacity:

- a. (*Center/hub*) *unloading* (*loading*) links connect arrival (staging/transit) nodes to staging/transit (departure) nodes with associated unit unloading (loading) cost. While the elapsed times for center unloading (loading) links are respective unloading times (the end of staging to the vehicle departure), it is the end of arrival (consolidation) to the beginning of consolidation (departure) at hubs.
- b. *Vehicle wait* links connect arrival and departure nodes of their respective stops. Each link has an associated vehicle operating cost that represents a nonproductive idle expenditure of the vehicle drivers (aircraft pilots). The associated duration is the elapsed time from the arrival to the departure of the vehicle.

- c. *Transportation* links connecting two facilities with associated vehicle transportation cost and travel time. Furthermore, it has an associate unit carry cost for each unit of freight carried.
- d. *Staging* links stack the commodity for the respective facility. They include three types. *Pickup staging* and *delivery staging* stacks pickups and delivery respectively, while *mixed staging* connects a pair of consecutive staging/transit stacks pickup and/or delivery and/or transit. All staging links have a property that the start time of the tail (from) node must be no earlier than the head (to) node. This ensures that pickups (delivery) can be loaded (unloaded). It further ensures that the transit freight can only be transferred from an earlier to the later vehicle with sufficient time interval for both unloading and loading at the stop. All links have no costs, while the elapsed times are the end of staging (pickup) to the start of staging/delivery (staging) nodes.

All possible freight paths are feasible in the generalized H/S network (Fig. 4). (1) Direct feed for hub consolidation, for example, freight loaded on secondary (S1) for hub. (2) Feeder via secondary for hub consolidation, for example, freight loaded on feeder route (F1) at center 1, transferred at (*feeder–secondary*) transit center 2 unto secondary route (S1) before ending at hub. (3) Center direct, for example, freight loaded on secondary route (S1) at center 2 and unloaded at center 3. (4) Note that this network configuration allows center via (*secondary–secondary*) transit center for center directs, for example, freight loaded on secondary (S1) at center 2 may transfer at center 3 onto secondary (S2) and unload at center 4 for center 2-to-center 4 direct. However, if this practice is disfavored by some freight carriers, we may set all the unloading (loading) cost to be infinite on all secondary routes inbound to (outbound from) hub. This resulted in an H/S network with stopovers and feeders.

3. Network design problem

Consider a set S of s centers, and a set H of h hubs with one consolidation operation for each hub. As the result, there are S respective pickup and delivery nodes and H consolidation nodes in the space–time network specified in Section 2. Furthermore, there are at most $|M| = |S + H| * |S + H|$ OD pairs with the generic element $m \in M$. Denote (m_o, m_d) the origin and destination for OD pair m . Origins (pickups) and destinations (delivery) form sets O and D , respectively. The fleet of vehicles consists of k types, each with a size of \hat{H}^k , $k \in K$. In addition, let a set P^k with index p , contains all possible vehicle routes for type- k , $k \in K$. The start (end) of the route has associated departure (arrival) node, while each intermediate stop has both associated nodes. In summary, the respective capacitated and directed network consists of N of i nodes and A of ij links. Node unit handling cost and duration are denoted as C_i and t_i , while C_{ij}^v , C_{ij} and t_{ij} are associated vehicle transportation and freight carrying costs and duration for each link. Each OD pair $m \in M$ has an associated demand Q^m and service level, T^m , the elapsed time from the cutoff of pickup to the beginning of the delivery.

A *vehicle route* consists of a sequence of facilities with their associated arrival and departure nodes and links connecting those nodes. The operating cost for a vehicle k on path p is the sum of vehicle's fixed, C^k , $k \in K$ and vehicle transportation costs, that is, $C_p^k = C^k + \sum_{ij \in A} \delta_{ij,p}^k C_{ij}^v$. The former cost is independent of the route assigned, while the latter is the sum of expenses for transporting with no carrying commodity between the facilities on the route. In practice, the common carriers amortize the capital expenditure of feeder fleet by a yearly depreciation rate during its usage span or taxed years. In this research, we applied the constant amount linear approach, the most common practical approach for accounting purpose. Thus, the vehicle's fixed cost was the yearly depreciation. On the other hand, a *directed freight path* of an OD pair consists of a sequence of nodes that starts at its origin and ends at its destination, and nodes and links in between. The unit freight carrying cost for an OD pair m on path $\{\hat{x}_{ij}^m = 1, ij \in \hat{A}\}$ is the sum of handling and carrying costs, that is, $C^m = \sum_{ij \in \hat{A}} (C_i + C_{ij}) \hat{x}_{ij}^m + C_{m_d}$. The network design problem in a generalized hub-and-spoke network is defined as follows:

$$\text{Objective: } \min \psi(h, x) = \sum_k \sum_p C_p^k h_p^k + \sum_m Q^m \left[\sum_{ij \in A} (C_i + C_{ij}) x_{ij}^m + C^{m_d} \right] \quad (1)$$

