Routing Traffic at Hub Facilities

Morton E. O'Kelly

Abstract Flows arriving at a hub or a transshipment facility may need to be switched from one path to another to complete their journey. These transfer aspects of hub and spoke systems are widely recognized as a hindrance to efficient completion of transit trips. For example, time-consuming delays at transfer points for bus passengers are a major reason for poor levels of service between some nodes when the origin and destination are on different network lines. Such transfers also arise in multimodal interaction systems. This paper outlines a simple notation and analytical framework for optimizing flows within a hub (i.e. at nodes with transfers) and discusses several variants of the problem. This paper addresses prototype models for the efficient allocation of resources to facilitate the operation of interactions at the hub. The paper is primarily a conceptual and methodological overview, but well-recognized existing optimization models are suggested as being useful for some of the related tasks. Small numerical examples are used to illustrate some of the ideas.

Keywords Optimal routing · Efficiency · Surveillance · Dock-door assignment · Hub network · Transfers

1 Introduction

Despite a great deal of attention paid to hubs in spatial networks, it appears that there are unexplored issues in the analysis of flows within the hub itself. This paper outlines a simple typology, notation, and analytical framework for routing traffic at

M. E. O'Kelly (✉)
Department of Geography, The Ohio State University, 1036 Derby Hall, 154 N Oval Mall, Columbus, OH 43210, USA
e-mail: okelly.1@osu.edu

transfer nodes. The fundamental idea is that flows between network nodes interconnect at various types of hubs. Transfer aspects of hub and spoke systems are widely recognized as a hindrance to efficient completion of transit trips. For example, time-consuming delays at transfer points for bus passengers are a major reason for poor levels of service between some nodes when the origin and destination are on different network spokes. In discussing models for this situation, this paper adapts ideas from diverse types of transportation situations, such as subway systems, internet backbone peering and connectivity, logistics facilities, and border crossings. The ideas here are flexible enough to suggest a framework for further research on a variety of interesting optimization problems that might arise in connection with a hub: the minimization of transfer delay, the design of strong surveillance schemes, or the decision to link certain paths directly.

By focusing attention on intra-hub connectivity, some new aspects of hub network efficiency may be realized. There is a precedent for this scale of analysis in Koopmans and Beckmann (1957), who looked at interdependent flow and location decisions in plant layout and showed that they lead to quadratic programs. Their work inspired further exploration of internal economies of scale (Reiter and Sherman 1962). By analyzing internal linkages at hubs, this paper opens a series of questions about network efficiency, in the same way that internal economies of scale are studied in industrial economics.

Three approaches to this issue are apparent: a descriptive (and data intensive) study could be carried out to recover historical data from the operational records of the throughput conditions at a hub. A second approach would be to derive generalized models of the flows from spatial interaction theory (Fotheringham and O'Kelly 1989; O'Kelly 1986). A third approach, and the main one adopted in the present paper, is to combine the expected flows with a network design problem to potentially optimize certain aspects of the operational management of the transport system (see also Campbell 1994; Crainic and Kim 2007). This paper deals with a relatively simple case where lines come into the hub at distinct points, and there the flows (passengers, packages, etc.) are either passed through on that same line or are switched to another line.

2 Typology of related transfer problems

In the present paper it is assumed that *exogenously given interactions* flow from origins to destinations, via hubs which act as switching points. Focusing on the hub facility itself, note that the *spokes* that come to the facility can be regarded as either (1) terminating at the hub or (2) part of a line passing through the hub. In both cases however, exogenously given flows may pass through the facility, either after making a transfer (Case 1), or on the same line (Case 2). These types of interchange and transfer problems give rise to two different analytical cases which are discussed next.

Case 1: Hub Terminus The first case is exemplified by spokes that *terminate* at a hub; i.e. doors at a cross-dock, airline gates at an airport, or transit routes that enter an urban area and stop at a terminus. In these cases the flows need to transfer to a connecting segment to continue their journey. The facility in each of these examples is the cross-dock, the airport, or the city center. Assuming that several terminating

spokes connect *at the facility*, one then has the problem of optimizing the organization of flows from one door/gate/terminal to another. To emphasize the need to connect stages of the trip at the facility, these intra-facility connections will be referred to as "links" and distinguished from lines and spokes. The problem in this case is typically one of minimizing transaction costs.

Case 2: Hubs with through connections The second case is exemplified by transfer hubs on bus, train, or subway lines where several lines (e.g. red, green, blue etc.) pass through the station. These lines meet at the common facility, perhaps occurring at various levels (i.e. a third dimension). Someone on a red line destined for another red line station stays on the train and makes no transfer and has a "through" connection. On the other hand a red line passenger needing to transfer to the blue line, exits the red train at the station, moves to the other platform (which may be some distance away) and then boards the blue train. The questions that arise in this case tend to involve the volume of required transfers and decisions about connections.

Although the two cases and the literatures that relate to them are perhaps rather separate, a number of useful cross-fertilizations occur when the problems are juxtaposed. The paper poses a variety of optimization problems for making efficient joins between incoming and outgoing flows at the hub. One particular optimization problem occurs when deciding which spokes may be bridged together allowing for a smoother transfer or more favorable connection from one spoke to another; in essence this converts two (or more) terminating spokes (Case 1) into linked or connected lines (Case 2). In a planning context this might be the problem of deciding which inbound and outbound spokes should be given a favored path through the terminal. Operationally, there are of course logical and physical limitations on the spokes that can be joined: sharp changes in line direction or level being impossible. Spokes, by definition, terminate at the hub (just like wire spokes attached to a hub on a bicycle wheel). However, whenever spokes can be interconnected to form continuous lines, the decision problem would be to select pairs of spokes to join. One variant, with special relevance for Case 2, is whether the spokes contain multiple nodes: invariably such *feeder* lines have multiple station stops on either side of the transfer station. The distinction between independent and interdependent nodes (such as might occur on regular spokes and feeders respectively) and a distinction between terminating and through lines gives rise to as many as four special cases, of which two are considered here in detail, as shown below.

	Terminating spokes	Through lines
Independent nodes	Case 1	See Case 1
Interdependent nodes (feeders)	See Case 2	Case 2

The two instances that are not explicitly dealt with (single nodes on through lines, and feeders arriving at terminals) are both useful potential extensions, though they would seem to be readily handled as parts of the two cases that are dealt with. The nodes on through lines for example, lead to expected inter-gate flows, and it seems that whether this is a feeder blending several flows, or a single node, the analysis is basically the same. The case of feeders arriving at terminals is in fact similar to Case 1 so long as the feeder flows are bundled/aggregated appropriately.

