Reliable Hub Network Design: Formulation and Solution Techniques

Trung Hieu Tran[a] Jesse R. O'Hanley[b*] M. Paola Scaparra[b]

[a] *Department of Statistics, University of Warwick, Coventry CV4 7AL, U.K.*

[b] *Kent Business School, University of Kent, Canterbury, Kent CT2 7PE, U.K.*

Abstract

In this paper, we investigate the issue of unreliability in hub location planning. A mixed integer nonlinear programming model is formulated for optimally locating p uncapacitated hubs, each of which can fail with a site-specific probability. The objective is to determine the location of hubs and the assignment of demand nodes to hubs in order to minimize expected demand weighted travel cost plus a penalty if all hubs fail. A linear version of the model is developed employing a specialized flow network called a probability lattice to evaluate compound probability terms. A Tabu search algorithm is proposed to find optimal to near optimal solutions for large problem instances. A parallel computing strategy is integrated into the Tabu search process to improve performance. Experimental results carried out on several benchmark instances show the efficiency of our linearized model and heuristic algorithm. Compared to a standard hub median model that disregards the potential for hub failures, our model produces solutions that serve larger numbers of customers and at lower cost per customer.

Keywords: hub location; reliability; linearization; probability lattice; Tabu search; parallel computing

1 Introduction

Hub network design involves the location of hub facilities through which flows from different origins to destinations must be routed. Many practical applications of hub location exist for transportation, telecommunications, and other logistics systems. The classic uncapacitated, single allocation p-hub median problem was first introduced by O'Kelly (1987). Alternative formulations have been proposed by Campbell (1994), Skorin-Kapov et al. (1996), Ernst and Krishnamoorthy (1996), and Ebery (2001), among others. Given the difficulty of solving p-hub median problems of even moderate size, a significant amount of research has been aimed at the development of efficient heuristic methods. This includes work by Klincewicz (1992), Skorin-Kapov and Skorin-Kapov (1994), Ernst and Krishnamoorthy (1996), Abdinnour-Helm (1998), Chen (2007), Silva and Cunha (2009), and Ilić et al. (2010). Recent review papers highlight the wide variety of models and solution methods that have been examined for this important problem (Alumur and Kara, 2008; Campbell and O'Kelly, 2012; Farahani et al., 2013).

*Correspondence email: j.ohanley@kent.ac.uk

In this paper, we focus on a critical issue in hub location planning that has, thus far, received little attention: hub reliability. Specifically, we propose an uncapacitated, single allocation hub location model that accounts for the random failure of hubs. A total of p hubs can be located. Hubs are assumed to fail independently with site-specific probabilities. Assignment of customer nodes to hubs follows a sequential order with a customer going to its level-s hub only if the first $s-1$ assigned hubs have all failed. At any given assignment level, a customer can allocate to only one hub. In the event that all hubs fail, a system-wide penalty is incurred. The goal is to minimize, over all hub failure scenarios, the expected demand weighted cost of routing customer-to-customer flows plus the penalty. We refer to the problem formally as the uncapacitated, single allocation, unreliable p-hub median problem (UHMP).

Within the location science literature, the topic of reliability has received considerable attention, particularly in the past decade. Reliable versions exist for the fixed-charge location problem (Snyder and Daskin, 2005; Cui et al., 2010), the p-median problem (Drezner, 1987; Berman et al., 2007), the maximum covering problem (Daskin, 1983; Camm et al., 2002), and various types of supply chain systems (Qi et al., 2010; Peng et al., 2011). Simpler reliability models assume there is an equal facility failure probability (Drezner, 1987; Snyder and Daskin, 2005), while more sophisticated ones allow for unequal failure probabilities (Berman et al., 2007; Cui et al., 2010; O'Hanley et al., 2013a). A detailed review of facility location problems under uncertainty can be found in Snyder (2006).

In the context of hub design and operations, examples of work dealing with reliability are limited. Exceptions include Janić (2005), Ball et al. (2007), and O'Kelly et al. (2006), all which propose response strategies (e.g., delaying, canceling, rerouting, and network peering) if and when disruptions occur. Such measures are important for coping with disruption, but are reactive in nature (as opposed to strategic) and can be expensive to implement, especially in the event of multiple disruptions.

A generally more robust approach for dealing with hub disruption is to consider it when deciding where to locate hubs in the first place. Kim and O'Kelly (2009), for example, consider the reliability of routes in the design of telecommunications networks. Failure is assumed to occur along arcs and at hubs. Critically, no rerouting is assumed to occur. Consequently, the success of sending messages between an origin and destination requires that all intermediary arcs (i.e., origin to hub, hub to hub, and hub to destination) be operational. Models with and without dispersion requirements are proposed for locating hubs in order to maximize expected flows. Following the same arc disruption approach, Kim (2012) proposes a series of hub location models to mitigate against hub failures, including two variants in which disrupted flows can be rerouted through a single intermediate backup hub. The backup hubs (q in total), unlike primary hubs (p in total), are assumed not to fail and are only used in the event of primary hub failure. Non-hubs are further allowed to have multiple primary and backup hub allocations.

The use of backup hubs has also been employed by An et al. (2011) and Azizi et al. (2014). In both studies, nonlinear and linear integer models are devised to minimize transportation costs under pre- and post-disruption conditions plus a penalty for lost demand (when the source or destination is a hub). Unlike Kim (2012), non-hubs must be assigned to a single hub and disrupted flow can pass through up to two hubs. Critically, as opposed to having a set of dedicated backup hubs, backup hubs are chosen among the existing primary hubs. It is assumed that at most one hub can fail at any one time so that the backup is guaranteed to be operational. The models of An et al. (2011) and Azizi et al. (2014) differ in how disrupted flow is rerouted. In Azizi et al. (2014), all flow passing through the disrupted hub is rerouted through the backup (i.e., each non-hub assigned to a particular hub must assign to the same backup hub), whereas in An et al.

(2011), individual flows between different origin-destination pairs can be allocated to different backups (i.e., non-hubs can assign to different backup hubs). In terms of solution methodology, An et al. (2011) propose Lagrangian relaxation combined with branch and bound to find optimal solutions to their problem. In Azizi et al. (2014) a genetic algorithm is used to produce heuristic solutions.

Our current work, as with the papers cited above, takes an expected-value approach for dealing with hub failures. An alternative would be to plan specifically for the worst-case scenario. Using an interdiction type framework, hubs/arcs are not assumed to fail randomly but in a systematic way which causes maximum disruption to the system. Parvaresh et al. (2012), for example, propose a bi-level mixed integer formulation for optimally locating p hubs in order to minimize the maximum increase in transportation costs as a result of losing any subset of r hubs. Similar location-interdiction and protection-interdiction models have been proposed for the well-known maximum covering (O'Hanley and Church, 2013), p-median (Scaparra and Church, 2008; Losada et al., 2012a; Liberatore et al., 2012) and fixed-charge facility location (Aksen and Aras, 2012) problems.

We offer a few observations about previous work on reliable hub location. One obvious critique of using dedicated backups is the significant cost involved. Such a strategy only makes sense when there is an appreciably high chance of disruption. If not, backups would only rarely be relied upon. The assumptions that dedicated backups are fail-proof and that all disrupted flow needs to be rerouted exclusively through backups are two other somewhat dubious assumptions. Rerouting flows through currently operational hubs seems a far more sensible and efficient option. However, even when this has been attempted, the rather unrealistic assumption has been made that only one hub can fail at a time, meaning that a hub is allowed to fail when considered as the primary hub but not when it is a backup. Logically, multiple hub failures are a distinct possibility unless the probability of any single hub failing is extremely small, in which case using a reliable hub location model probably serves little purpose. On the other hand, if failure probabilities are sufficiently large and multiple hubs can fail, it makes sense to recognize this fact and incorporate it into the decision making framework.

Our present work is aimed directly at addressing some of the shortcomings of existing hub location planning models. To this end, we begin by presenting a nonlinear integer model for locating unreliable hubs that not only considers the possibility of multiple hub failures but also allows disrupted flows to be rerouted through remaining operational hubs. We succeed in linearizing this model through the use of a novel network flow structure referred to as a *probability lattice*. A probability lattice extends the concept of probability chains recently introduced by O'Hanley et al. (2013a) for evaluating compound probability terms by interlinking multiple probability chains together. Related work on linearizing compound probability terms using alternative network flow structures includes that of Morton et al. (2007), Losada et al. (2012b) and O'Hanley et al. (2013b). The probability lattice model allows us to produce optimal solutions, but only for relatively small problem instances. To efficiently solve medium to large instances, we propose a Tabu search heuristic coupled with a parallel computing strategy.

The remainder of this paper is organized as follows. In Section 2, we present a nonlinear formulation of our unreliable hub location model and then show how to linearize the it through the introduction of a probability lattice. A description of the Tabu search algorithm, a node-to-hub assignment heuristic, and parallel computing strategy are provided in Section 3. In Section 4, we report computational results of our linearized model and heuristic algorithm on two test datasets. We further discuss some insights about reliable hub network design. Finally, we give some concluding remarks and suggestions for future work in Section 5.

2 The Unreliable p-Hub Median Problem

2.1 Nonlinear Formulation

We assume that p hubs need to be located and that non-hub nodes (aka spokes) need to be singly allocated to a hub with all inbound as well as outbound flows for a spoke being routed through the assigned hub. The hubs fail independently with site-specific probabilities. In the event of hub failure, the spokes assigned to a failed hub must be reassigned to another operational hub. For simplicity, spokes assign to hubs in a sequential fashion based on a level-set approach. Specifically, a spoke will assign to its level-s hub only when the hubs at levels $1, \ldots, s-1$ have all failed. In the event that a spoke cannot be allocated to a hub (i.e., all hubs fail), a penalty cost ϕ is incurred per unit demand. Some other basic assumptions of our model are that: (i) the cost of locating hubs is equal; (ii) hubs have unlimited capacities; (iii) all hubs in the network are fully connected; (iv) planners have complete information about the operational status of hubs; (v) a spoke can be allocated to only one hub at each assignment level; (vi) direct transportation between spoke nodes is not allowed; and (vii) transportation costs satisfy the triangle inequality (i.e., flow between two different spokes can pass through one or at most two hubs). The goal of our problem is to determine both the location of hubs and level assignments for spokes to hubs.

To formulate the problem mathematically, let N be a set of demand nodes, indexed by i, j, k and m. Hubs, of which there are p, can be located at any of the $n = |N|$ demand sites. For notational purposes, indices i and j will be used to index non-hub sites; indices k and m will be used to index hub sites. Indexes $s = 1, \ldots, p$ and $t = 1, \ldots, p$, meanwhile, will be used to denote the level-s and level-t hub assignments for origin node i and destination node j, respectively. The demand flow from node i to node j is denoted by h_{ij}. The unit transportation cost along the link connecting nodes i and j is given by c_{ij}. The overall unit cost of transporting demand from node i to node j via hubs at sites k and m is given by $c_{ijkm} = \chi c_{ik} + \gamma c_{km} + \delta c_{mj}$, where γ is the unit cost for inter-hub transportation and χ and δ are unit collection and distribution costs, respectively. Typically, γ is assumed to be less than χ and δ to account for economies of scale associated with consolidating flows through hubs. A hub located at node k is assumed to fail (not fail) with probability q_k ($\bar{q}_k = 1 - q_k$).