Subject to:

$$\sum_i x_{ij}^m - \sum_i x_{ji}^m = \begin{cases} 1; & j = m_d \\ -1; & j = m_o \\ 0; & \text{otherwise} \end{cases} \quad \forall j \in N, m \in M \tag{2}$$

$$\sum_j \sum_{m_d} Q^{m_o m_d} x_{m_o j}^{m_o m_d} \leq \hat{V}_{m_o} \quad \forall m_o \in O \tag{3}$$

$$\sum_{i:j \in A} \sum_m Q^m x_{ij}^m \leq \hat{V}_j \quad \forall j \in N \setminus O \tag{4}$$

$$\sum_m Q^m x_{ij}^m \leq \sum_k V^k \left(\sum_p \delta_{ij,p}^k h_p^k \right) \quad \forall ij \in A \tag{5}$$

$$\sum_j \sum_p \delta_{ji,p}^k h_p^k - \sum_j \sum_p \delta_{ij,p}^k h_p^k = 0 \quad \forall i \in N, k \in K \tag{6}$$

$$x_{ij}^m, h_p^k \in \{0, 1\} \quad \forall ij \in A, m \in M, p \in P, k \in K \tag{7}$$

Other side constraints

$$\sum_{ij} (t_i + t_{ij}) x_{ij}^m + t_{m_d} \leq T^m \quad \forall m \in M \tag{8}$$

$$\sum_j z_{ij}^{m_d} \leq 1 \quad \forall i \in N \setminus D, \forall m_d \in D \tag{9}$$

$$Bz_{ij}^{m_d} - \sum_{m_o} x_{ij}^{m_o m_d} \geq 0 \quad \forall ij \in A, \forall m_d \in D \tag{10}$$

$$\sum_j \sum_k \sum_p \delta_{ij,p}^k h_{p(h)}^k \leq 1 \quad \forall i \in N \setminus H \tag{11}$$

$$\sum_i \sum_k \sum_p \delta_{ij,p}^k h_{(h)p}^k \leq 1 \quad \forall j \in N \setminus H \tag{12}$$

$$\sum_p h_p^k \leq \hat{H}^k \quad \forall k \in K \tag{13}$$

$$z_{ij}^{m_d} \in \{0, 1\} \quad \forall ij \in A, \forall m_d \in D \tag{14}$$

with decision variables:

$x_{ij}^m = 1$, if the freight of m th OD pair transverse on link ij ; $x_{ij}^m = 0$, otherwise.

$h_p^k(h_{p(h)}^k, h_{(h)p}^k) = 1$, if a vehicle of type- k is dispatched on route p (terminated at hub h , originated from hub h); $h_p^k(h_{p(h)}^k, h_{(h)p}^k) = 0$, otherwise.

$z_{ij}^{m_d} = 1$, if j is the outbound node for destination d at node i ; $z_{ij}^{m_d} = 0$, otherwise.

and parameters:

$\delta_{ij,p}^k = 1$, if link ij on vehicle route p of type- k ; $\delta_{ij,p}^k = 0$, otherwise.

The objective function (1) is to minimize the sum of the fixed and transportation costs for vehicle routes and the handling and carrying costs for freight paths. Constraints (2) enforce the *flow conservation*. All freight must depart from their origins and arrive at their destinations. Any intermediate node, vehicle arrival, departure, staging, or hub sort, may never stage any freight. Constraints (3) and (4) are the *node capacity* constraints. The former constrains all origin nodes, while the latter constrains all other nodes. Constraints (5) are the *bundling* constraints, coupling freight flows and vehicle routes, which require that the freight can only be flown on the vehicle routes. In addition, they serve as the *link capacity* constraints, that is, the freight flows may not exceed vehicle's carrying capacities assigned on any link on any vehicle route. Constraints (6) are the *balancing*

constraints state that the incoming vehicles must be equal to the outgoing vehicles for all facilities. All decision variables, the vehicle routes and freight paths are binary, 1 and 0 otherwise, as stated in constraints (7). They form a set of *common* constraints applicable for almost of all applications (Current, 1988; Bryan and O’Kelly, 1999; Kim et al., 1999; Armacost et al., 2002).

On the other hand, there are other side constraints shown in various applications. Constraints (8) enforce the service commitments (Lin, 2001; Lin and Chen, 2004). The freight path from the origin to the destination for any OD pair must meet the service level. Constraints (9) are the restriction that there is at most one outbound node for each destination at any node. Constraints (10) state that all the freight paths can be routed via that outbound node for that destination, if assigned. Parameter B is a large number. Constraints (9) and (10) together ensure that there must be a *directed in-tree* rooted at any destination (Lin, 2001). Constraints (11) and (12) are the *maximum connecting* constraints (Lin et al., 2003). They, respectively, state that at most, there is a single vehicle route departing (arriving) from (at) a center to (from) any hub. Constraints (13) are the fleet size constraints on each of all types. Lastly, constraints (14) restrict the decision variables, one outbound node per destination at any node are binary. Collectively, constraints (8)–(14) are typical operational constraints for the time-definite freight delivery common carriers (Lin, 2001; Lin et al., 2003).

4. Solution algorithm

The space–time network configuration allows us to design the following separable algorithmic scheme. We branch on the vehicle route binary variable, dispatching or not, resulting in a vehicle route plan. If operationally feasible, we may solve the freight routing subproblem which is an integral and time constrained multi-commodity min-cost flow problem (IMCP). An associated lower bound may subsequently be computed to decide whether or not the branch should be fathomed for computational efficiency, a depth-first approach. In addition, the separable algorithmic scheme also allows the modularization in coding and future expansion. Its complete flowchart and modularization is presented in Fig. 6.