Thus, while terminating spokes can have independent nodes and several interdependent nodes (as on a feeder line), the primary focus here is on the problem of terminating spokes with single nodes (Case 1). Through lines can also have independent nodes and feeder lines, and in this paper the focus is on the essential problem (Case 2) of through lines with multiple feeder nodes.

3 Related literature

While this paper is not a review of the voluminous research on network routing, some existing papers that have examined parts of the problems outlined above are cited here as an introductory guide to further conceptual development. Kleinrock (1964) simulates the flows in *communications networks* between lines that join at a central node; the primary focus of his approach is to emphasize stochastic models for queuing. He develops considerable numerical insight into the impact of network structure on the throughput of flows. Kara and Tansel's (2001) and Yaman et al.'s (2007) work on the 'latest arrival' problem is a merger of the hub flow problem and the earlier communications network ideas of Kleinrock. When there are multiple destinations for flows in each inbound vehicle, one cannot dispatch an outbound vehicle until the latest arrival containing material for that destination occurs. In a related case airlines also hold departures to accommodate late incoming connections (Lim et al. 2005). In distinction to this however, one notes that the waves of transit vehicle departures have to adhere to a schedule and so cannot usually wait for the last inbound passenger (i.e. in bus or train networks). The idea of optimally pulsing or organizing the waves of interconnecting paths is beyond the scope of this paper though clearly there is ample opportunity for further research to combine the network concepts here with the large body of literature on optimal transit scheduling (see Chang and Hsu (2001) and Chowdhury and Chien (2002) for representative samples). A recent detailed treatment of the synchronization of flows is in Bobzin (2005).

In many models of hub and spoke networks, the links to the hub are typically just a single path from an origin. However, Kuby and Gray (1993) expand this idea to include hub networks with feeders, especially for air freight, where the useful notion of feeder is introduced to allow multiple pick-up points along a spoke entering a hub. While this is clearly too inconvenient for most air passenger applications, it does make a great deal of sense for the amalgamation of flows in surface transportation networks. It also gives rise to the kind of flow aggregation problem that could potentially strain the capacity of the link and the terminal. Although this present paper does not focus on capacity or flow estimation per se, it is important to note that in modeling the flows arriving at a hub from a spoke, one needs to be cognizant of the volume of aggregated flows.

A hub node can also be used to represent an intermodal facility such as a connection between ground and air transportation (Hall 2001; Crainic and Kim 2007). In the area of logistics and facility layout, cross-dock facility planning is very relevant to the present situation. When flows meet at a location it is often required to optimize assignments to particular gates to minimize intra-facility costs. Researchers on this topic derive a special type of quadratic assignment problem called the *dock-door assignment* problem (Tsui and Chang 1992; see also review of assignment

problems in general by Pentico 2007). The typical cross-docking system occurs for example in a mail distribution center with receiving doors and shipping doors (Oh et al.'s 2006). "The assignment of destinations to shipping doors, clustering of destinations to form groups, and determination of the number of groups are major operational problems directly related with the efficiency of the center." While there are strong similarities between the goals of the present paper and Oh et al.'s (2006) objective of minimizing travel within the center, the present paper suggests variations on the model that might be adapted to intra-hub movements of a more general nature. One distinction between the typical cross-dock model and the present case is that it is envisioned that multiple origins and destinations connect to each of the terminal gates. These multiple spokes can arise either from the feeder functionality, or the hub-like amalgamation of multiple incoming flows through a particular gate.

Finally, a series of developments in the hub location literature (Campbell 1994) guide solution procedures for network optimization involving hub activity levels: these include simply enumerating the levels of activities at the hub (O'Kelly 1986), to a quadratic assignment-type problem, where a variety of optimization techniques have been applied to provide a heuristic driven approach to what can be an unwieldy problem (Skorin-Kapov et al. 1996).

In summary then, the literature suggests that this kind of problem domain will entail integer programs, probably with quadratic objective functions. In fact the full complexity of the interdependent strategic and operational decisions on clustering and linking paths is likely to require some variants of the classical quadratic assignment or generalized assignment problem. Along with these challenges, the routing of flows is expected to give rise to some substantively interesting problems.

4 Case 1: Optimal allocation and matching

In considering Case 1 (optimization of connections), the important decision variables are (1) the ways in which the incoming and outgoing spokes are connected at the hub gates, (2) which spokes (representing nodes) connect to the same gates, and (3) the arrangement of these allocations to minimize the amount of inter-gate costs. The scale or distance between gates at the terminal varies somewhat by application. When the gates are separated and require longer walks between stations (e.g. St. Pancras to King's Cross in London), there is simply a larger intervening movement between the facilities, but the connection between the lines is still appropriately represented as a form of movement between lines or stations. It is not clear from the literature what if any impact the length of the intra-facility linkage might have other than the obvious one that longer links pose greater costs and delays.

The optimal allocation and matching problem is related to the task of selecting improved interline connections. As will be seen in this section, many location and optimization issues that arise under the topic of this paper are variants on the quadratic assignment problem; the organization of connections at the hub play a key role in the throughput. Imagine a set of origins connected to one side of the facility (IN), and the destinations connected on the other side (OUT). The flow connects through inbound and outbound gates at the central hub. Each origin and each destination needs to be handled by a particular pair of hub gates, k and m

respectively. The material is transferred between the gates without intervening storage, but with a cost of movement. That cost is assumed to be small relative to the total cost of accessing the hub from the originating point; the flow is presumed to have arrived at the exterior of the hub, and ready for assignment to a gate.

Although the feeder problem could be applied to this part of the paper also (a la Kuby and Gray 1993), here we focus on a simple version of the problem, as all the nodes are considered to be *independent*, and thus each spoke has one associated node; this further leads to the fact that one can refer interchangeably to node or spoke i (as each spoke has just one unique node). The issue of combining these ideas with a more complete treatment of feeders, realistic flows, and so on, is left for future research.

A question arises then as to the optimal allocation of nodes to gates. The merit of assigning node i to gate k and node j to gate m depends on whatever else is assigned to those same gates, especially if each node is assigned to a unique gate. This is a key similarity to the so-called single-gate assignment constraint in the hub literature. The ultimate assignment of node i to k will of course depend on the components of its interactions, and the assignment of the nodes with which it interacts: once the material reaches the hub at gate k it is then routed through the facility to the connections for node j etc. It is this interdependence of the routing decision variable for the inbound and outbound gates that make this a fundamentally quadratic problem (routing depends on the product of X_{ik} and X_{jm}). However, the product of the two assignment variables is sometimes represented as a linear term with four subscripts: these routes are defined by R_{ijkm} which is the routing variable from i to k and from m to j.