The set of decision variables is given by:

$$X_k = \begin{cases} 1 & \text{if a hub is located at site } k \\ 0 & \text{otherwise} \end{cases}$$

$$Y_{ik}^s = \begin{cases} 1 & \text{if a hub located at site } k \text{ is the level-}s \text{ hub of node } i \\ 0 & \text{otherwise} \end{cases}$$

$$V_{ijm}^{st} = \begin{cases} 1 & \text{if site } m \text{ is the level-}t \text{ hub for node } j \text{ but not the level-}s \text{ hub for node } i \\ 0 & \text{otherwise} \end{cases}$$

$$W_{ijm}^{st} = \begin{cases} 1 & \text{if site } m \text{ is the level-}t \text{ hub for node } j \text{ and the level-}s \text{ hub for node } i \\ 0 & \text{otherwise} \end{cases}$$

λ^{st}_{ijkm} = the joint probability that node i can allocate to site k at its level-s hub and node j can allocate to site m as its level-t hub

Given the assumption of single allocation, it necessarily holds that $\lambda^{st}_{ijkm} = \lambda^{ts}_{jimk}$. Accordingly, let $h'_{ij} = h_{ij} + h_{ji}$ and $c'_{ijkm} = h_{ij}c_{ijkm} + h_{ji}c_{jimk}$. A nonlinear formulation of the uncapacitated, single allocation, unreliable p-hub median problem (UHMP) is then given as follows.

$$\text{[UHMP1]} \ \min \sum_i \sum_{j>i} \sum_k \sum_m \sum_s \sum_t c'_{ijkm}\lambda^{st}_{ijkm} + \sum_i \sum_{j>i} \phi h'_{ij} \prod_k (1 - \bar{q}_k X_k) \tag{1}$$

s.t.

$$\sum_k X_k = p \tag{2}$$

$$\sum_k Y^s_{ik} = 1 \qquad \forall i, s \tag{3}$$

$$\sum_s Y^s_{ik} \le X_k \qquad \forall i, k \tag{4}$$

$$\lambda^{st}_{ijkm} = \begin{cases} \prod_{r<s} \prod_\ell (1 - \bar{q}_\ell Y^r_{i\ell})\, \bar{q}_k Y^s_{ik} \times \prod_{r<t} \prod_\ell \left(1 - \bar{q}_\ell V^{sr}_{ij\ell}\right) \left(1 - W^{sr}_{ij\ell}\right) \bar{q}_m V^{st}_{ijm} & m \ne k \\ \prod_{r<s} \prod_\ell (1 - \bar{q}_\ell Y^r_{i\ell})\, \bar{q}_k Y^s_{ik} \times \prod_{r<t} \prod_\ell \left(1 - \bar{q}_\ell V^{sr}_{ij\ell}\right) \left(1 - W^{sr}_{ij\ell}\right) W^{st}_{ijm} & m = k \end{cases}$$
$$\forall i, j > i, k, m, s, t \tag{5}$$

$$1 + V^{st}_{ijm} \ge Y^t_{jm} + \sum_{r>s} Y^r_{im} \qquad \forall i, j > i, m, s, t \tag{6}$$

$$V^{st}_{ijm} \le Y^t_{jm} \qquad \forall i, j > i, m, s, t \tag{7}$$

$$V^{st}_{ijm} \le \sum_{r>s} Y^r_{im} \qquad \forall i, j > i, m, s, t \tag{8}$$

$$1 + W^{st}_{ijm} \ge Y^t_{jm} + Y^s_{im} \qquad \forall i, j > i, m, s, t \tag{9}$$

$$W^{st}_{ijm} \le Y^t_{jm} \qquad \forall i, j > i, m, s, t \tag{10}$$

$$W^{st}_{ijm} \le Y^s_{im} \qquad \forall i, j > i, m, s, t \tag{11}$$

$$X_k \in \{0, 1\} \qquad \forall k \tag{12}$$

$$Y^s_{ik} \in \{0, 1\} \qquad \forall i, k, s \tag{13}$$

$$V^{st}_{ijm}, W^{st}_{ijm} \ge 0 \qquad \forall i, j > i, m, s, t \tag{14}$$

The objective function (1) minimizes total expected transportation cost over all hub disruption scenarios (the series of summations involving $c'_{ijkm}\lambda^{st}_{ijkm}$) plus a penalty $\sum_i \sum_{j>i} \phi h'_{ij}$ in the event all hubs fail, which occurs with probability $\prod_k (1 - \bar{q}_k X_k)$. Note that intrahub flows h_{ii} are not taken into consideration as these are assumed to occur regardless of any possible hub disruptions. Constraint (2) requires exactly p hubs be located. Constraints (3) ensure for any given assignment level s that each node is allocated to exactly one hub. Constraints (4) allow node i to assign to site k at any level s only if site k has been selected as a hub ($X_k = 1$).

Equations (5) determine for any origin-destination pair the joint probability that origin node i will allocate to site k as its level-s hub and destination node j will allocate to site m as its level-t hub. Specifically, for each node i, the product $\prod_{r<s} \prod_\ell (1 - \bar{q}_\ell Y^r_{i\ell})$ in (5) determines the probability that node i cannot allocate to any of its first $s-1$ assigned hubs (i.e., hubs at levels $r = 1, \ldots, s-1$ all fail). Note that this product is taken over all sites ℓ since, according to (3), only one $Y^r_{i\ell}$ will be equal to 1 at each level $r < t$. The product $\bar{q}_k Y^s_{ik}$ gives the probability that level-s hub for node i is located at site k and does not fail. Multiplying the two terms together gives the marginal probability that node i will allocate to site k as its level-s hub.

The product to the right of $\prod_{r<s} \prod_\ell (1 - \bar{q}_\ell Y^r_{i\ell})$ in (5), either $\prod_{r<t} \prod_\ell \left(1 - \bar{q}_\ell V^{sr}_{ij\ell}\right) \left(1 - W^{sr}_{ij\ell}\right) \bar{q}_m V^{st}_{ijm}$ if $m \neq k$ or $\prod_{r<t} \prod_\ell \left(1 - \bar{q}_\ell V^{sr}_{ij\ell}\right) \left(1 - W^{sr}_{ij\ell}\right) W^{st}_{ijm}$ if $m = k$, determines the conditional probability that destination node j will allocate to site m as its level-t hub given that origin node i is allocated to site k as its level-s hub. The derivation of this expression is explained as follows. In order for node j to allocate to its level-t hub, all hubs at levels $r < t$ have to fail, which occurs with probability $\prod_{r<t} \prod_\ell \left(1 - \bar{q}_\ell V^{sr}_{ij\ell}\right)$ assuming node j and node i do not share any hub assignments in common at levels $r < t$ ($W^{sr}_{ij\ell} = 0, \forall \ell, r < t$). If, however, node j has the same hub assignment ℓ at level $r < t$ as node i does at level s ($W^{sr}_{ij\ell} = 1$), then the probability of node j actually allocating to its level-t hub is 0, since node j would allocate to site ℓ at level r. Consequently, for a given level t, either this term will be equal to 0 (i.e., if $W^{sr}_{ij\ell} = 1$ for any ℓ, $r < t$) or equal to the product of the hub failure probabilities at levels $r < t$ (i.e., the series of q_ℓ values such that $V^{sr}_{ij\ell} = 1$, $r < t$). Accordingly, $\prod_{r<t} \prod_\ell \left(1 - \bar{q}_\ell V^{sr}_{ij\ell}\right) \left(1 - W^{sr}_{ij\ell}\right)$ gives the correct conditional probability that all hubs assigned to node j fail at levels $r < t$.

Continuing on, it is necessary to distinguish whether at level t destination node j would allocate to the same or to a different hub as origin node i. If the level-t assignment for node j is different from the level-s assignment for node i ($V^{st}_{ijm} = 1$), then with conditional probability $\prod_{r<t} \prod_\ell \left(1 - \bar{q}_\ell V^{sr}_{ij\ell}\right) \left(1 - W^{sr}_{ij\ell}\right) \bar{q}_m V^{st}_{ijm}$ node j will allocate to a particular site m. This is simply the probability that all hubs at levels $r < t$ fail $\prod_{r<t} \prod_\ell \left(1 - \bar{q}_\ell V^{sr}_{ij\ell}\right) \left(1 - W^{sr}_{ij\ell}\right)$ times the probability that the level-t hub is located at m and operational $\bar{q}_m V^{st}_{ijm}$. On the other hand, if the level-t and level-s assignments for nodes j and i are the same ($W^{st}_{ijm} = 1$), then the probability of node j allocating to its level-t hub is conditional only on the failure of hubs at levels $r < t$ (i.e., $\prod_{r<t} \prod_\ell \left(1 - \bar{q}_\ell V^{sr}_{ij\ell}\right) \left(1 - W^{sr}_{ij\ell}\right)$). The operational status of the level-t hub for node j is already factored into the probability of node i being able to assign to its level-s hub. Specifically, with probability $\bar{q}_k Y^s_{ik}$ the level-s hub for i and level-t hub for j will be operational, as given in the first part of (5).

To conclude, constraints (6) - (11) help to determine the correct values for the auxiliary variables V^{st}_{ijm} and W^{st}_{ijm} feeding into (5). Inequalities (6) force variable V^{st}_{ijm} to be 1 whenever site m is the level-t hub for node j but node i assigns to m at some level $r > s$ (an equivalent definition for variable V^{st}_{ijm}). Inequalities (7) and (8), on the other hand, force variable V^{st}_{ijm} to be 0 whenever this condition is not met. Similarly, inequalities (9) force W^{st}_{ijm} to be 1 whenever site m is both the level-t hub for node j and the level-s hub for node i, while inequalities (10) and (11) force variable W^{st}_{ijm} to be 0 whenever this does not hold. Lastly, constraints

(12) and (13) place binary restrictions on the X_k and Y_{ik}^s variables, respectively, while constraints (14) place non-negativity restrictions on the V_{ijm}^{st} and W_{ijm}^{st} variables. Given the nature of the objective function and the constraint set, these variables are guaranteed to take on binary values.

2.2 Linear Reformulation

To linearize UHMP, we introduce the following auxiliary variables.