A *feasible* vehicle fleet size and route plan is a set of vehicle routes, $\{\hat{h}_p^k \in (0, 1), \forall p \in P, k \in K\}$, in which the vehicle balancing constraints (6) $\sum_j \sum_p \delta_{ji,p}^k \hat{h}_p^k - \sum_j \sum_p \delta_{ij,p}^k \hat{h}_p^k = 0, \forall i \in N, k \in K$, the maximum covering constraints (11) and (12) $\sum_j \sum_k \sum_p \delta_{ij,p}^k \hat{h}_p^k \leq 1, \forall i \in N \setminus H$ and $\sum_i \sum_k \sum_p \delta_{ij,p}^k \hat{h}_p^k \leq 1, \forall j \in N \setminus H$, and fleet size constraints (13) $\sum_p \hat{h}_p^k \leq \hat{H}^k, \forall k \in K$ are met. Observe that for any given feasible vehicle fleet and their routes, the network design problem becomes an IMCP. That is, $\psi(x) = \min \sum_m Q^m \left[\sum_{ij \in A} (C_i + C_{ij}) x_{ij}^m + C_{m_d} \right]$ subject to (2)–(5), (8)–(10) for $x_{ij}^m, z_{ij}^{m_d} \in \{0, 1\}$. Based on this observation, in this research, we develop an implicit enumeration algorithm with embedded IMCP subproblem for network design problem.

First, we construct a search matrix (shown in Fig. 5) for implicit enumeration algorithm. It consists of rows for facilities and columns for vehicle routes. Each facility has two rows, outbound and inbound vehicle routes. Outbound (inbound) vehicle routes are vehicle departures (arrivals) at the facility and travels to (originates from) a hub. Each cell represents one of the possible *outbound (inbound) fleet* and routes that meet the maximum covering constraints (11) and (12). That is, it may contain two vehicle routes for a two-hub network, but at most one vehicle to (from) any hub. Since, any center may be stopped over by another vehicle route started at another facility, the first node of each row is a *null* vehicle. To speed up the computation, we order the vehicle fleet in each of all rows in an ascending order according to their total costs.

A vehicle fleet size and route plan, called *combination*, is a set of cells, one for each row (facility). The one under evaluation, $\{\hat{h}_p^k \in (0, 1), \forall p \in P, k \in K\}$ is called the *current* combination. The implicit enumeration algorithm basically consists of two procedures, the determination of the current combination (branching) and the computation of the IMCP subproblem. The former is to determine vehicle fleet size and routes, while the latter determines the freight paths.

The algorithm starts with an initial vehicle route plan (current combination) that consists of the first cell of all rows. The detailed steps of the algorithm shown in Fig. 6, are discussed as follows:

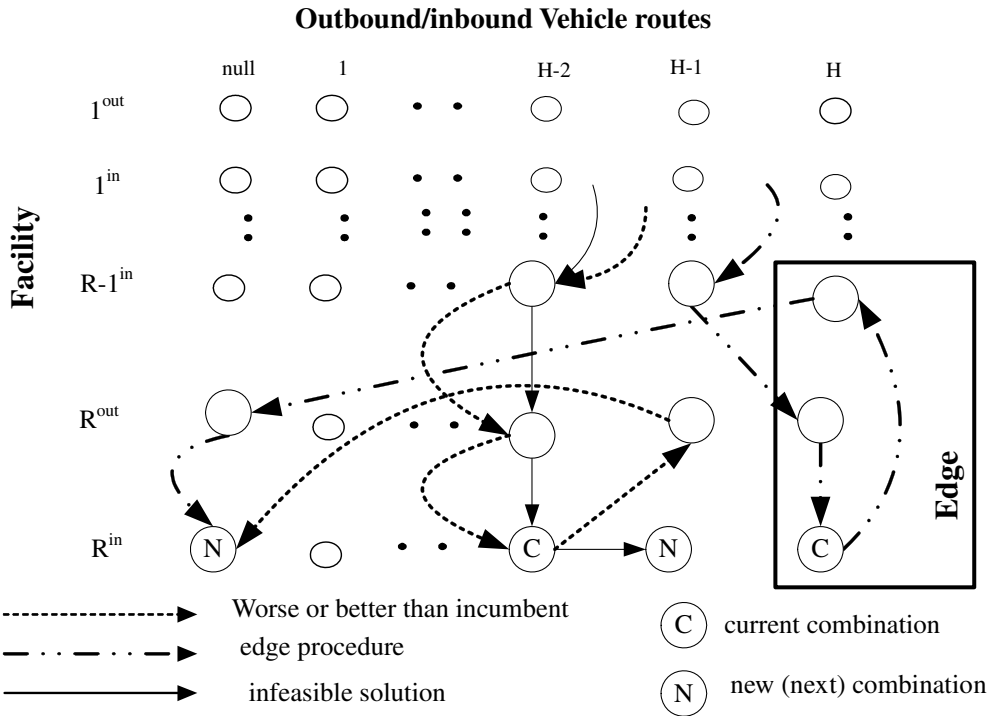


Fig. 5. The search matrix.