In very broad terms, the objective is to assign the incoming node or spoke to a gate, and assign the outgoing node or spoke to a gate, to minimize the total movement of the transfer operation. Unlike some prior research, this model simultaneously assigns the inbound and outbound spokes under a variety of protocols, noting the fact that an incoming flow from an origin can contain items for multiple destinations [see other examples applied to less-than-truckload shipping by Bartholdi and Gue (2000, 2004) and Oh et al. (2006)].

Assuming a desire to optimize the allocation of origin node i to gate k and the allocation of destination node j to gate m, the following integer program chooses $\{X_{ik}^O, X_{jm}^D, R_{ijkm}\}$ to

$$\text{Minimize} \sum_{i \in O} \sum_{j \in D} \sum_{k \in \text{IN}} \sum_{m \in \text{OUT}} W_{ij} R_{ijkm} \left[C_{ik}^O + C_{km}^H + C_{jm}^D \right] \quad (1)$$

Subject to:

$$\sum_{k \in \text{IN}} X_{ik}^O = 1 \qquad \text{for all } i \in O \quad (2)$$

$$\sum_{m \in \text{OUT}} X_{jm}^D = 1 \qquad \text{for all } j \in D \quad (3)$$

$$\sum_{m \in \text{OUT}} R_{ijkm} - X_{ik}^O = 0 \quad \text{for all } i \in O, j \in D, k \in \text{IN} \quad (4)$$

$$\sum_{k \in \text{IN}} R_{ijkm} - X_{jm}^D = 0 \quad \text{for all } i \in O, j \in D, m \in \text{OUT} \quad (5)$$

$$X_{ik}^O, X_{jm}^D, R_{ijkm} \in \{0, 1\}. \quad (6)$$

The data associated with the problem are primarily driven by W_{ij}, the pre-specified flows between the origins and destinations. The variables to describe these decisions are as follows:

X_{ik}^O the allocation of origin node i to inbound gate k; every unique origin is assigned to one gate; O is the set of origin nodes.

X_{jm}^D the allocation of destination node j to outbound gate m; D is the set of destination nodes.

Constraint (2) ensures that every unique origin is assigned to one gate. Constraint (3) ensures that every unique destination is assigned to one gate. In constraints (4) and (5) the routing variables R_{ijkm} are constrained by the allocation variables, following a formulation in O'Kelly et al. (1995).

The unit cost coefficients associated with these movements are as follows: C_{ik}^O is the cost to assign origin i to inbound dock gate k; C_{jm}^D is the cost to assign destination j to outbound dock gate m; and C_{km}^H is the inter-gate cost between gates k and m at the hub.

The objective can then be re-written as follows (O'Kelly 1987) where the range of summation of i, j, k, m is simplified (e.g. $\Sigma_{i \in O}$ is written as Σ_i)

$$\sum_i \sum_j \sum_k \sum_m W_{ij} R_{ijkm} \left[C_{ik}^O \right] + \sum_i \sum_j \sum_k \sum_m W_{ij} R_{ijkm} \left[C_{jm}^D \right] \\ + \sum_i \sum_j \sum_k \sum_m W_{ij} R_{ijkm} \left[C_{km}^H \right] = \tag{7}$$

$$\sum_i \sum_k C_{ik}^O \sum_j W_{ij} \sum_m R_{ijkm} + \sum_j \sum_m C_{jm}^D \sum_i W_{ij} \sum_k R_{ijkm} \\ + \sum_i \sum_j \sum_k \sum_m W_{ij} R_{ijkm} \left[C_{km}^H \right] = \tag{8}$$

$$\sum_i \sum_k O_i X_{ik}^O C_{ik}^O + \sum_j \sum_m D_j X_{jm}^D C_{jm}^D \tag{9}$$
$$+ \sum_i \sum_j \sum_k \sum_m W_{ij} R_{ijkm} C_{km}^H$$

and the three components in Eq. (9) are referred to as P1 (the inbound cost of moving materials from the terminal entrance to the assigned gates), P2 (the outbound cost of moving materials from the assigned gates to the terminal exit), and P3 (the inter-gate costs or distances) respectively. Note that $O_i = \Sigma_j W_{ij}$, and $D_j = \Sigma_i W_{ij}$ are outflows from origins and inflows to destinations respectively. It is assumed that the costs of getting from the origin and destination to the hub are calculated outside the model, and in terms of orders of magnitude they are likely to be much larger than the intra-facility costs.

Depending on the number and configuration of incoming and outgoing lines, the potential gate locations may or may not all be active in a solution. The decision as to the active subset of the gates to use could result in some gates getting no assignment. This suggests a generalized form of quadratic assignment where the active gates and the flows between them are determined, thereby producing a configuration for the

intra-hub flows and linkages. If one wished to force the use of all the available gates, for reasons of balance, the following constraints could be added:

$$\sum_i X_{ik}^O \geq 1 \qquad\qquad \text{for all } k \in \text{IN} \qquad\qquad (10)$$

$$\sum_j X_{jm}^D \geq 1 \qquad\qquad \text{for all } m \in \text{OUT} \qquad\qquad (11)$$

Adding constraints (10) and (11) gives rise to an interesting variant on the problem, where we might explore the increased cost associated with forcing gates to have an assignment. It also turns out in the numerical examples to be essential to enforce the binary restriction on the allocation variables X^O and X^D.

5 Case 2: Optimal interline connection

To consider Case 2 (interline connections), examine a network with a number of spokes entering a central switching point (see Fig. 1), so the flows from the nodes on the spokes reach nodes on other spokes by traversing the hub at the center of the network. If permitted by the network protocol, which pairs of spokes should be joined so as to have "through" connections? One simple optimization rule would be to join the busiest pairs, assuming a feasible connection. However, from the point of view of interaction through a switch or a border crossing, the apparent simplicity of the calculations hides many nuances which need to be considered in more detail [as a simplifying assumption, the flows proceed from left to right (i.e. from set O to set D) and so in this case, the nodes in O cannot interact with each other; this is not a major limitation however as the sets O and D can be expanded to include all interacting places].