$\alpha_{ik}^s =$ the marginal probability that node i can allocate to site k as its level-s hub

$\beta_{ik}^s =$ the marginal probability that node i cannot allocate to site $\ell \leq k$ as its level-s hub

$\mu_{ijkm}^{st} =$ the joint probability that node i can allocate to site k as its level-s hub but node j cannot allocate to site $\ell \leq m$ as its level-t hub

A linear formulation for UHMP is then given by:

$$\text{[UHMP2]} \ \min \sum_i \sum_{j>i} \sum_k \sum_m \sum_s \sum_t c'_{ijkm} \lambda_{ijkm}^{st} + \phi \sum_i \sum_{j>i} h'_{ij} \beta_{in}^p \tag{15}$$

subject to (2)-(4), (6)-(14) and the following set of constraints:

(*Backbone Chain*)

$$\alpha_{ik}^s + \beta_{ik}^s = \begin{cases} 1 & k=1, s=1 \\ \beta_{in}^{s-1} & k=1, s>1 \\ \beta_{i(k-1)}^s & k>1, s>1 \end{cases} \qquad \forall i,k,s \tag{16}$$

$$\alpha_{ik}^s \leq \bar{q}_k Y_{ik}^s \qquad \forall i,k,s \tag{17}$$

$$\alpha_{ik}^s \leq \begin{cases} \bar{q}_k & k=1, s=1 \\ \bar{q}_k \beta_{in}^{s-1} & k=1, s>1 \\ \bar{q}_k \beta_{i(k-1)}^s & k>1, s>1 \end{cases} \qquad \forall i,k,s \tag{18}$$

$$\beta_{ik}^s \leq \begin{cases} q_k + \bar{q}_k\left(1 - Y_{ik}^s\right) & k=1, s=1 \\ q_k \beta_{in}^{s-1} + \bar{q}_k\left(1 - Y_{ik}^s\right) & k=1, s>1 \\ q_k \beta_{i(k-1)}^s + \bar{q}_k\left(1 - Y_{ik}^s\right) & k>1, s>1 \end{cases} \qquad \forall i,k,s \tag{19}$$

(*Spur Chains*)

$$\lambda_{ijkm}^{st} + \mu_{ijkm}^{st} = \begin{cases} \alpha_{ik}^{s} & m=1, t=1 \\ \mu_{ijkn}^{s(t-1)} & m=1, t>1 \\ \mu_{ijk(m-1)}^{st} & m>1, t>1 \end{cases} \quad \forall i, j>i, k, m, s, t \tag{20}$$

$$\lambda_{ijkm}^{st} \leq \begin{cases} \bar{q}_m V_{ijm}^{st} & m \neq k \\ \bar{q}_m W_{ijm}^{st} & m = k \end{cases} \quad \forall i, j>i, k, m, s, t \tag{21}$$

$$\lambda_{ijkm}^{st} \leq \begin{cases} \bar{q}_m \alpha_{ik}^{s} & m=1, t=1 \\ \bar{q}_m \mu_{ijkn}^{s(t-1)} & m=1, t>1 \\ \bar{q}_m \mu_{ijk(m-1)}^{st} & m>1, t>1 \end{cases} \quad \forall i, j>i, k, m \neq k, s, t \tag{22}$$

$$\mu_{ijkm}^{st} \leq \bar{q}_m \left(1 - W_{ijm}^{st}\right) \qquad \forall i, j>i, k, m=k, s, t \tag{23}$$

$$\mu_{ijkm}^{st} \leq \begin{cases} q_m \alpha_{ik}^{s} + \bar{q}_m \left(1 - V_{ijm}^{st}\right) & m=1, t=1 \\ q_m \mu_{ijkn}^{s(t-1)} + \bar{q}_m \left(1 - V_{ijm}^{st}\right) & m=1, t>1 \\ q_m \mu_{ijk(m-1)}^{st} + \bar{q}_m \left(1 - V_{ijm}^{st}\right) & m>1, t>1 \end{cases} \quad \forall i, j>i, k, m \neq k, s, t \tag{24}$$

Our linear reformulation of UHMP is based on evaluating the nonlinear terms in (1) and (5) using continuous variables α_{ik}^s, β_{ik}^s, λ_{ijkm}^{st} and μ_{ijkm}^{st} together with flow-conservation constraints (16) and (20) and bounding constraints (17)-(19) and (21)-(24). These variables and constraints form a series of specialized flow networks, one for each origin-destination pair (i, j), which we refer to as a *probability lattice*. A probability lattice represents an extension of the probability chain concept proposed by O'Hanley et al. (2013a) in which multiple probability chains are interlinked together. In our particular problem, the lattice for a given (i, j) pair is composed of a *backbone chain* (16)-(19) for origin node i and multiple *spur chains* (20)-(24) emanating from the backbone for destination node j. A graph representation of a probability lattice is shown in Figure 1.

The backbone chain serves to evaluate marginal probabilities α_{ik}^s and β_{ik}^s while the spur chains are used to compute joint probabilities λ_{ijkm}^{st} and μ_{ijkm}^{st}. In Figure 1, the backbone for origin node i is shown as a series of square nodes along the top with a node defined for each site $k = 1, \ldots, n$ and assignment level $s = 1, \ldots, p$. The dashed out-flow arcs emanating from and the solid transverse arcs connecting the backbone nodes correspond to variables α_{ik}^s and β_{ik}^s, respectively. Connected to each backbone out-flow arc α_{ik}^s is a separate spur chain. Every spur chain for destination node j is made up of a series of round nodes for each site $m = 1, \ldots, p$ and assignment level $t = 1, \ldots, p$ along with a matching set of dashed out-flow and solid transverse arcs corresponding to λ_{ijkm}^{st} and μ_{ijkm}^{st}, respectively.

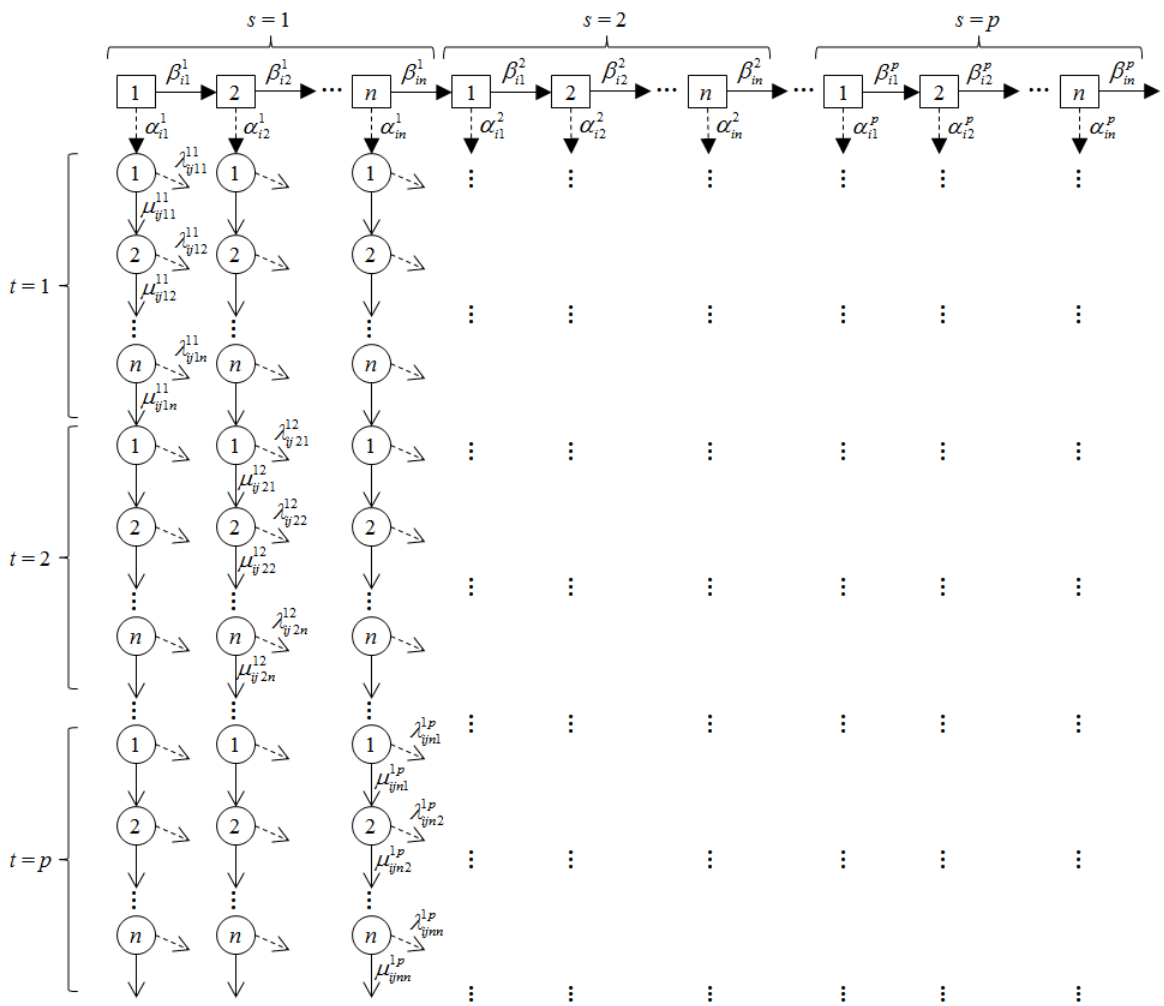


Figure 1: Graph representation of a probability lattice for a hypothetical origin-destination pair (i, j) in a network with n nodes and p hubs.

With respect to the probability lattice's underlying structure and function, the backbone and spur chains are individually composed of p interlinking probability subchains (one for each hub assignment level $s = 1, \ldots, p$ and $t = 1, \ldots, p$, respectively) with each subchain being comprised of n nodes (one for each site $k = 1, \ldots, n$ and $m = 1, \ldots, n$, respectively). Internal flow β^s_{ik} within the backbone chain is implicitly determined by all preceding out-flows $\alpha^r_{i\ell}$, $r \leq s$, $\ell \leq k$. Internal flow μ^{st}_{ijkm} of a spur chain, meanwhile, is determined by both the initial in-flow α^s_{ik} coming from the backbone chain and any intervening out-flows $\lambda^{sr}_{ijk\ell}$, $r \leq t$, $\ell \leq m$.

Looking specifically at the set of constraints (16)-(19) for a given backbone chain, inequalities (17) and (18) place bounds on the probability α^s_{ik} that node i will allocate to site k as its level-s hub. If site k is not selected as the level-s hub ($Y^s_{ik} = 0$), then (17) forces α^s_{ik} to be 0. If site $k = 1$ is selected as the level $s = 1$ hub ($Y^1_{i1} = 1$), (18) restricts α^1_{i1} to be less than or equal to the probability site 1 does not fail or $\bar{q}_1$. If instead site $k = 1$ is selected at a higher level ($Y^s_{i1} = 1$, $s > 1$), the probability of node i assigning to site 1 is bounded above by $\bar{q}_1$ times the probability that node i cannot allocate to any lower level hub β^{s-1}_{in}. For any other site $k \neq 1$ selected as an level-s hub ($Y^s_{ik} = 1$, $k > 1$), the bound on α^s_{ik} is $\bar{q}_k$ times $\beta^s_{i(k-1)}$, the probability that node i cannot allocate to any other site $\ell < k$ at level s. Inequalities (19) take on an analogous role to

(18) by specifying permissible values for the β^s_{ik} variables. They are limiting only in the case when $Y^s_{ik} = 1$. Finally, inequalities (16) force α^s_{ik} and β^s_{ik} to take on binding values as determined by (17)-(18) and (19), respectively.