1. *Feasible vehicle route plan.* At the current combination, $\{\hat{h}_p^k, p \in P, k \in K\}$, we verify whether or not it is a feasible plan. A feasible vehicle route plan must satisfy the balancing constraints (6), maximum connecting constraints (11) and (12) and fleet size constraints (13). When infeasible go to step 4 for branching. Note that, a vehicle feasible route plan may not necessarily guarantee a feasible freight path. A trivial case is that no vehicle routes at the current combination is a feasible route plan.
2. *Lower bound.* Whenever the current combination is a feasible vehicle plan, we calculate its lower bound, $\psi(\hat{h}_p^k) = \sum_k \sum_p C_p^k \hat{h}_p^k + \sum_m \min C^{m*} Q^m$. It is the sum of the fixed cost of the current vehicle fleet and their routes, and the *least* handling and carrying costs for all OD pairs. However, we might underestimate it, since some of the centers are not connected unto the network. Thus, we determine a minimum-spanning tree, $ij \in \hat{A}$ to connect all the unconnected centers. Thus, the minimum transportation cost is $\sum_{ij \in \hat{A}} C_{ij}^v$. As the result, the lower bound becomes $\psi(\hat{h}_p^k) = \sum_k \sum_p C_p^k \hat{h}_p^k + \sum_{ij \in \hat{A}} C_{ij}^v + \sum_m \min C^{m*} Q^m$. Whenever the lower bound is greater than the operating cost of incumbent solution, go to step 4 for branching.
3. *IMCP.* We solve the subproblem to determine the freight paths for all OD pairs. The detailed steps are as follows:
 - 3.1. *Service feasible.* First, we verify whether there exists at least one freight path that satisfies the service level for all OD pairs. We implemented the dequeue implementation, a label-correcting algorithm to determine the least time path for all OD pairs (Ahuja et al., 1993). If this fails, go to step 4 for branching.
 - 3.2. *Least cost path.* We determine the least cost path using the same dequeue implementation code. If available, typically this is the center-to-center direct path with no additional hub consolidation costs, since it is superior to paths that require hub rehandling. However, the least time path is used instead, if it violates the required service level.
 - 3.3. *Network capacity feasible.* Load the demand unto the network according to their respective freight paths. If the flows on any node or link exceed its respective capacity, we move to step 4 for branching.

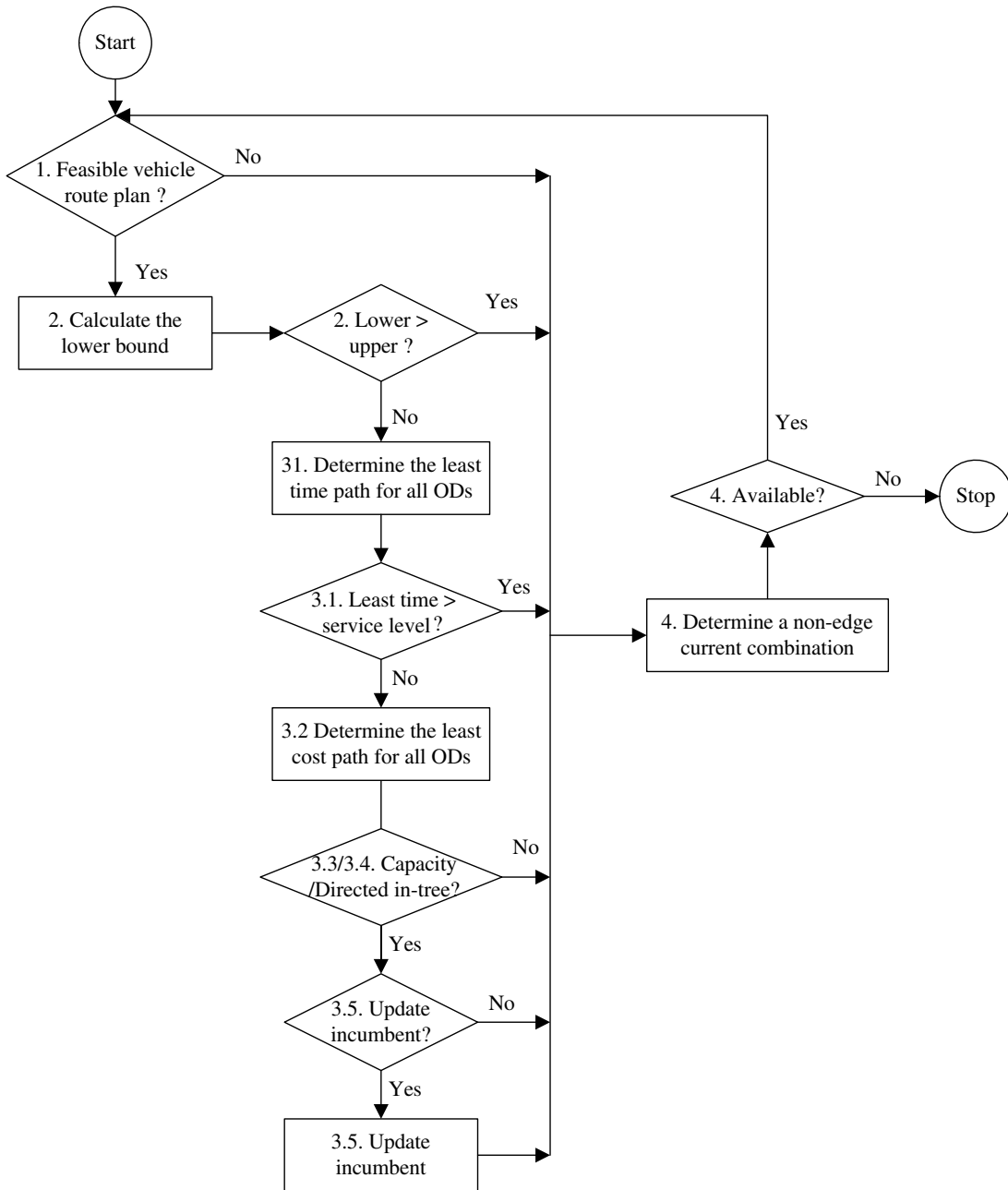


Fig. 6. The flowchart of implicit enumeration algorithm.