Transit routes coming into a terminal usually have many stops. Each of those various stations gives rise to potential interactions with nodes along the other lines, as well as with the central terminus. The terminating point for the portion of the journey on one line links to the origination for the other line: in essence then the transfer appears to occur within one facility. It is primarily the traffic flow through the hub that is the subject of the analysis. The interactions are weighted by the amount of flow between nodes on spokes and the flows on intra-hub links are then the accumulation of all the traffic that crosses those connections.

An interesting problem is to decide if any spokes ought to be bridged together so that the passengers transiting the hub can stay at a brief station stop and then continue on the same line to their destination, (such as might occur at Union Station in Toronto). In a related area, route choice in a subway system with transfers (Lee and Lee 1998) is considered for quite complex subway systems (such as in Seoul South Korea) where dozens of interconnected rail lines provide regional coverage and connection. A comprehensive atlas of such systems has recently been provided by Ovenden (2007). Estimating travel time in these networks typically requires some measure of the probable number of transfers. Optimizing connectivity in such
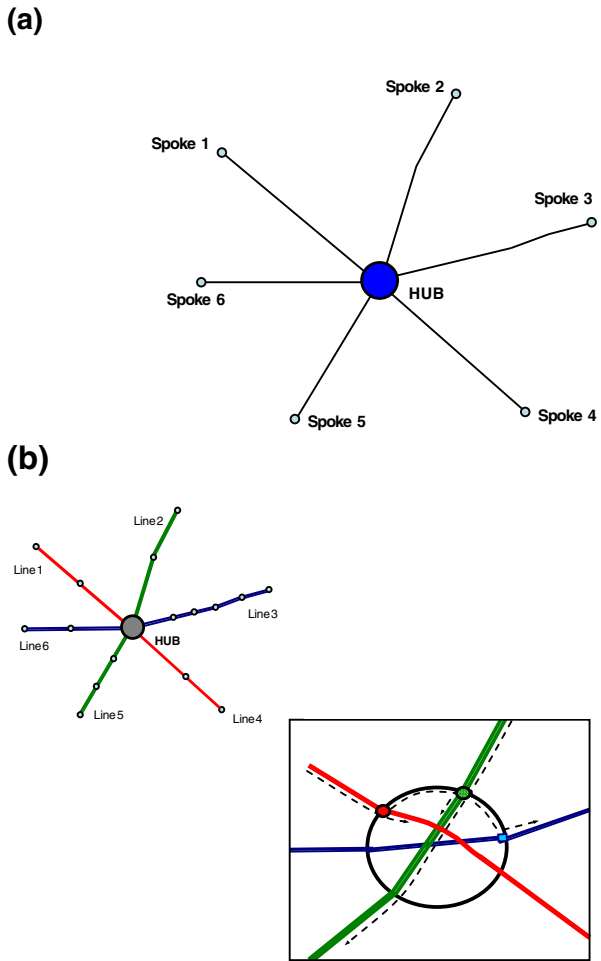
systems is equivalent to eliminating the necessity for a transfer between some in and
outbound lines. One simple ad hoc way to decide these joins would be to mesh
together the segments that have the highest interaction (not necessarily the largest
numbers of pairs of interactions). For example if lines 2 and 5 in Case 2 of Fig. 1
were to be designated the 'green' line all five stations could have easier access to
each other through the hub. It is the model problem to decide if other pairs of lines
could produce a better pairing—for example the three stops on spoke 5 could be
better joined to the five stops on spoke 3 depending on the flows.

The *interline improvement problem* can then be seen as an application of this
measurement scheme. Suppose that there is a fixed budget for upgrading interline
connections. A simple knapsack problem would be suited to the task of selecting the
subset of the activities that maximizes return subject to a cost constraint for the
improvements. If the link between the spokes terminating at k and m respectively, is

modeled as having benefit or flow W_{km} and cost C_{km}, the decision problem might be to choose $\{X_{km}\}$ to

$$\text{Max} \qquad \sum_k \sum_m X_{km}\, W_{km} \qquad (12)$$

$$\text{Subject to} \qquad \sum_k \sum_m X_{km}\, C_{km} \leq C \qquad (13)$$

$$X_{km} \in \{0, 1\} \qquad (14)$$

As this is a simple variant of the knapsack problem (X binary), we could essentially rank the interlink flows in terms of their ratios of W_{km}/C_{km} and progress through these in a greedy fashion to add the improvements with the greatest "bang for the buck." Infeasible connections could be blocked by setting a large cost coefficient. As is usual with knapsack problems, ranking flows in order of these ratios does not usually produce an optimal solution, so the task would have to be completed with some branch and bound steps (Nemhauser and Wolsey 1988). The knapsack problem is a well studied subroutine in many optimization texts and will be simply incorporated here by reference (Martello and Toth 1990).

The problem with this knapsack is that it does not deal with matching or assignment constraints, and these are needed to ensure that each gate is paired with at most one other gate. Disregarding intra-facility costs and capacity, one sees that a feeder arriving at a hub contains all the traffic that originated on the feeder nodes destined for other nodes on other feeders. Thus a simple aggregation of all the flows gives the expected interaction between gate k and m. A direct approach is to preprocess the W_{ij} flows such that all flows from the spokes assigned to gates k and m are found by summing

$$W_{km} = \sum_{i \in \text{spoke } k} \sum_{j \in \text{spoke } m} W_{ij} \qquad (15)$$

Then, the following assignment problem could be solved to determine the best links to connect:

$$\text{Max} \qquad \sum_k \sum_m X_{km}\, W_{km} \qquad (16)$$

$$\text{Subject to} \qquad \sum_k X_{km} = 1 \qquad (17)$$

$$\sum_m X_{km} = 1 \qquad (18)$$

$$X_{km} \in \{0, 1\} \qquad (19)$$

6 Solution methods

Generally, for small test cases, the models can be explored with commercial software such as CPLEX. However, because the problems outlined above (especially for

Case 1) can become large for practical instances, it is important to devise some additional solution techniques (see also Oh et al. 2006). Lower bounding is a common device in optimization. It is used to set a floor under the objective function: while we have no interest in an underestimate of the true cost per se, we do benefit from knowing that the optimal solution *must be* greater than the lower bound.

One possible heuristic approach is to work out a decoupled version of the allocation variables, using modified coefficients. Having solved by inspection for those allocations, the original objective can be evaluated, and a bound on its accuracy established. Details of this substitution follow from exploiting the triangle inequality (for the relationship between C_{ik} and C_{km} for example). This leads to two substitutions for the inter-gate assignment cost, leading in turn to a lower bound.