Spur chain constraints (20)-(24) perform the same role for the λ^{st}_{ijkm} and μ^{st}_{ijkm} flow variables that (16)-(19) do for the α^s_{ik} and β^s_{ik} variables. The primary difference lies in the incorporation of variables V^{st}_{ijm} and W^{st}_{ijm} in place of Y^s_{ik} and the addition of (23), which further restricts μ^{st}_{ijkm} to be 0 whenever $W^{st}_{ijm} = 1$. In particular, if both $V^{st}_{ijm} = 0$ $(m \neq k)$ and $W^{st}_{ijm} = 0$ $(m = k)$, then (21) forces out-flow λ^{st}_{ijkm} to be 0, while (20) requires any non-zero flow to be sent via μ^{st}_{ijkm}. If $V^{st}_{ijm} = 1$ $(m \neq k)$, however, equations (20) force (22) to be binding for λ^{st}_{ijkm} and (24) to be binding for μ^{st}_{ijkm}. Different bounds for the λ^{st}_{ijkm} and μ^{st}_{ijkm} variables are specified in (22) and (24) depending on the site m and assignment level t in question. Lastly, if $W^{st}_{ijm} = 1$, then (23) forces internal flow μ^{st}_{ijkm} to be 0, while (20) requires any non-zero flow to be sent along the out-flow λ^{st}_{ijkm}. It is worth noting that constraints (3) imply for given backbone subchain s, that exactly one out-flow variable α^s_{ik} can be non-zero (i.e., for $Y^s_{ik} = 1$). Consequently, only one spur chain with potentially non-zero flow can be active per backbone subchain for any feasible set of hub locations and hub assignments.

Returning to the linearized objective function (15), expected transportation cost is given by the term $\sum_i \sum_{j>i} \sum_k \sum_m \sum_s \sum_t c'_{ijkm} \lambda^{st}_{ijkm}$, the same as in (1). Variable β^p_{in}, meanwhile, is simply the probability that all hubs fail (specifically, all p hubs assigned to node i fail) and is equivalent to the product $\prod_k (1 - \bar{q}_k X_k)$ in (1).

2.3 Problem Size Reduction

As observed by O'Kelly et al. (1996), the set of feasible inter-node paths can, for any level s and t, be restricted to $A = \{(i, j, k, m) : (i < j) \wedge [(k = i) \vee (k \neq i \wedge k = m = j) \vee (k \neq i \wedge k \neq j \wedge m \neq i)]\}$. Accordingly, it possible to reduce the size of UHMP2 by adding the following constraints, which effectively remove any λ^{st}_{ijkm} variables not in set A from the formulation.

$$\lambda^{st}_{k\ell mk} = \lambda^{st}_{\ell kkm} = 0 \qquad \forall k, \ell > k, m \neq k, s, t \tag{25}$$

This variable fixing scheme was employed by Marín et al. (2006) to a standard multiple allocation hub median problem in the special case where transportation costs are symmetric ($c_{ijkm} = c_{jimk}$). Applying reduction rule (25) produced a marginal decrease in solution time for the datasets we tested. This is not particularly surprising given that we had, prior to applying this rule, already eliminated origin-destination pairs such that $j \leq i$, thus resulting in few additional routes in set A that could subsequently be eliminated.

To further reduce UHMP2, we can also include the following constraints:

$$\mu^{sp}_{ijkn} = 0 \qquad \forall i, j > i, k, s \tag{26}$$

which are based on the observation that if node j cannot allocate to its level-p hub, then the probability of node i allocating to any level-s hub must be 0 (i.e., since all hubs must have failed). Experimental results showed that this simple variable fixing rule was successful in significantly reducing branch and bound solution times by up to 40-fold in some cases.

2.4 Bounding Constraints

To help tighten the linear formulation UHMP2, additional bounding constraints can be imposed on the α_{ik}^s and β_{ik}^s variables that are particularly effective when the hub failure probabilities q_k are all close in value (i.e., $q_k \approx q$, $\forall k$). To begin, note that variable α_{ik}^s can be evaluated directly by the equation $\alpha_{ik}^s = \prod_{r<s} \prod_\ell (1 - \bar{q}_\ell Y_{i\ell}^r)\, \bar{q}_k Y_{ik}^s$. For $Y_{ik}^s = 1$, it is straightforward to show that the product $\prod_{r<s} \prod_\ell (1 - \bar{q}_\ell Y_{i\ell}^r)$ is bounded above (below) by the product of the $s-1$ largest (smallest) hub failure probabilities excluding site k. Letting $Q_{ks}^{max} = \prod_{r \le s} q_{[r]}^{max}(k)$, such that $q_{[r]}^{max}(k)$ is the r^{th} largest hub failure probability in the set $N \backslash \{k\}$, and letting $Q_{ks}^{min} = \prod_{r \le s} q_{[r]}^{min}(k)$, such that $q_{[r]}^{min}(k)$ is the r^{th} smallest hub failure probability in the set $N \backslash \{k\}$, this naturally leads to the following bounds on α_{ik}^s.

$$\alpha_{ik}^s \le Q_{k(s-1)}^{max} \bar{q}_k Y_{ik}^s \qquad \forall i, k, s \tag{27}$$

$$\alpha_{ik}^s \ge Q_{k(s-1)}^{min} \bar{q}_k Y_{ik}^s \qquad \forall i, k, s \tag{28}$$

In a similar way, variable β_{ik}^s can be evaluated using the equation $\beta_{ik}^s = \prod_{r<s} \prod_\ell (1 - \bar{q}_\ell Y_{i\ell}^r)(1 - \bar{q}_k Y_{ik}^s)$. For $Y_{i\ell}^s = 1$, such that $\ell \le k$,β_{ik}^s is bounded above and below by $Q_{\ell(s-1)}^{max} q_\ell$ and $Q_{\ell(s-1)}^{min} q_\ell$, respectively. On the other hand, if $Y_{i\ell}^s = 0$, then β_{ik}^s is bounded above by product of the $s-1$ largest hub failure probabilities (denoted $Q_{(s-1)}^{max}$) and bounded below by the $s-1$ smallest hub failure probabilities (denoted $Q_{(s-1)}^{min}$). Accordingly, we have the following bounds on theβ_{ik}^s variables.

$$\beta_{ik}^s \le Q_{(s-1)}^{max} + \sum_{\ell \le k} \left(Q_{\ell(s-1)}^{max} q_\ell - Q_{(s-1)}^{max} \right) Y_{i\ell}^s \qquad \forall i, k, s \tag{29}$$

$$\beta_{ik}^s \ge Q_{(s-1)}^{min} + \sum_{\ell \le k} \left(Q_{\ell(s-1)}^{min} q_\ell - Q_{(s-1)}^{min} \right) Y_{i\ell}^s \qquad \forall i, k, s \tag{30}$$

Initial testing showed that adding constraints (27)-(30) to UHMP2 usually resulted in a 20% or more reduction in solution time. Moreover, better lower bounds and better feasible solutions (in cases where the model could not be solved to optimality) could be obtained. We note that bounds similar to (27)-(30) can also be derived for variables λ_{ijkm}^{st} and μ_{ijkm}^{st}. Unfortunately, preliminary experiments showed that placing individual bounds on λ_{ijkm}^{st} and or μ_{ijkm}^{st} led to a marked increase in solution times for larger problem instances due to the considerable number of additional constraints.

In spite of improvements in solution time due to the addition of the bounding constraints and variable fixing rules, it was found that even fairly small problems (≤20 nodes) were difficult to solve to optimality. This motivated us to explore the use of a heuristic solution method, as discussed in the following section.

3 Tabu Search for UHMP

Tabu search (Glover, 1989) is a local search method that has been employed for solving a variety of optimization problems, including hub location problems (HLPs). For the standard uncapacitated, single allocation

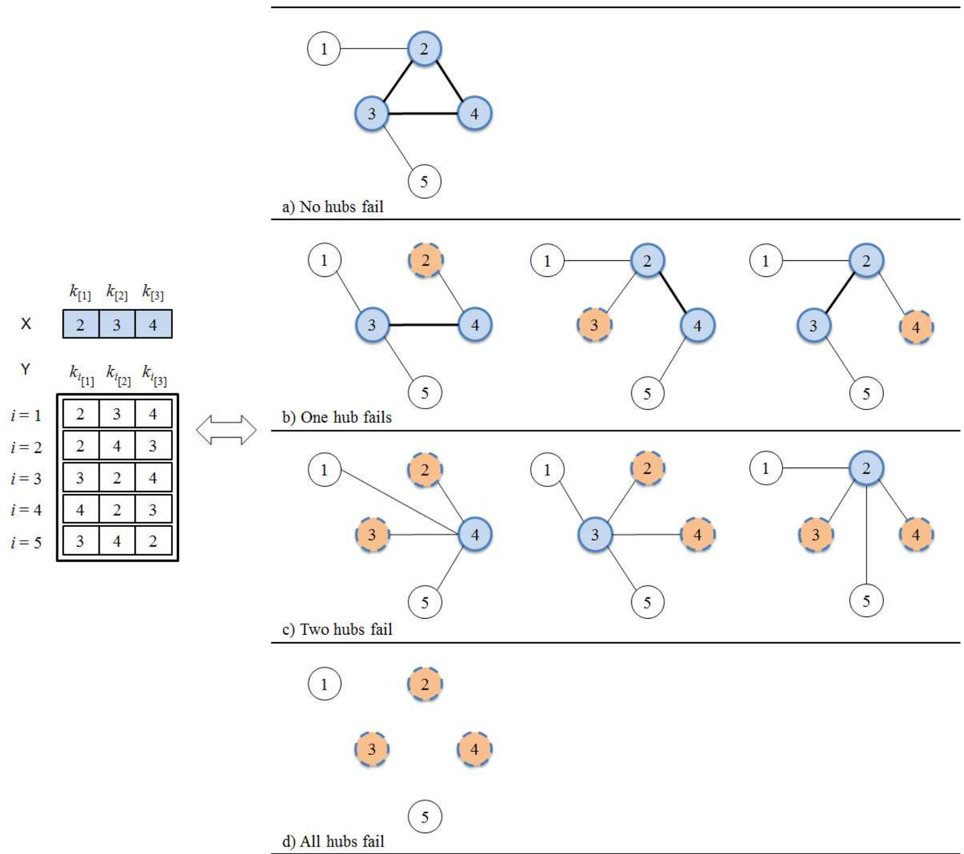


Figure 2: An illustrative example of a Tabu search solution representation.

p-hub median problem (USApHMP), a simple but efficient Tabu search was developed by Skorin-Kapov and Skorin-Kapov (1994). Numerous hybrid approaches that combine Tabu search with some other metaheuristic have also been proposed in attempt to improve its efficiency. This includes Tabu search with genetic algorithms (Abdinnour-Helm, 1998); Tabu search with simulated annealing (Chen, 2007); and multiple start Tabu search (Silva and Cunha, 2009). Tabu search algorithms which have been devised to solve various types of HLPs can be found in Farahani et al. (2013). In what follows, we introduce a Tabu search integrated into a parallel computing strategy to efficiently solve our unreliable hub location model UHMP.