- 3.4. *Directed in-tree feasible.* We verify whether or not the freight paths for any destination form a directed in-tree configuration. Similarly, any violation is resulted in going to step 4 for branching.
- 3.5. *New feasible solution.* The operating cost for the new feasible solution is $\psi(\hat{h}_p^k) = \sum_k \sum_p C_p^k \hat{h}_p^k + \sum_m Q^m \left[\sum_{ij \in A} (C_i + C_{ij}) x_{ij}^m + C_{m_d} \right]$. If superior, we update the incumbent solution, and go to step 4 for branching.
- 4. *Branching.* There are three outcomes at the current combination. Each is associated with a different branching scheme as follows (see Fig. 5).

- (1) *Infeasible*. An infeasible vehicle combination or infeasible freight paths, we branch to the next cell of the current row which becomes the current combination, only when it is not an edge cell. Otherwise, we perform edging procedure.
- (2) *Inferior*. Since, we sequence the vehicle routes by ascending order, for a worse than the incumbent solution, we move to the next cell of the one above the current row. If an edge cell is encountered, we perform edge procedure.
- (3) *Superior*. For a superior solution, we update the incumbent solution. Upon completion, the same as the outcome in (2), we back track one row up until a non-edge of a row is determined. Otherwise, we perform edge procedure.

Whenever we encounter an edge combination, we perform the *edge procedure* as follows. We move upward to the one above the current row until a row with non-edge combination. When determined, the inclusion of all the first cell of the lower rows will result in a new current combination. When this fails, the current incumbent is the (local) optimal solution. Otherwise, there are no feasible solutions to the current demand. An exact algorithm to solve IMCP may be adopted in step 3 (Lin, 2001), if desired. That is, this method (or approach) explores all possible time-feasible paths to meet directed in-tree and capacity constraints with the minimal cost. We instead propose a heuristic approach that is to determine and implement the least-cost time-feasible path. This still may not guarantee optimality. However, in practice, the vehicle fleet cost is the dominating costing factor. This simplification in computational efficiency may still determine a solution very close to optimality.

5. Computational results

In this research, we used the Federal Express Far East's AsiaOne air network for numerical testing. The AsiaOne air network is depicted in Fig. 7. Federal Express assigned 7 MD-11, 2 DC-10 and 4 A310, a total of 13 aircraft to serve the Asia Pacific regional demand and their shipments to and from the rest of the world

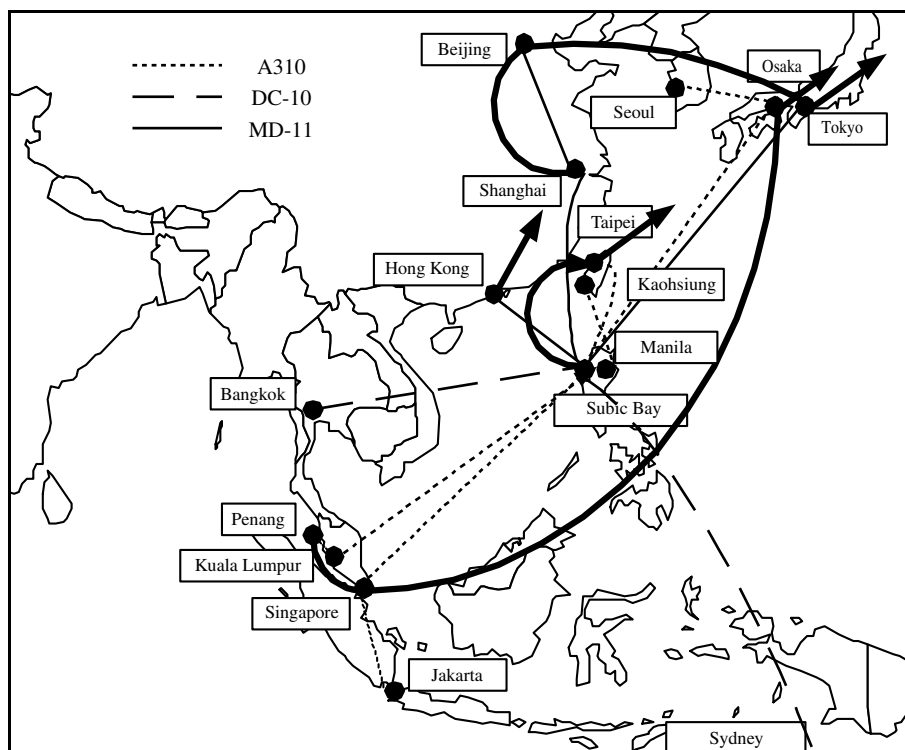


Fig. 7. The Federal Express existing AsiaOne air network.

(Lin et al., 2003). Studies of their flights showed that: (1) the A310 serves the Asia Pacific regional operations, while the MD-11/DC-10, a longer travel range and higher carrying capacity, serves the Asia Pacific and North America/European routes; (2) even though the carrier makes no center direct loads, more than half of the aircraft stop over and collect (drop off) additional pickups (delivery) on their routes to (from) the hubs; (3) whenever two aircraft routes terminated for two FedEx current regional hubs, Subic Bay and Anchorage, but stopping over in a common city, it makes such a city a candidate for cargo transit. *Seoul → Osaka → Subic Bay* (A310) and *Penang–Singapore–Osaka–Memphis* (MD-11) routes stop over Osaka, while routes *Shanghai → Beijing → Tokyo → Anchorage* (MD-11) and *Tokyo → Subic Bay* (MD-11) stop over Tokyo. As the result, Osaka and Tokyo may serve as transit cities for Seoul's non-Asia Pacific pickups, and Shanghai and Beijing's Asia Pacific pickups, respectively (see Fig. 7).