Let C_{im}^M = distance from origin i to the *outbound* gate. It is clear from the triangle inequality that $C_{im}^M \le C_{ik}^O + C_{km}^H$. This observation leads to the first substitution

$$C_{km}^H \ge C_{im}^M - C_{ik}^O. \tag{20}$$

Let C_{jk}^K = distance from destination j to *inbound* gate. Again, it is clear from the triangle inequality that $C_{jk}^K \le C_{km}^H + C_{jm}^D$. This observation leads to the second substitution

$$C_{km}^H > C_{jk}^K - C_{jm}^D. \tag{21}$$

Since there are two alternative sets of substitutions leading to two contenders for the lower bound, the *larger* resulting objective function becomes the largest lower bound. Solving the problem (to be derived below) with these artificial substitutions for the inter-gate costs can lead to a new set of allocations. To distinguish these from the primal variables, use Z_{ik}^O which is an allocation variable from origin to gate k (decision variable in relaxation) and Z_{jm}^D an allocation variable from destination to gate m (decision variable in relaxation).

Approaching the objective from the first substitution (20) into P3 we obtain the following result: LB1 $= \sum_j \sum_m Z_{jm}^D \sum_i W_{ij} \left[C_{jm}^D + C_{im}^M \right]$

Proof From (20) $C_{km}^H > C_{im}^M - C_{ik}^O$ and inserting in the objective:

$$\text{P3} > \sum_i \sum_j \sum_k \sum_m W_{ij} Z_{ik}^O Z_{jm}^D \left[C_{im}^M - C_{ik}^O \right]$$
$$= \sum_i \sum_j \sum_m W_{ij} Z_{jm}^D \left[C_{im}^M \right] - \text{P1}$$

combining P1, P2, and lower bound estimate of P3, noting the ability to cancel P1:

$$\text{LB1} = \sum_j \sum_m \sum_i W_{ij} Z_{jm}^D C_{jm}^D + \sum_i \sum_j \sum_m W_{ij} Z_{jm}^D \left[C_{im}^M \right]$$
$$= \sum_j \sum_m Z_{jm}^D \sum_i W_{ij} \left[C_{jm}^D + C_{im}^M \right] \qquad \blacksquare$$
$$\tag{22}$$

Define $C_{jm}^J = \sum_i W_{ij} \left[C_{jm}^D + C_{im}^M \right]$. The first lower bound based allocation can be found from setting $Z_{jm*}^D = 1$ for m such that m^* is the index with $C_{jm*}^J < C_{jm}^J$ over all m, and repeated for each j. This provides a feasible allocation of the destination nodes to outbound gates.

Approaching the objective from the second substitution (21) into P3 we obtain the following result (similar proof omitted):

$$LB2 = \sum_i \sum_k Z^O_{ik} \sum_j W_{ij} \left[C^O_{ik} + C^K_{jk} \right]. \tag{23}$$

Define $C^I_{ik} = \sum_j W_{ij} \left[C^O_{ik} + C^K_{jk} \right]$. The second lower bound based allocation can thus be found from setting $Z^O_{ik*} = 1$ for k such that $k*$ is the index with $C^I_{ik*} < C^I_{ik}$ over all k, and repeated for each i. This provides a feasible allocation of the origin nodes to the inbound gates.

Having assigned the Z^D_{jm} and Z^O_{ik} variables according to the *lower bound* coefficients, it is possible to evaluate the primal objective from the Z variables. Of course this solution provides an initial solution (labeled as a BKS), but may still leave a gap between the BKS and the lower bound. It remains then to gauge the quality of the lower bound in comparison to the BKS.

## 7 Numerical results

7.1 Case 1

The flow matrix illustrated in Table 1 has ten basic origins treated separately, and eight destinations. There are four inbound hub gates, and four outbound hub gates. The problem is to determine the allocation of the nodes or spokes to the gates at the hub.

In the present context the models are solved to optimality using CPLEX. Some comparative bounds are also discussed briefly later in this section.

The numerical solutions are interesting because the assignments provide a routing solution for each OD pair. A sample optimal solution for this 10×8 problem is shown in Fig. 2. Note that the gate assignments of the ten origins may be denoted [1 1 2 2 3 3 4 4 4 4] and of the eight destinations may be denoted [1 1 2 2 3 4 4 4]. The resulting inter-gate flows are easily found by adding up the interactions and are

Table 1 Sample input data for 10×8 problem	Origin	Destination							
		a	b	c	d	e	f	g	h
	1	1	2	3	4	5	6	7	8
	2	2	0	4	0	0	4	0	0
	3	1	2	5	7	9	11	13	15
	4	5	0	6	0	0	6	0	0
	5	1	2	7	9	12	15	18	21
	6	6	0	8	0	0	8	0	0
	7	1	2	9	12	16	20	24	28
	8	7	0	10	0	0	10	0	0
	9	8	2	11	15	20	25	30	35
	10	1	0	12	0	0	12	0	0
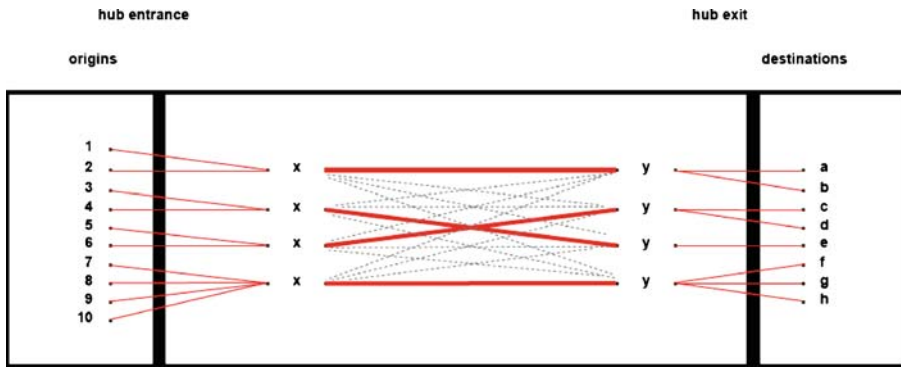
FROM / TO	a,b	c,d	e	f,g,h
1,2	5	11	5	25
3,4	8	18	9	45
5,6	9	24	12	62
7,8,9,10	21	69	36	184

Fig. 2 10×8 Node example. Notes: The *thicker lines* represent upgraded inter-gate linkages. The 4×4 table shows the interactions between the gates, with the upgraded inter-gate linkages highlighted

shown on Fig. 2. Examining the resulting solution shown in Fig. 2, note that all four inbound and outbound gates are used. Note that the four incoming spokes and the four outgoing spokes might be joined to form through connections, one possible solution would be to solve a small assignment problem [as in Eqs. (16), (17), (18), and (19) above]. In this case joining 1 to 1, 2 to 3, 3 to 2 and 4 to 4 is shown to be the solution that maximizes flows on the connections and those flows (5, 9, 24, 184) are emphasized on the diagram.