3.1 Solution Representation, Neighborhood Structure and Tabu List

A feasible solution to UHMP is determined entirely by its hub locations $\hat{\boldsymbol{X}}$ and hub assignments $\hat{\boldsymbol{Y}}$; the corresponding $\hat{\boldsymbol{V}}$, $\hat{\boldsymbol{W}}$ and $\hat{\boldsymbol{\lambda}}$ variables can be computed directly from $\hat{\boldsymbol{X}}$ and $\hat{\boldsymbol{Y}}$ via constraints (5)-(11) and (14). To represent a feasible solution in our Tabu search heuristic, therefore, we only need to consider the indices k in $\hat{\boldsymbol{X}}$ and $\hat{\boldsymbol{Y}}$ with value equal to 1. These are stored, respectively, in a $1 \times p$ row vector $\mathsf{X} = [k_{[1]}, \ldots, k_{[z]}, \ldots, k_{[p]}]$ of hub locations and a $n \times p$ matrix $\mathsf{Y} = [[k_{1_{[1]}}, \ldots, k_{1_{[s]}}, \ldots, k_{1[p]}], \ldots, [k_{i_{[1]}}, \ldots, k_{i_{[s]}}, \ldots, k_{i_{[p]}}], \ldots, [k_{n_{[1]}}, \ldots, k_{n_{[s]}}, \ldots, k_{n_{[p]}}]]$ of hub assignments, where $k_{[z]}$ is the index of the z^{th} hub site in the set $\{k : \hat{X}_k = 1\}$ and $k_{i_{[s]}}$ is the index of the level-s hub assignment for node i in the set $\mathsf{Y}_i = \{k : \hat{Y}^s_{ik} = 1\}$, respectively.

An example of our Tabu search solution representation for an example hub network with $n = 5$ demand nodes and $p = 3$ hubs under each possible disruption scenario is shown in Figure 2. Hubs are located at nodes 2, 3, and 4; nodes 1 and 5 are non-hubs. Operational hubs are indicated as blue-filled solid nodes, nonoperational hubs as orange-filled dashed nodes, and non-hubs as white-filled solid nodes. Hub assignments are based on the assignment matrix Y. With all hubs operational, we obtain the hub-and-spoke allocation shown in Figure 2. Here, node 1 is assigned to hub site 2 and node 5 is assigned to hub site 3. If hub site 3 were to fail (middle of Figure 2b), then according to the hub assignment matrix Y, hub 3 would allocate to hub 2 ($k_{3_{[2]}} = 2$) and non-hub 5 would allocates to hub 4 ($k_{5_{[2]}} = 4$).

The neighborhood of solutions used in our Tabu search consists of all 1-opt moves from the current solution X (i.e., any exchange which drops a hub location and replaces it with a non-hub). The first improving move that decreases the objective is chosen and the incumbent solution updated. Following a successful move, the new solution is added to a Tabu list for a specified number of iterations (i.e., the full set of hub locations as opposed to the selected 1-opt move). The Tabu list prohibits recent combinations of moves from being chosen in an attempt to prevent the search from cycling back to previously visited solutions. The length of the Tabu list was determined based on preliminary testing.

Critically, when evaluating any potential solution X' within the neighborhood of X, it is necessary to determine a corresponding set of hub assignments Y'. Given that there are $p!$ possible hub assignment orderings alone for each node, this poses an especially challenging problem. To find a set of hub assignments Y' for candidate solution X', we perform a secondary 2-opt search of each node's hub assignments (i.e., exchanges involving two hub assignments for any non-hub). This is explained in detail in Section 3.2.

We also considered employing larger neighborhoods in our Tabu search, specifically 2-opt moves for hub locations (swapping two hubs with two non-hub) and or 3-opt moves for hub assignments (swapping three hub assignments). Larger neighborhoods can sometimes help by allowing a local search to escape from a local optimum. Preliminary testing, however, showed that the use of these larger neighborhoods only increased run times of the Tabu search without any improvement in solution quality.

3.2 Hub Assignment Search Heuristic

Given a set of hub locations X, we would like to determine an optimal or near-optimal set of hub assignments for each demand node. An initial set of hub assignment for a given node i is obtained by sorting hubs in order of unit transportation costs $c_{ik_{[z]}}$ (i.e., the cost of transporting from node i to the z^{th} hub). The hub with the lowest unit cost is assigned as the level-1 hub, the one with next highest unit cost the level-2 hub, and so on. We also considered ordering hubs based on $\sum_j c'_{ijk_{[z]}k_{[z]}}$, a measure of aggregate demand-weighted transportation cost in which it is presumed both nodes i and j assign to the same hub $k_{[z]}$, but this did not result in substantially better solutions. After constructing an initial ordering of hubs, a 2-opt search is applied to improve the initial allocation order. This 2-opt search is performed repeatedly until no improvement of the assignments can be made.

During a 2-opt search, we need to recalculate the objective value associated with swapping two of a node's hub assignments. To save on the computational effort of this operation, we only recalculate the λ^{st}_{ijkm} probabilities affected by this change. Specifically, assume that we swap the level-r and level-r' hub assignments for node ℓ such that $r' > r$. Only the probabilities $\lambda^{st}_{i\ell k_{i_{[s]}} k_{\ell_{[t]}}}$ $(i < \ell, s = 1, \dots, p, r \leq t \leq r')$ and $\lambda^{st}_{\ell j k_{\ell_{[s]}} k_{j_{[t]}}}$ $(j > \ell, r \leq s \leq r', t = 1, \dots, p)$ are affected by the swap (see Figure 3).

Accordingly, let $\tilde{\lambda}^{st}_{ijk_{i_{[s]}}k_{j_{[t]}}}$ denote the probability of node i allocating to its designated level-s hub $k_{i_{[s]}}$ and node j allocating to its designated level-t hub $k_{j_{[t]}}$ *given* that the level-r and level-r' hub assignments for

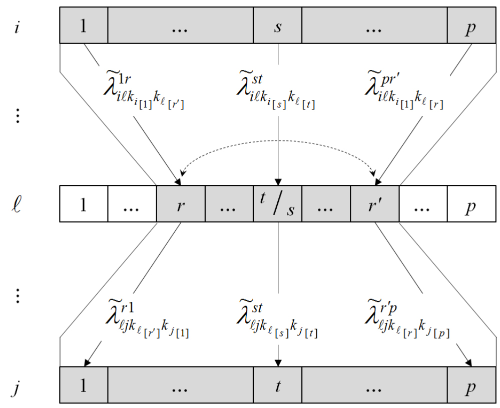


Figure 3: Representation of the required probability updates involved with swapping the level-r and level-r' hub assignment of node ℓ. The indices in each box represent a specified assignment level.

node ℓ (as currently defined in Y) have been swapped. Note that probabilities $\lambda^{st}_{ijk_{i_{[s]}}k_{j_{[t]}}}$ and $\tilde{\lambda}^{st}_{ijk_{i_{[s]}}k_{j_{[t]}}}$ can be calculated via (5) with variables Y^s_{ik}, V^{st}_{ijm}, and W^{st}_{ijm} implicitly determined for solution Y. With this in place, we then update the objective by adding Δc_1, Δc_2, Δc_3, and Δc_4, as defined below.

$$\triangle c_1 = \sum_{i<\ell}\sum_{s}\left(c'_{i\ell k_{i_{[s]}}k_{\ell_{[r']}}}\tilde{\lambda}^{sr}_{i\ell k_{i_{[s]}}k_{\ell_{[r']}}} - c'_{i\ell k_{i_{[s]}}k_{\ell_{[r]}}}\lambda^{sr}_{i\ell k_{i_{[s]}}k_{\ell_{[r]}}}\right) + \sum_{i<\ell}\sum_{s}\left(c'_{i\ell k_{i_{[s]}}k_{\ell_{[r]}}}\tilde{\lambda}^{sr'}_{i\ell k_{i_{[s]}}k_{\ell_{[r]}}} - ac'_{i\ell k_{i_{[s]}}k_{\ell_{[r']}}}\lambda^{sr'}_{i\ell k_{i_{[s]}}k_{\ell_{[r']}}}\right) \quad (31)$$

$$\triangle c_2 = \sum_{j>\ell}\sum_{t}\left(c'_{\ell jk_{\ell_{[r']}}k_{j_{[t]}}}\tilde{\lambda}^{rt}_{\ell jk_{\ell_{[r']}}k_{j_{[t]}}} - c'_{\ell jk_{\ell_{[r]}}k_{j_{[t]}}}\lambda^{rt}_{\ell jk_{\ell_{[r]}}k_{j_{[t]}}}\right) + \sum_{j>\ell}\sum_{t}\left(c'_{\ell jk_{\ell_{[r]}}k_{j_{[t]}}}\tilde{\lambda}^{r't}_{\ell jk_{\ell_{[r]}}k_{j_{[t]}}} - c'_{\ell jk_{\ell_{[r']}}k_{j_{[t]}}}\lambda^{r't}_{\ell jk_{\ell_{[r']}}k_{j_{[t]}}}\right) \quad (32)$$

$$\triangle c_3 = \sum_{i<\ell}\sum_{s}\sum_{r<t<r'} c'_{i\ell k_{i_{[s]}}k_{\ell_{[t]}}}\left(\tilde{\lambda}^{st}_{i\ell k_{i_{[s]}}k_{\ell_{[t]}}} - \lambda^{st}_{i\ell k_{i_{[s]}}k_{\ell_{[t]}}}\right) \quad (33)$$

$$\triangle c_4 = \sum_{j>\ell}\sum_{t}\sum_{r<s<r'} c'_{\ell jk_{\ell_{[s]}}k_{j_{[t]}}}\left(\tilde{\lambda}^{st}_{\ell jk_{\ell_{[s]}}k_{j_{[t]}}} - \lambda^{st}_{\ell jk_{\ell_{[s]}}k_{j_{[t]}}}\right) \quad (34)$$

Values Δc_1 and Δc_2 correspond to the incremental cost (for any node $i < \ell$ and $j > \ell$, respectively) associated with node ℓ allocating to hub location $k_{\ell_{[r']}}$ at level-r and hub location $k_{\ell_{[r]}}$ at level-r'. Values Δc_3 and Δc_4, meanwhile, correspond to the the incremental cost (for any node $i < \ell$ and $j > \ell$, respectively) associated with node ℓ allocating to hubs at levels strictly between r and r'. Preliminary experiments showed that using (31)-(34) reduced overall run time by roughly 10-fold compared to recomputing all λ^{st}_{ijkm} following a swap.