Since the candidate aircraft routes could be numerous, to make the problem tractable, we selected and restricted our candidate routes based on similar rules suggested by Kuby and Gray (1993). The rules are: (1) time feasible aircraft routes, meeting cut-off and delivery times for centers and consolidation start and end times for hubs; (2) only major centers in terms of pickup and deliveries can make direct connection to hubs; (3) any aircraft may only stop at other centers within a 150 miles radius. As the result, we determine, in total, 16 candidate aircraft routes as listed in Table 2. Furthermore, since FedEx only used 2 DC-10 out of 9 (with additional 7 MD-11) long range fleet, we selected MD-11 as an exclusive long range aircraft type.

To conduct the numerical testing, we also need OD demands, flight times, aircraft carrying capacities and pickup cutoff and delivery times. Unfortunately, OD demands are proprietary company data, and consequently were not available. Therefore, we used the estimated OD demands and public domain operational data that were all tabulated in Lin et al. (2003). They used the aggregated OD demands with an estimated FedEx market share to estimate FedEx (time-definite) *express* cargo. In practice, FedEx will fill up its carry capacities with (time-indefinite) *general* cargo, when vacant. Thus, they proportionally increased all-pairs

Table 2
Optimal fleet size and routes with various demands

Number	Candidate flight routes	Aircraft	Volume					
			50%	60%	70%	80%	90%	
1	Hong Kong–Subic Bay	MD11	✓	✓	✓	✓	✓	–
2	Penang–Kuala Lumpur–Subic Bay	A310	✓	✓	✓	✓	✓	–
3	Beijing–Shanghai–Subic Bay	MD11	✓	✓	✓	✓	✓	–
4	Bangkok–Subic Bay	MD11	✓	✓	✓	✓	✓	–
5	Malina–Kaohsiung–Taipei–Subic Bay	A310	✓	✓	✓	✓	✓	–
6	Seoul–Osaka–Subic Bay	A310	✓	✓	✓	✓	✓	–
7	Tokyo–Subic Bay	MD11	✓	✓	✓	✓	✓	–
8	Jakarta–Singapore–Subic Bay	A310	✓	✓	✓	✓	✓	–
9	Sydney–Subic Bay	MD11	✓	✓	✓	✓	✓	–
10	Subic Bay–Taipei–Anchorage	MD11	✓	✓	✓	✓	✓	–
11	Hong Kong–Anchorage	MD11	✓	✓	✓	✓	✓	–
12	Penang–Singapore–Osaka–Memphis	MD11	✓	✓	✓	✓	✓	–
13	Shanghai–Beijing–Tokyo–Anchorage	MD11	✓	✓	✓	✓	✓	–
14	Seoul–Tokyo–Anchorage	MD11		✓				–
15	Shanghai–Beijing–Seoul–Osaka–Subic Bay	MD11		✓				–
16	Kaohsiung–Taipei–Subic Bay	MD11						–
		Volume						
			50%	60%	70%	80%	90%	>95%
Fleet size	MD11	8	9	9	9	9	–	
	A310	4	3	4	4	4	–	
Cost (USD)	Transportation cost	2,042,018	2,114,166	2,186,288	2,186,288	2,186,288	–	
	Handling cost	138,186	159,450	184,716	211,108	237,496	–	
	Total cost	2,180,204	2,273,616	2,371,004	2,397,396	2,423,784	Infeasible	
CPU time (s)	2,039.17	2,041.48	2,064.53	2,086.75	2,077.12	2112.8		
Overall payload factor (%)	50.53	53.02	60.25	68.86	77.47	–		

OD express cargo till some of flight segments were saturated. This results in the maximum possible payload with respect to the aircraft design capacities.

The implicit enumeration algorithm is coded in C and tested on PC equipped with 3.0 GHz Pentium IV chip. The computational results are tabulated in Table 2 with additional results of sensitivity analysis on volumes. The computational times on all tests were all in approximately 35 min, a reasonable time for planning tasks. Some observations on solutions: (1) still, A310s and MD-11s are, respectively, the primary aircraft type for the intra-Asia Pacific region and inter-continent services. (2) Depending on demand scenarios, Tokyo, Osaka, and Seoul serve as transit cities for Chinese and South Korean shipments. Tokyo and Seoul are, respectively, the transit cities for China's Asia Pacific and China's inter-continent shipments on 50% and 60% scenarios (Fig. 8), while Osaka is the transit city for South Korea's inter-continent shipments on 50% and over 70% scenarios. (3) On an average, 15–20% of pickups (or deliveries) were loaded (unloaded) by stopping over vehicles (Fig. 9). (4) On an average, 5.19% and 1.33% of the total OD pairs are made center directs and pass through a non-hub transit city with no additional hub rehandling. Their respective demands are 6% and 0.8%. (5) All the solutions are in a perfectly symmetric network, even though the demands are not.