Once the model is set up as an optimization task in standard software, experiments can be performed to test the result of adding extra constraints (like the added generalized assignment constraints) or modified data. Two added variants on the test data set are now introduced. The problems are set up to consider a number of variants on the location and clustering of the spokes and variations on the inter-facility distance.

First, various spatial clusters of the ten inbound and eight outbound nodes are taken into consideration: bundling the ten origins and eight destinations into groups, as shown in Fig. 3 where there are two, three, and four clusters respectively. Flows are regarded as arriving from *separate* originating sources, but they arrive near the hub at the same physical location (like having north and south regions send their component flows to the hub through a north and south entrance). From that entrance or arrival point the nodal flows are assigned to gates to minimize intra-facility distances. Unlike the nodes on feeders, considered below in Case 2, these flows may have distinct allocations.

Second, the fact that the intra-facility distances may be more or less important is simulated by considering the in- and out-bound sides of the facility to be progressively closer together: i.e. at a distance of 1.0, 0.8, 0.6, 0.4, 0.2, and 0.0
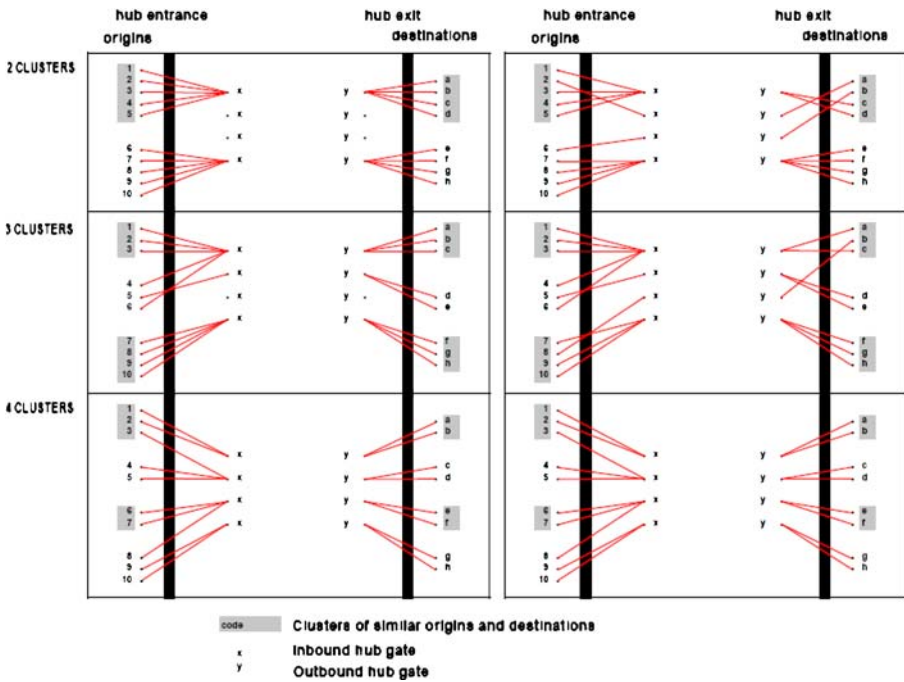
Fig. 3 Case 1 results for 2, 3, and 4 clusters of flows

units apart with this last case obviously signifying a long narrow linear facility. A total of 18 cases are therefore considered [three node cluster configurations × six facility layouts], and these are run without and with the gate-usage restriction. Numerical results for the 18 test runs are shown in Table 2. These show that the objective function is somewhat higher (of course) as a result of the added gate-usage restrictions, but more interesting is the explicit routing instructions that emerge as needed to force certain gate assignments.

The models are run for a variety of intra-facility distance parameter choices, one of which (intra-facility distance of 0.4) is shown in Fig. 3. In the top panel (i.e. results for two clusters) because the base model solution leaves several gates unused, the imposition of a gate usage rule requires at least a minimal amount of shifting of the allocations to insure that gates 2 and 3 are used. This also occurs in the middle panel which shows the results for three clusters; since every gate is not used, some adjustments are needed to accommodate constraints (10) and (11). Finally, in the bottom panel of that figure which shows results for four clusters, and because the model allocations actually use all the gates the extra constraint has no effect.

Comparative empirical results for the bounds on the objective functions are shown in Table 3 for the basic 10×8 example. Knowing the optimal solution in the present case allows us to gauge the quality of the lower bound, as well as determining if a reasonably good solution was derived as a by product of the Z variables. The interesting result of this small example experiment is that the lower bound quantity is tight compared to the initial BKS (ranging from 2.8% to 6.8%). However as the result of evaluating the primary objective function, note that an

Table 2 Case 1 for 18 test data instances

Number of clusters	Inter-gate distance	Without assignment constraints	With constraints (10) and (11)
2	1.0	3,050.03	3,097.60
2	0.8	3,039.90	3,086.54
2	0.6	3,036.63	3,082.96
2*	0.4	3,039.99	3,086.88
2	0.2	3,049.78	3,098.31
2	0.0	3,065.75	3,117.24
3	1.0	2,802.66	2,831.40
3	0.8	2,801.54	2,829.91
3	0.6	2,807.96	2,836.28
3*	0.4	2,820.76	2,849.49
3	0.2	2,841.39	2,869.46
3	0.0	2,869.89	2,895.33
4	1.0	2,514.75	2,514.75
4	0.8	2,524.33	2,524.33
4	0.6	2,543.49	2,543.49
4*	0.4	2,572.10	2,572.10
4	0.2	2,603.90	2,605.41
4	0.0	2,635.30	2,639.81

Inter-gate distance is the distance between in and outbound gates. These variants are designed to represent the different levels of intra-facility costs. Table entries with asterisks are shown in Fig. 3

improved solution is found in many cases (i.e. a better solution is found between the BKS and the LB). These comparative bounds are given below in Table 3.