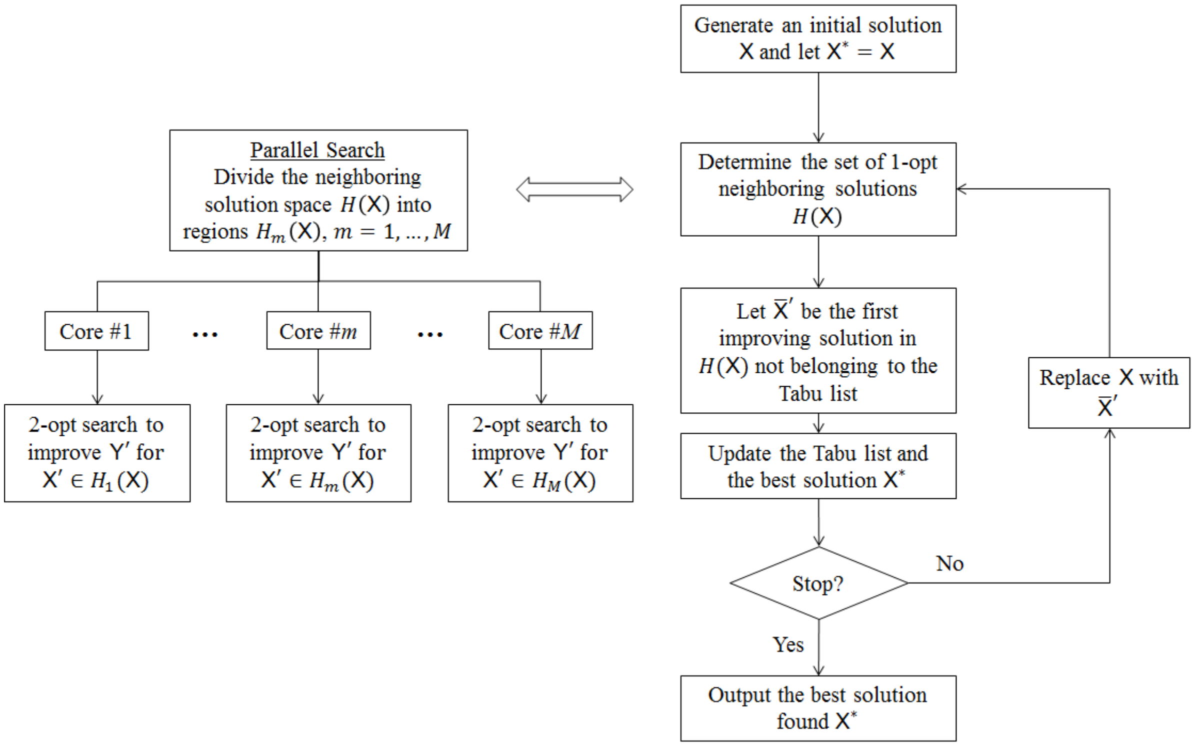


Figure 4: Flow chart of our proposed Tabu search with parallel computing for UHMP.

3.3 Parallel Computing Strategy

We implemented a parallel computing strategy for the Tabu search heuristic aimed at reducing the time involved with evaluating neighboring solutions. By allocating a subset of neighbors to each CPU core and then performing hub assignment searches for each neighboring solution on different cores, we were able to save considerable computation time on finding an improving Tabu search move. Preliminary results showed that run times were roughly 6 times faster by using a parallel computing strategy.

A flow chart of our parallelized Tabu search procedure is shown in Figure 4. The set of neighboring solutions to X is denoted by $H(\mathsf{X})$, while $H_m(\mathsf{X})$, $m = 1, ...M$, signifies a "region" of the full neighborhood $H(\mathsf{X}) = \bigcup_{m \in M} H_m(\mathsf{X})$. The number of regions M is determined based on the number CPU cores. The parallel search of the neighboring solution space $H(\mathsf{X})$ stops as soon as an improving move in any region $m = 1, \ldots, M$ is found.

To construct an initial solution for the Tabu search, we first determine for each possible hub site $k \in N$ the total unit transportation cost $\sum_i c_{ik}$ between each demand node i and the hub site. We then sort the hub sites in ascending order of total unit transportation cost and select the first p sites as hub locations. Alternatively, an initial solution can be obtained by applying the same constructive procedure but with hub sites sorted according to $\sum_i \sum_j c'_{ijkk}$, a measure of total demand-weighted transportation cost. We implemented both constructive procedures and select the better of the two solutions to be the initial solution of the Tabu search. Termination conditions for the Tabu search include a maximum number of iterations and a time limit. If either condition is met, the Tabu search terminates and the best solution found is reported.

4 Results and Analysis

In this section, we investigate the computational efficiency of solving the unreliable p-hub median problem UHMP using either the linear model UHMP2 or the parallelized Tabu search algorithm (hereafter referred to as PCTS). We also analyze the benefits of incorporating facility reliability into the design of hub systems by comparing our solutions with those obtained by solving a standard single allocation p-hub median problem (denoted HMP). The comparison is made with respect to various performance measures.

The model UHMP2 and the algorithm PCTS were both implemented in C++ and run on the same Microsoft Windows 7 Enterprise PC with an Intel Core i7-3770 processor (3.40 GHz per chip) and 24 GB of RAM. UHMP2 was formulated and solved using the IBM ILOG CPLEX version 12.5 callable library. In our computational analysis, we set a time limit of 24 hours for UHMP2 and 3 hours for PCTS. Based on preliminary experiments, parameter values for PCTS were chosen as follows: the maximum number of iterations was set to 30, the length of the Tabu list was 5, and the number of regions M in the parallel computing implementation was 8 (the same as the number of cores on the Intel Core i7 processor). Increasing the length of Tabu list did not produce any substantive changes in solution quality given the small number of iterations that were performed.

4.1 Datasets and Parameter Settings

Our computational experiments were run on two well-known benchmark hub location datasets available from the OR-Library (Beasley, 1990).

- The Civil Aeronautics Board (CAB) dataset (O'Kelly, 1987) consists of airline passenger flow volumes and travel distances between 25 major cities in the United States. From this dataset, we extracted 4 problem instances of different size: number of demand nodes $n = 10, 15, 20, 25$. Each instance was solved with the number of hubs $p = 2, 3, 4, 5$. We set the inter-hub transfer cost rate $\gamma = 0.7$ and the collection and distribution unit costs $\chi = \delta = 1$.
- The Australia Post (AP) dataset (Ernst and Krishnamoorthy, 1996) consists of the locations of 200 postcode districts and their associated flow volumes. In order to generate smaller sized problems, we aggregated the nodes following the procedure described in (Ernst and Krishnamoorthy, 1996). We produced instances with $n = 10, 20, 25, 40, 50, 100, 200$. As with the CAB dataset, we considered the number of hubs $p = 2, 3, 4, 5$. In accordance with the settings used in other articles, the cost parameter values were set as follows: $\gamma = 0.75$, $\chi = 3$, and $\delta = 2$.

For both datasets, failure probabilities q_k were generated uniformly in the range 0.02 to 0.10. The penalty term ϕ was set equal to $\tau \times \max\{c_{ijkm}\}$, with$\tau = 10$. In the computation of ϕ, only the cost of feasible routes were considered (i.e., routes (i, j, k, m) in set A defined in Section 2.3).

4.2 Computational Performance of UHMP2 and PCTS

In this part of our analysis, we compare the computational performance of UHMP2 and the PCTS algorithm on the CAB and AP test problems. The results, shown in Table 1, are only reported for instances $n \leq 20$, as

Table 1: Results for UHMP2 and PCTS on the CAB and AP instances.

			UHMP2				PCTS		
Dataset	n	p	Obj	LB	Time (s)*	Gap (%)	Obj	Time (s)	Gap (%)
CAB	10	2	819.28	-	11.81	0.00	819.28	0.38	0.00
		3	691.80	-	127.64	0.00	691.80	0.88	0.00
		4	634.04	-	1,279.93	0.00	634.04	2.12	0.00
		5	578.74	-	8,290.94	0.00	578.74	3.90	0.00
	15	2	2,898.62	-	1,143.15	0.00	2,898.62	1.15	0.00
		3	2,580.08	-	10,520.50	0.00	2,580.08	4.89	0.00
		4	2,361.88	-	29,735.68	0.00	2,361.88	12.34	0.00
		5	2,200.42	-	70,332.89	0.00	2,200.42	23.59	0.00
	20	2	6,909.40	-	27,745.07	0.00	6,909.40	3.85	0.00
		3	6,143.70	-	81,125.66	0.00	6,143.70	10.89	0.00
		4	5,690.25	5,471.39	-	3.85	**5,615.72**	27.13	**2.57**
		5	5,564.31	4,796.82	-	13.79	**5,215.34**	72.76	**8.02**
AP	10	2	175.09	-	31.84	0.00	175.09	0.36	0.00
		3	129.39	-	212.02	0.00	129.39	1.05	0.00
		4	107.22	-	1,211.69	0.00	107.22	3.16	0.00
		5	92.36	-	33,574.35	0.00	92.36	5.47	0.00
	20	2	179.58	-	37,145.80	0.00	179.58	2.55	0.00
		3	150.02	141.53	-	5.66	**147.40**	10.24	**3.98**
		4	137.85	128.83	-	6.54	**134.11**	24.06	**3.94**
		5	134.11	111.76	-	16.67	**121.61**	46.40	**8.10**
Avg						2.33		12.86	**1.33**

* A "-" indicates that the 24-hour time limit was reached without obtaining a proven optimal solution.

larger problems could not be solved by UHMP2 due to the large number of variables and constraints. Note that for problem instances with $n = 20$ and $p = 5$, the UHMP2 model has about 4 million variables (2020 binary) and 8 million constraints. This is a clear challenge for any available commercial solver.

For each combination of the parameters n and p, Table 1 shows for each of the two solution approaches the expected cost (Obj), CPU computing time in seconds (Time), and the percentage optimality gap (Gap) relative to the linear programming lower bound (LB) given by CPLEX. Expected costs in columns Obj and LB were divided by 10^6 for the CAB instances and by 10^3 for the AP instances.

The results show that UHMP2 is able to solve to optimality on all of the $n < 20$ instances. For the $n = 20$ instances, optimality gaps were still significant in some cases, up to 13.79% to 16.67%, even after 24 hours of computing time. Overall, the average optimality gap for UHMP2 was 2.33%. Optimality gaps for the PCTS algorithm, in contrast, were much smaller, just 1.33% on average and 8.10% in the worst case. We note that the non-zero optimality gaps for PCTS are, in all cases, based on the linear programming lower bounds to UHMP2 rather than proven optimal solutions. The reported gaps, as such, do not give a definitive picture of PCTS solution quality. Indeed, PCTS found optimal solutions for instances that could be solved to optimality by UHMP2 and identified better feasible solutions (highlighted in bold) for instances that could not be solved to optimality by UHMP2, giving us some confidence that the heuristic is able to produce high quality solutions.

Computing times for the PCTS algorithm were negligible compared to UHMP2 (12.86 seconds on average). The performance of PCTS on larger problems is discussed further in the next section.