In addition, there are several observations on the results of sensitivity on demands. To accommodate higher demands, if possible, the economical operational plan is to hold the same fleet size (as the demands increased from 50% to 60%). But this strategy requires a shift to larger carrying capacity aircraft. Furthermore, the operations plan requires alterations on the transit cities and changes on some freight paths and flight routes. This experiment revealed a measurable interaction among the design of operations network, and the determination of the freight paths and flight routes. Sequential planning practices as studied in the previous research may result in a sub-optimal solution. In addition, although it may not always be the case, this experiment showed quite large demand range resulted in the same optimal operational plan, in terms of network configuration, freight paths and flight routes. Such a plan may provide the flexibility for uncertain demands without plan alternations.

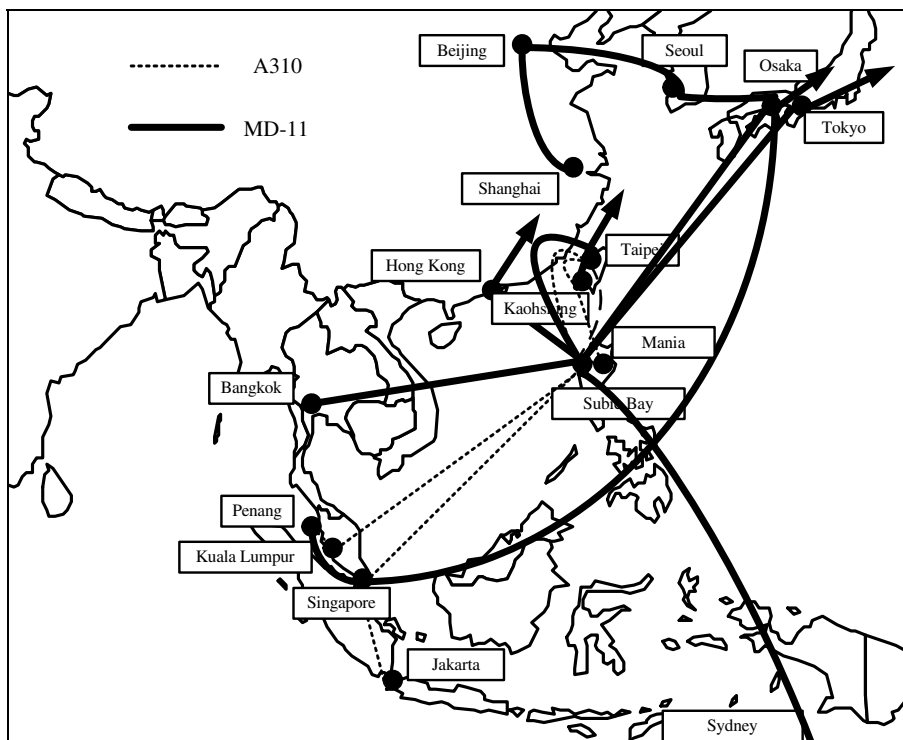


Fig. 8. The optimal solution with 60% of the base demands.

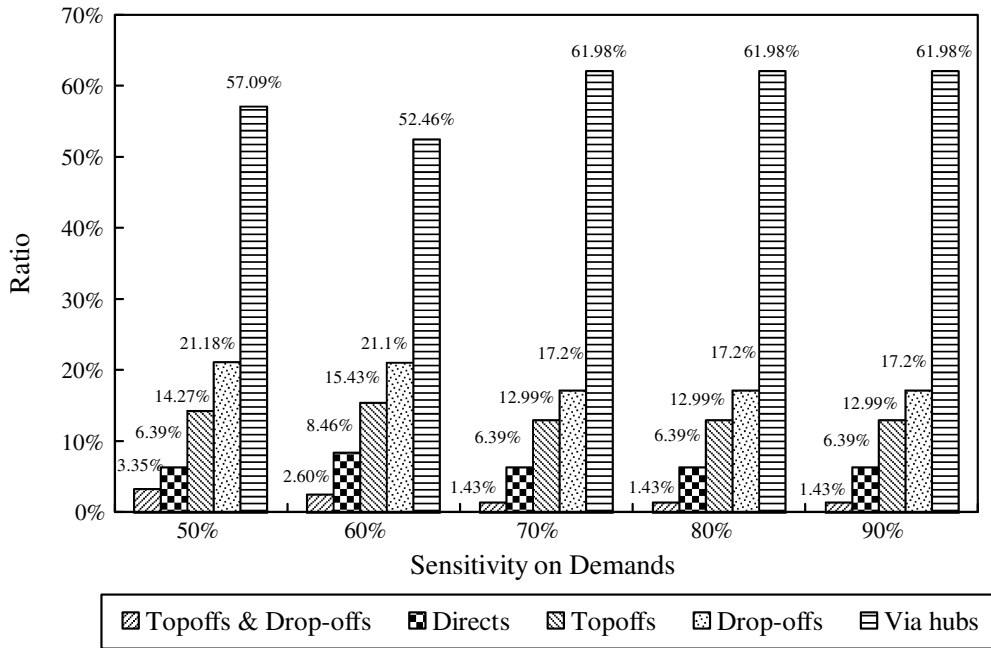


Fig. 9. The volume distribution of types of freight paths.