7.2 Case 2

When the ten and eight nodes are clustered the resulting aggregate inter-gate flows are obtained from Eq. (9). The model is solved for the aggregated flows data (see Table 1) and produces new gate assignments. Considering these inter-gate flows the

Table 3 Case 1 model runs for 10×8 system, with lower bound for comparison

Number of spokes	Intergate distance	A	B	C	D	E	F
		Lower Bound	LB as % of final optimal value	Gap (%)	Initial upper bound (a feasible solution from LB)	CPLEX OPT MODEL	Percent greater than the true optimum using heuristic (%)
10×8	1.0	2,569.49	96.1	3.9	2,581.02	2,673.51	0.28
10×8	0.8	2,577.09	96.4	3.6	2,687.22	2,674.13	0.49
10×8	0.6	2,586.57	96.4	3.6	2,700.57	2,683.69	0.63
10×8	0.4	2,597.82	96.2	3.8	2,721.69	2,699.85	0.81
10×8	0.2	2,510.77	95.8	4.2	2,751.27	2,725.35	0.95
10×8	0.0	2,624.74	95.1	4.9	2,768.40	2,759.86	0.31

Number of spokes ten inbound, eight outbound. Intergate distance = Infra facility distance from inbound to outbound sides

A Lower Bound From Eqs. (22) and (23); *B* LB as % of final optimal value from column E; $C = 1 - A/E$; *D* Initial upper bound (a feasible solution from LB) uses *Z* variables; *E* CPLEX optimal solution; *F* Percent greater than the true optimum using heuristic = $D/E - 100$

optimal allocation of the improved linkages are shown in the colored boxes corresponding to the colored lines on the diagram (Fig. 4). The 2×2 case has a very high level of flow aggregation basically taking place on a pair of through lines, and a criss-cross pattern for the red to green line interchanges, as shown to the right of the diagram. below: For example, in the 3×3 case, the interactions between nodes 4,5,6 and a,b,c account for a total of 35 units of flow, and these are connected in the diagram on the "green" line. Clearly 7,8,9,10 to f,g,h is a very large flow (blue line) and this might require some capacity balance in a more elaborate case. When the system is organized into four incoming and outgoing feeder lines, the optimal organization of red, green, blue, and grey lines shown to the lower right of the figure. This chart shows which pairs of gates might be connected or given improved service capability. To be clear, the remaining inter-gate interactions still take place, but need to connect by movement from one gate area to another, in a manner akin to "walking from one platform to another" in the transit example.

8 Conclusions and future work

The paper has suggested a class of problems that may have been overlooked in the literature: i.e. the handling of intra-facility flows at a hub or switching point.
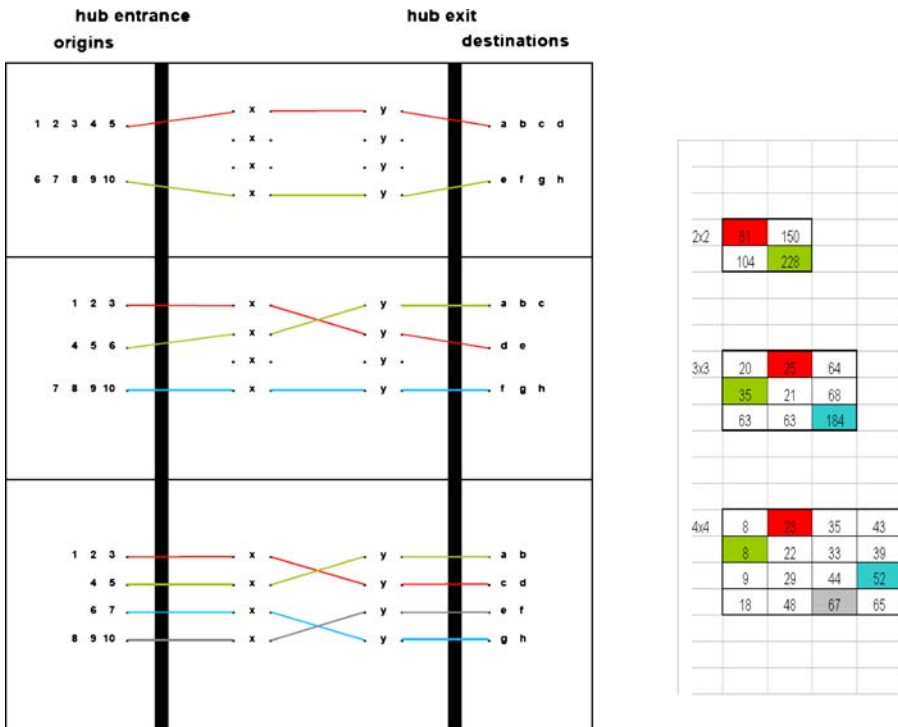

Fig. 4 Case 2 results for 2, 3, and 4 feeder lines. Notes: the connections between the *colored lines* are not shown for clarity, but it should be clear that the x and y gates are fully interconnected

Handling such flows has been shown to give rise to a number of interesting matching and optimization problems. Among these, the idea of adapting the dock-door assignment problem to the hub flow issue, and the idea of selectively joining pairs of spokes to form through connections are especially ripe for further realistic modeling. Some simplifications adopted here would need to be expanded in an application: one would be a much larger scale of the number of nodes. As suggested in the review, subway and cross-dock problems pose considerable complex challenges in the real world. Secondly, the results and conclusions in this paper are dependent on the very simple numerical set up of the sample problem, but even within that framework a reasonable range of results is found and these are expected to be easy to generalize: depending on the spatial arrangement of the flows and the location of the gates relative to the inbound and outbound spokes, the actual usage and levels of flow between gates reflects intuitive levels of spatial organization.

The two special cases that are modeled here also seem to encompass the major representative problems. Variants on these problems are potential topics for further work, and would depend on the particular protocols that govern the network: a lot might depend for example on whether all the incoming flows on a feeder are permitted to disperse to the most suitable gates for the completion of their journeys, or if some kind of single allocation to a transfer node is a more fitting protocol.

For future expansion of these ideas, more realistic situations will need to be examined. Synchronizing and organizing the timing of arrivals and departures at hubs in a type of pulsed arrival and departure pattern is quite difficult (Bobzin 2005). Airlines manage this interdependence by organizing a set of inbound and outbound flights (Trietsch 1993). The difficulty of making those connections work in a very small time window (with the consequence of possible missed connections and the delay of outbound flights to accommodate the late arriving inbound passengers) has led to more relaxed timing. Recently, the idea has been modified to provide more slack between waves of arrivals to make sure that the time is sufficient to make the connections even if there are inbound delays. Also it is clear that a choreographed arrival and departure rate might be very sensitive to random disruption, and in a sense impossible to "get it right" all the time.