4.3 Comparison of Standard and Unreliable p-Hub Median Models

To assess the utility of our modeling approach, solutions to UHMP and a standard single allocation p-hub median problem (HMP) were evaluated with respect to the following performance measures: normal operating cost without disruptions (denoted $NCost$), expected transportation cost including the penalty cost (denoted $ECost$), network capacity (denoted Cap), and expected unit cost (denoted $UCost$). The normal operating cost $NCost$ of a solution to UHMP can be easily computed by considering only the probabilities associated with first level assignments λ^{11}_{ijkm}. $NCost$ for a standard p-hub median solution is simply the value of the objective function. Expected transportation cost $ECost$ of an unreliable p-hub solution is given by the objective function value of UHMP. To compute $ECost$ for a solution to HMP, the solution's hub assignments were used as the level-1 hub assignments in UHMP; the PCTS algorithm was then run to determine the higher-level assignments ($s > 1$) with hub locations and level-1 hub assignments fixed. The optimized assignments were used to derive the assignment probabilities λ^{st}_{ijkm} and the overall expected cost for an HMP solution.

The performance measure Cap (Klincewicz, 1998) is defined as the ratio of the expected value of served demand over the total demand in the network:

$$Cap = \frac{EDemand}{\sum_i \sum_j h'_{ij}} \times 100. \tag{35}$$

The $UCost$ measure was calculated as the ratio of the expected transportation cost (including the penalty cost) over the expected value of served demand:

$$UCost = \frac{ECost}{EDemand}. \tag{36}$$

In both formulas (35) and (37), the expected value of served demand is computed as:

$$EDemand = \sum_i \sum_{j>i} \sum_k \sum_m \sum_s \sum_t h'_{ij} \lambda^{st}_{ijkm} \tag{37}$$

Note that for HMP, the assignment probabilities λ^{st}_{ijkm} in formula (37) were obtained using the procedure described above for computing $ECost$.

The results of our comparative analysis for the CAB dataset are displayed in Table 2. The solutions to HMP are known to be optimal in all cases. We observe that the solutions of the two models differ only for $p = 2$ and in only one case for $p = 3$. When the solutions do differ, our model has lower expected cost, lower unit cost, and higher Cap in all instances. On the other hand, normal operating costs for the standard model are lower. On balance, our model gives up a little on normal operating cost (0.83% higher on average) but improves more on expected cost (2.47% lower) and unit cost (2.52% lower). The most noticeable gain for the reliable model is observed for the $n = 25$ and $p = 2$ instance, where our model reduces expected cost by 24.64%, while increasing normal operating cost by just 4.13%. The fact that UHMP performs much better than HMP in terms of expected cost when the number of hubs is small is quite easily explained by the fact that even one hub failure can, in such circumstances, produce a substantial increase in transportation costs because of the few remaining hubs through which flows can be routed.

Table 2: Comparison between UHMP (solved by PCTS) and HMP for the CAB instances.

		UHMP						HMP				
n	p	Hubs	$ECost$	$NCost$	Cap	$UCost$	Time (s)	Hubs	$ECost$	$NCost$	Cap	$UCost$
10	2	(4, 7)	**819.28**	779.24	**99.95**	**820.45**	0.38	(7, 9)	855.10	**761.04**	99.89	856.88
	3	(4, 6, 7)	691.80	681.22	100.00	692.48	0.88	(4, 6, 7)	691.80	681.22	100.00	692.48
	4	(4, 6, 7, 8)	634.04	619.24	100.00	634.66	2.12	(4, 6, 7, 8)	634.04	619.24	100.00	634.66
	5	(1, 4, 6, 7, 8)	578.74	561.78	100.00	579.30	3.90	(1, 4, 6, 7, 8)	578.74	561.78	100.00	579.30
15	2	(4, 7)	**2,898.62**	2,803.34	**99.95**	**1,226.23**	1.15	(4, 11)	3,058.56	**2,778.40**	99.82	1,295.58
	3	(4, 7, 8)	**2,580.08**	2,518.48	100.00	**1,091.01**	4.89	(4, 7, 12)	2,580.98	**2,512.38**	100.00	1,091.39
	4	(1, 4, 7, 12)	2,361.88	2,303.36	100.00	998.70	12.34	(1, 4, 7, 12)	2,361.88	2,303.36	100.00	998.70
	5	(1, 4, 6, 7, 12)	2,200.42	2,145.30	100.00	930.43	23.59	(1, 4, 6, 7, 12)	2,200.42	2,145.30	100.00	930.43
20	2	(2, 4)	**6,909.40**	6,670.40	**99.95**	**1,201.25**	3.85	(4, 17)	7,127.98	**6,547.48**	99.88	1,240.19
	3	(4, 7, 17)	**6,143.70**	6,007.25	**100.00**	**1,067.64**	10.89	(4, 12, 17)	6,176.38	**5,942.26**	100.00	1,073.42
	4	(1, 4, 12, 17)	5,615.72	5,410.40	100.00	975.87	27.13	(1, 4, 12, 17)	5,615.72	5,410.40	100.00	975.87
	5	(1, 4, 6, 12, 17)	**5,215.34**	5,063.37	**100.00**	**906.29**	72.76	(4, 7, 12, 14, 17)	5,254.32	**5,032.94**	100.00	913.07
25	2	(2, 4)	**11,487.50**	11,095.34	**99.95**	**1,345.79**	5.58	(12, 20)	15,244.16	**10,654.90**	99.45	1,794.93
	3	(2, 4, 12)	9,645.22	9,361.54	100.00	1,129.47	18.33	(2, 4, 12)	9,645.22	9,361.54	100.00	1,129.47
	4	(2, 4, 7, 12)	**8,999.56**	8,768.18	**100.00**	**1,053.81**	52.95	(1, 4, 12, 18)	9,053.62	**8,662.86**	100.00	1,060.43
	5	(1, 2, 4, 7, 12)	**8,493.44**	8,247.97	**100.00**	**994.55**	113.22	(1, 4, 7, 12, 18)	8,537.60	**8,179.14**	100.00	1,000.53
Avg					**99.99**	**978.00**	22.12				99.94	1,016.71
Avg Difference (%)			- 2.47	0.83	0.05	- 2.52						

Table 3: Comparison between UHMP (solved by PCTS) and HMP for the AP instances.

		UHMP						HMP				
n	p	Hubs	$ECost$	$NCost$	Cap	$UCost$	Time (s)	Hubs	$ECost$	$NCost$	Cap	$UCost$
10	2	(3, 7)	175.09	167.49	99.82	44.09	0.36	(3, 7)	175.09	167.49	99.82	44.09
	3	(3, 7, 8)	**129.39**	136.77	99.99	**32.52**	1.05	(3, 4, 7)	131.19	**136.01**	99.99	32.98
	4	(3, 4, 7, 8)	107.22	112.40	100.00	26.95	3.16	(3, 4, 7, 8)	107.22	112.40	100.00	26.95
	5	(1, 3, 4, 7, 8)	92.36	91.10	100.00	23.21	5.47	(1, 3, 4, 7, 8)	92.36	91.10	100.00	23.21
20	2	(6, 14)	179.58	172.82	99.89	45.18	2.55	(6, 14)	179.58	172.82	99.89	45.18
	3	(6, 11, 14)	**147.40**	152.65	99.99	**37.05**	10.24	(6, 12, 14)	148.51	**151.53**	99.99	37.32
	4	(2, 6, 11, 14)	**134.11**	136.88	100.00	**33.71**	24.06	(2, 6, 12, 14)	134.97	**135.62**	100.00	33.92
	5	(2, 6, 12, 13, 14)	121.61	123.13	100.00	30.56	46.40	(2, 6, 12, 13, 14)	121.61	123.13	100.00	30.56
25	2	(13, 18)	**188.82**	190.74	**99.94**	**47.48**	7.46	(8, 18)	199.57	**175.54**	99.79	50.26
	3	(7, 14, 18)	154.16	155.26	99.99	38.75	17.42	(7, 14, 18)	154.16	155.26	99.99	38.75
	4	(2, 7, 14, 18)	138.00	139.20	100.00	34.68	46.61	(2, 7, 14, 18)	138.00	139.20	100.00	34.68
	5	(2, 7, 14, 17, 18)	122.88	123.57	100.00	30.88	119.97	(2, 7, 14, 17, 18)	122.88	123.57	100.00	30.88
40	2	(20, 28)	**201.33**	192.34	**99.91**	**50.65**	34.15	(12, 28)	237.60	**177.47**	99.61	59.95
	3	(12, 22, 28)	162.71	158.83	99.99	40.90	83.48	(12, 22, 28)	162.71	158.83	99.99	40.90
	4	(12, 22, 26, 28)	145.90	143.97	100.00	36.67	251.89	(12, 22, 26, 28)	145.90	143.97	100.00	36.67
	5	(3, 12, 22, 26, 28)	135.61	134.26	100.00	34.08	463.44	(3, 12, 22, 26, 28)	135.61	134.26	100.00	34.08
50	2	(14, 36)	**201.58**	179.18	**99.86**	**50.73**	65.15	(14, 35)	211.40	**178.48**	99.81	53.23
	3	(14, 33, 36)	**164.39**	162.58	**99.99**	**41.32**	186.03	(14, 28, 35)	166.00	**158.57**	99.98	41.73
	4	(14, 28, 33, 35)	146.56	143.38	100.00	36.83	318.88	(14, 28, 33, 35)	146.56	143.38	100.00	36.83
	5	(4, 14, 28, 33, 35)	136.19	132.37	100.00	34.23	462.58	(4, 14, 28, 33, 35)	136.19	132.37	100.00	34.23
Avg					**99.97**	**37.52**	107.52				99.94	38.32
Avg Difference (%)			- 1.45	1.11	0.03	- 1.47						

Table 4: Impact of failure probability q on solution diversity and quality for the CAB instances.

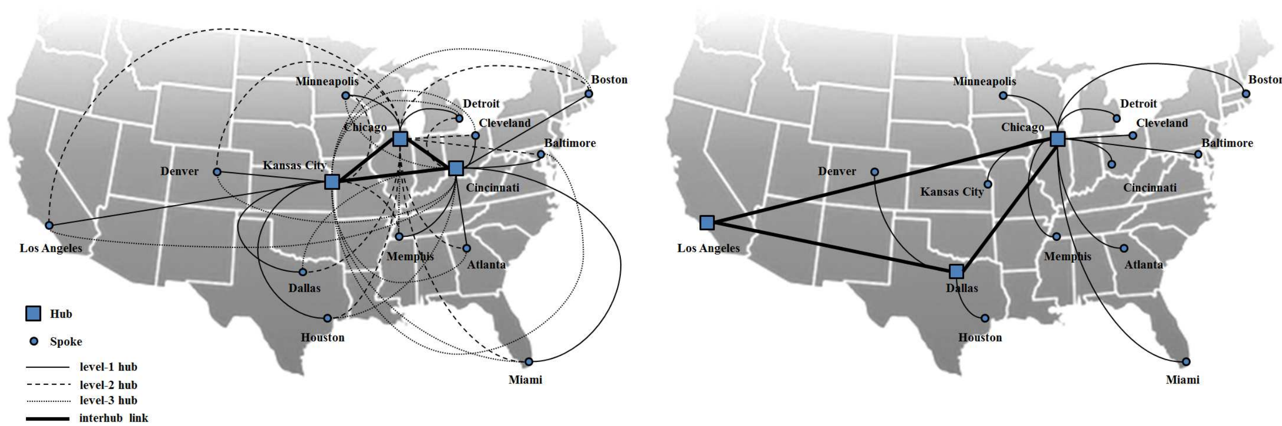
q	No. of Different Solutions / No. of Instances	Avg Change in $ECost$ (%)	Avg Change in $NCost$ (%)
0.05	4 / 16	- 0.30	0.33
0.10	9 / 16	- 0.86	0.88
0.15	13 / 16	- 1.60	2.50
0.20	15 / 16	- 2.29	3.41
0.25	15 / 16	- 2.83	3.83
0.30	15 / 16	- 3.16	5.00

Table 3 presents a comparison for the AP dataset. Note that optimal solutions to HMP can be obtained for problem instances with $n \leq 50$. For $n = 100$ and 200, we used the best known solutions found by Ilić et al. (2010). The same general observations made for the CAB dataset also hold for the AP instances, although there is slightly greater variability in the solutions identified by the two models. Three of the solutions differed for $p = 3$ and one solution was different for $p = 4$. The gain in expected costs relative to the increase in normal costs is less pronounced for this dataset. However, there are a few cases where the improvement in expected cost is still significant. An example is the instance with $n = 40$ and $p = 2$, where our model reduces expected cost by 15.27% with an 8.38% increase in normal operating cost.