6. Conclusions and future research

Time-definite freight delivery common carriers, a key third-party logistics service provider in the supply chains, provide door-to-door on time delivery services for small shipment shippers. Centers are points of pickup and delivery while hubs are points of consolidation in hub-and-spoke networks. Three H/S networks, pure, stopovers and center directs, are widely used by the carriers to fit their specific demands on operations which were separately studied. In this research, we proposed a generalized H/S network in a capacitated and directed space-fleet network configuration that is capable to capture and integrate the vehicle and freight operational features of all three H/S networks. This allows carriers to simultaneously evaluate different operational possible freight operations, hub rehandling, toffoffs/drop-offs, transit, and center directs and vehicle routes, stopovers and feeders, to determine the most cost-effective operations plan. Its configuration also allows us to develop a separable algorithmic scheme. To be operational as a planning tool, an implicit enumeration algorithm was developed. It is a separable approach; implicitly enumerating on vehicle routes resulted in an IMCP. IMCP may be solved by an exact algorithm (Lin, 2001), however, the vehicle fleet cost is the dominating costing factor in practice. Thus, in the numerical testing, we instead implemented it with the *least cost* but time feasible paths. We tested the approach in the FedEx AsiaOne air network.

The results are encouraging. (1) They showed that our proposed generalized network structure allow carriers with operational flexibility, center directs, toffoffs/drop-offs, transits and rehandling via hubs, to determine the most cost-effective operations plan. (2) They also revealed the mutual interactions in the design of operations network, that is, the determination of the freight paths and flight routs. The simultaneous operational determination enhances the sequential planning practices as studied in the previous research which may result in a sub-optimum. (3) Our detailed studies on the operations plans showed, on an average of all demand scenarios, the hub rehandling remained the most cost-effective freight handling method 93%, with toffoffs/drop-offs of (15–20%) and center directs the least (7%). It showed that the hub-and-spoke networks provide the economies of scale in the line-haul operations for the carriers who serve numerous small shippers. (4) The sensitivity on demand showed that the demand scenarios may impact the design of operations network, specifically, the designation of transit cities. However, a steady state may exist for a range of demand fluctuation if there is a constant percentage of increase/decrease over all pickups. This allows the carrier to implement a

sustainable operations plan, however, a re-planning may be required, if increases/decreases were varied by locations.

The network design model and the proposed generalize H/S network may well be applicable to and adopted by the passenger transport industry. Since, the passengers may transit flight-to-flight on their itineraries by themselves; no requirements of singular path for each OD pair are imposed. As a result, the embedded network subproblem becomes a multi-commodity min-cost flow with no integral constraints. Various algorithms are available and can be applied, such as Lagrangian relaxation, Dang–Wolfe decomposition, primal-dual, etc. (Ahuja et al., 1993). One further extension of this research is to use link-based approach to construct, instead of exogenously predetermined vehicle routes. This allows planners to explore all the possible vehicle routes in pursuit of optimality. Some of the network techniques, such as minimum spanning tree to determine vehicle route plan may be incorporated.

Acknowledgement

This research was supported by a grant from the National Science Council, Taiwan, ROC under Grant NSC-92-2416-H-006-004.

References

- Ahuja, R.K., Magnanti, T.L., Orlin, J.B., 1993. *Network Flows*. Prentice Hall, New Jersey.
- Armacost, A.P., Barnhart, C., Ware, K.A., 2002. Composite variable formulations for express shipment service network design. *Transportation Science* 36 (1), 1–20.
- Bryan, D.L., O’Kelly, M.E., 1999. Hub-and-spoke networks in air transportation: an analytical review. *Journal of Regional Science* 39 (2), 275–295.
- Current, J.R., 1988. The design of a hierarchical transportation network with transshipment facilities. *Transportation Science* 22 (4), 270–277.
- Ebery, J., 2001. Solving large single allocation p-Hub problems with two or three hubs. *European Journal of Operational Research* 128, 447–458.
- Ernst, A.T., Krishnamoorthy, M., 1998. An exact solution approach based on shortest-paths for p-hub median problems. *INFORMS Journal on Computing* 10 (2), 149–162.
- Groothedde, B., Ruijgrok, C., Tavasszy, L., 2005. Towards collaborative, intermodal hub networks: a case study in the fast moving consumer goods market. *Transportation Research Part E* 41, 567–583.
- Kim, D., Barnhart, C., Ware, K., Reinhardt, G., 1999. Multimodal express package delivery: a service network design application. *Transportation Science* 33 (4), 391–407.
- Kuby, M.J., Gray, R.G., 1993. The hub network design problem with stopovers and feeders: the case of Federal Express. *Transportation Research Part A* 27, 1–12.
- Leung, J.M., Magnanti, T.L., Singhal, V., 1990. Routing in point-to-point delivery system: formulations and solution heuristic. *Transportation Science* 24 (4), 245–260.
- Lin, C.-C., 2001. The freight routing problem of time definite freight delivery common carriers. *Transportation Research Part B* 35 (6), 1–23.
- Lin, C.-C., Chen, S.H., 2004. The hierarchical network design problem for time-definite express common carriers. *Transportation Research Part B* 38 (3), 271–283.
- Lin, C.-C., Lin, J., 2007. The multistage stochastic integer load planning problem. *Transportation Research Part E* 43 (2), 143–156.
- Lin, C.-C., Lin, Y.J., Lin, D.-Y., 2003. The economic effects of center-to-center directs on hub-and-spoke networks for air express common carriers. *Journal of Air Transport Management* 9 (4), 255–265.
- O’Kelly, M., Bryan, D., 2002. Interfacility interaction in models of hub and spoke networks. *Journal of Regional Science* 42 (1), 145–164.
- O’Kelly, M., Lao, Y., 1991. Mode choice in a hub-and-spoke network: a zero-one linear programming approach. *Geographical Analysis* 23 (4), 283–297.
- Sohn, J., Park, S., 2000. The single allocation problem in the interacting three-hub network. *Network* 35 (1), 17–25.