Future work such as the question of the optimal number of lines to switch at the same hub, the sensitivity of these issues to the basic number and interaction patterns of the underlying network nodes and the potential for the system to create a second or third hub will be dealt with in follow-up papers on particular extensions and applications.

A two hub version of this problem is similar to a single hub network except that the nodes can have interactions through a high capacity connector.

A security version of this problem would be to model the inter-terminal flows and to maximize the time spent in scrutiny—in other words to delay the traversal of the flows for the longest possible time. This is likely to be an inconvenient level of intervention however, and so the best practice might be to delay the highest risk material and permit the fastest passage for the least threat. As a contemporary and relevant security extension for the paper, note that the flows at a hub intermingle a variety of components. The inbound flows to a particular gate (or freight dock in logistics) comprise many elements that need to be dispersed to other nodes, and so traverse the hub to reach the appropriate outgoing gates. Suppose the components

have varying densities of potential threat levels. The inter-gate flows are then a *blend* of a variety of rates and intensities: scrutiny might be targeted most intensively where the flows have the highest probability of intercepting a threat. Among the previously understudied aspects of this problem are the pairs of interactions to be provided with especially favored flow. This aspect of the problem has potential applications to security issues where certain through flows receive heightened scrutiny while others are given an automated clear path (EZ Pass, Clear Path, Electronic Data Exchange; these are various methods to assure a quicker passage through a check point). Because terrorist attacks often target busy transport facilities, the ideas here have relevance for the detection of vulnerable links; i.e. ones that involve a high amount of transfer from one line to another.

Future applications are also likely for planning bus routes. For example a realistic application was studied for the LakeTran bus system in Lake Co. Ohio where the towns spread out along a linear network served as the hub collection points for the local communities. In planning transfers for this Dial-a-Ride bus system the goal was to assess the probable impact on the expected number of stops and waits for a flexible on demand system feeding into a main line or backbone route; the convenience vs efficiency trade off was of great concern to the system operators.

References

Bartholdi JJ III, Gue KR (2000) Reducing labor costs in an LTL crossdocking terminal. Oper Res 48 (6):823–832

Bartholdi JJ III, Gue KR (2004) The best shape for a crossdock. Transp Sci 38(2):235–244

Bobzin H (2005) Principles of network economics. Springer, Berlin

Campbell JF (1994) Integer programming formulations of discrete hub location problems. Eur J Oper Res 72:387–405

Chang SKJ, Hsu CL (2001) Modeling passenger waiting time for intermodal transit stations. Transp Res Rec 1753:69–75

Chowdhury SM, Chien SIJ (2002) Intermodal transit system coordination. Transp Plan Technol 25 (4):257–287

Crainic TG, Kim KH (2007) Intermodal transportation. In: Barnhart C, Laporte G (eds) Transportation, handbooks in operations research and management science. Elsevier Science, Amsterdam

Fotheringham AS, O'Kelly ME (1989) Spatial interaction models: formulations and applications. Kluwer, Amsterdam

Hall RW (2001) Truck scheduling for ground to air connectivity. J Air Transp Manag 7(6):331–338

Kara BY, Tansel BC (2001) The latest arrival hub location problem. Manage Sci 47(10):1408–1420

Kleinrock L (1964) Communication nets. McGraw Hill, New York

Koopmans TC, Beckmann M (1957) Assignment problems and the location of economic activities. Econometrica 25:53–76

Kuby MJ, Gray RG (1993) The hub network design problem with stopovers and feeders: the case of Federal Express. Transp Res A 27:1–12

Lee K, Lee H-Y (1998) A new algorithm for graph-theoretic nodal accessibility measurement. Geogr Anal 30(1):1–14

Lim A, Rodrigues B, Zhu Y (2005) Airport gate scheduling with time windows. Artif Intell Rev 24(1):5–31

Martello S, Toth P (1990) Knapsack problems: algorithms and computer implementations. Wiley, New York

Nemhauser GL, Wolsey LA (1988) Integer and combinatorial optimization. Wiley, New York

Oh Y, Hwang H, Cha CN (2006) A dock-door assignment problem for the Korean mail distribution center. Comput Ind Eng 51(2):288–296

O'Kelly ME (1986) Activity levels at hub facilities in interacting networks. Geogr Anal 18(4):343–356

O'Kelly ME (1987) A quadratic integer program for the location of interacting hub facilities. Eur J Oper Res 32:393–404

O'Kelly ME, Skorin-Kapov D, Skorin-Kapov S (1995) Lower bounds for the hub location problem. Manage Sci 41:713–721

Ovenden M (2007) Transit maps of the world. Penguin, Baltimore, MD

Pentico DW (2007) Assignment problems: a golden anniversary survey. Eur J Oper Res 176(2):774–793

Reiter S, Sherman GR (1962) Allocating indivisible resources affording external economies or diseconomies. Int Econ Rev 3:108–135

Skorin-Kapov D, Skorin-Kapov J, O'Kelly ME (1996) Tight linear programming relaxations of uncapacitated p-hub median problems. Eur J Oper Res 94(3):582–593

Trietsch D (1993) Scheduling flights at hub airports. Transp Res B Methodol 27(2):133–150

Tsui LY, Chang C-H (1992) An optimal solution to a dock door assignment problem. Comput Ind Eng 23:283–286

Yaman H, Kara BY, Tansel BC (2007) The latest arrival hub location problem for cargo delivery systems with stopovers. Transp Res B 41(8):906–919

Further Reading

Gu JX, Goetschalckx M, McGinnis LF (2007) Research on warehouse operation: a comprehensive review. Eur J Oper Res 177(1):1–21

Lee YH, Jung JW, Lee KM (2006) Vehicle routing scheduling for cross-docking in the supply chain. Comput Ind Eng 51(2):247–256

Lim A, Ma H, Miao ZW (2006a) Truck dock assignment problem with operational time constraint within crossdocks. Lect Notes Comput Sci 4031:262–271

Lim A, Ma H, Miao ZW (2006b) Truck dock assignment problem with time windows and capacity constraint in transshipment network through crossdocks. Lect Notes Comput Sci 3982:688–697

Matisziw T (2003) Spatial aspects of interregional truck-based commodity flow modeling. Unpublished PhD Dissertation Department of Geography, The Ohio State University

McWilliams DL, Stanfield PM et al (2005) The parcel hub scheduling problem: a simulation-based solution approach. Comput Ind Eng 49(3):393–412