Based on our analysis, it can be inferred that solutions to a standard p-hub median problem are fairly reliable and generally similar to the UHMP solutions (the ratio of different solutions to the total number of instances is 9 / 16 for CAB and 7 / 20 for AP). This is particularly true when the number of hubs being located is high (i.e., $p = 4$ and 5). This phenomenon can be attributed in part to the fairly low hub failure probabilities used in our tests. To investigate further the role failure probabilities play in affecting the similarity of solutions, we solved the instances of the CAB dataset for a range of equal failure probabilities q varying between 0.05 and 0.3 in steps of 0.05. Table 4 shows that the number of times solutions to UHMP and HMP differ increases with increasing values of the failure probability. For any value of q greater than 0.2, 15 out of the 16 solutions were different. This result shows, as one might intuitively expect, that the solutions to a standard p-hub median problem become less reliable when the probability of failure is high. While it is true that the average percentage increase in normal costs is higher relative to the average percentage decrease in expected costs (over the range of q values analyzed), solutions to our model nonetheless perform better in an expected sense and, in some cases, can be substantially better. As a final remark, we point out that expected costs for the HMP solutions were obtained by using our model to optimize customer assignments in case of failure. This certainly had a positive impact on the overall quality of the HMP solutions as measured in terms of $ECost$.

An illustrative example of how solutions to our model can differ from a standard p-hub median model is displayed in Figure 5 for one of the CAB instances used in Table 4. The figure shows how the hub locations based our model are clustered in the central part of US, an area with high customer demand. The standard hub network, in contrast, has a hub located at Los Angeles which only serves itself. The closer concentration of hubs guarantees that in case of failure, backup hubs will not be too far away from customers that need to be reassigned.

Based on our observation that the HMP solutions are in many cases quite reliable, we modified the PCTS algorithm in order to improve its efficiency at solving large-sized problems. Specifically, we considered using a two-phase approach where, in the first phase, a standard p-hub median solution was generated and in the



(a) Reliable hub network

(b) Standard hub network

Figure 5: Illustration of reliable versus standard hub networks for the CAB instance $n = 15$, $p = 3$ and a fixed hub failure probability $q = 0.15$

second phase this solution was used as the starting solution for the PCTS algorithm. To generate an HMP solution in phase 1, we modified the PCTS algorithm so that the sequential 2-opt search of hub assignments was restricted to the level-1 assignments only and moves were evaluated with respect to the objective of a standard HMP model.

The two-phase approach was tested on some large instances of the AP dataset with $n = 100$ and 200. We used 100 iterations for phase 1 and the same parameter settings described previously for PCTS in phase 2. Results obtained by solving UHMP using the PCTS algorithm with and without phase 1 are shown in Table 5. The table also shows a comparison with the solutions of the standard p-hub median problem based on our defined performance measures. We point out that in all cases, phase 1 of PCTS produced either an optimal solution or the best known solution to HMP (results not shown).

The benefits of using a two-phase approach are clearly evident. Without phase 1, there are two cases (i.e., $n = 200$, $p = 4, 5$) where the standard model finds better solutions than our model in terms of expected and unit costs. This implies that the solutions found by PCTS are sub-optimal. With phase 1, however, the PCTS algorithm produces better solutions in all cases in terms of expected and unit costs. Average reduction in expected cost improves from 4.91% to 6.70%, while unit cost improves on average from 4.99% to 6.78% lower. At the same time, the average increase in normal operating cost is reduced from 4.20% to 3.32% above HMP. The reduction in expected cost is significant for the instance $n = 100$, $p = 2$ (14.48%) and for the instance with $n = 200$, $p = 2$ (28.49%).

4.4 Discount Factor Sensitivity Analysis

In attempt to understand the influence of the discount factor γ on expected hub and spoke flows for UHMP and HMP, we carried out experiments on the AP dataset with $n = 20$ nodes in which parameters p and γ were systematically varied. Minimum and maximum expected hub arc flows, as well as the average percentage of spoke arcs with flow greater than hub arcs, are reported in Table 6.

Table 5: Performance of PCTS without phase 1 (upper portion) and with phase 1 (lower portion) on the large AP instances.

		PCTS						HMP				
n	p	Hubs	$ECost$	$NCost$	Cap	$UCost$	Time (s)*	Hubs	$ECost$	$NCost$	Cap	$UCost$
100	2	(26, 92)	**206.47**	197.33	**99.93**	**51.93**	542.33	(28, 71)	241.43	**180.22**	99.66	60.88
	3	(28, 65, 73)	**168.59**	166.30	**99.99**	**42.37**	964.67	(28, 55, 70)	170.14	**160.85**	99.98	42.77
	4	(28, 55, 66, 71)	**150.11**	146.34	100.00	**37.73**	3,215.61	(28, 55, 64, 70)	150.79	**145.90**	100.00	37.90
	5	(7, 28, 55, 66, 70)	**140.58**	137.09	100.00	**35.33**	-	(7, 28, 55, 64, 70)	141.21	**136.93**	100.00	35.49
200	2	(54, 147)	**213.02**	203.28	**99.94**	**53.57**	4,856.04	(56, 140)	297.90	**182.46**	99.40	75.32
	3	(53, 107, 140)	181.83	162.89	99.96	45.72	9,225.17	(53, 107, 140)	181.83	162.89	99.96	45.72
	4	(56, 106, 131, 140)	156.89	149.82	100.00	39.43	-	(56, 110, 131, 140)	**156.76**	**147.77**	100.00	**39.40**
	5	(61, 101, 112, 131, 141)	154.61	150.55	100.00	38.86	-	(14, 61, 113, 131, 140)	**146.64**	**140.06**	100.00	**36.85**
Avg					**99.98**	**43.12**					99.88	46.79
Avg Difference (%)			- 4.91	4.20	0.10	- 4.99						
100	2	(26, 92)	**206.47**	197.33	**99.93**	**51.93**	618.90	(28, 71)	241.43	**180.22**	99.66	60.88
	3	(28, 65, 73)	**168.59**	166.30	**99.99**	**42.37**	1,152.77	(28, 55, 70)	170.14	**160.85**	99.98	42.77
	4	(28, 55, 66, 71)	**150.11**	146.34	100.00	**37.73**	3,572.55	(28, 55, 64, 70)	150.79	**145.90**	100.00	37.90
	5	(7, 28, 55, 66, 70)	**140.58**	137.09	100.00	**35.33**	-	(7, 28, 55, 64, 70)	141.21	**136.93**	100.00	35.49
200	2	(54, 147)	**213.02**	203.28	**99.94**	**53.57**	5,190.68	(56, 140)	297.90	**182.46**	99.40	75.32
	3	(57, 103, 141)	**171.92**	165.23	**99.99**	**43.20**	-	(53, 107, 140)	181.83	**162.89**	99.96	45.72
	4	(57, 108, 130, 141)	**152.90**	148.05	100.00	**38.43**	-	(56, 110, 131, 140)	156.76	**147.77**	100.00	39.40
	5	(14, 61, 112, 130, 141)	**145.36**	140.42	100.00	**36.53**	-	(14, 61, 113, 131, 140)	146.64	**140.06**	100.00	36.85
Avg					**99.98**	**42.39**					99.88	46.79
Avg Difference (%)			- 6.70	3.32	0.11	- 6.78						

* A "-" indicates that the 3-hour time limit was reached.

Table 6: Discount factor effects on inter-hub flows of UHMP (solved by PCTS) and HMP for selected AP $n = 20$ instances.

		UHMP					HMP				
				Hub Arc Flows		Spoke Arcs			Hub Arc Flows		Spoke Arcs
p	γ	Hubs	*ECost*	Min	Max	with Flow > Hub Arcs (%)	Hubs	*ECost*	Min	Max	with Flow > Hub Arcs (%)
3	0.2	(6, 12, 14)	129.66	366.0	866.9	5	(6, 12, 14)	129.66	366.0	866.9	5
	0.4	(6, 12, 14)	136.56	318.1	813.3	10	(6, 12, 14)	136.56	318.1	813.3	10
	0.6	**(6, 11, 14)**	**143.36**	**366.7**	**801.3**	**5**	(6, 12, 14)	**143.38**	**318.1**	**813.3**	**10**
	0.8	**(6, 11, 14)**	**148.75**	**366.7**	**801.3**	**5**	(6, 12, 14)	**150.13**	**265.8**	**865.6**	**15**
4	0.2	(2, 6, 12, 14)	111.58	88.1	677.6	38	(2, 6, 12, 14)	111.58	88.1	677.6	38
	0.4	(2, 6, 12, 14)	120.13	88.1	725.2	43	(2, 6, 12, 14)	120.13	88.1	725.2	43
	0.6	(2, 6, 12, 14)	128.62	85.3	727.9	43	(2, 6, 12, 14)	128.62	85.3	727.9	43
	0.8	**(2, 6, 11, 14)**	**135.88**	**109.6**	**786.2**	**43**	(2, 6, 12, 14)	**137.04**	**70.9**	**672.7**	**44**
5	0.2	(2, 6, 12, 13, 14)	96.10	65.7	578.9	50	(2, 6, 12, 13, 14)	96.10	65.7	578.9	50
	0.4	(2, 6, 12, 13, 14)	105.45	67.5	626.8	52	(2, 6, 12, 13, 14)	105.45	67.5	626.8	52
	0.6	(2, 6, 12, 13, 14)	114.68	67.5	626.9	52	(2, 6, 12, 13, 14)	114.68	67.5	626.9	52
	0.8	**(2, 6, 12, 13, 14)**	**123.89**	**67.5**	**585.8**	**53**	(2, 4, 6, 11, 14)	**124.70**	**21.8**	**688.0**	**58**

As observed with various other hub location models (O'Kelly, 1987; Campbell, 1994; O'Kelly and Bryan, 1998), inter-hub flows for UHMP are imbalanced; a few inter-hub links tend to have very high flows, while others have comparatively small flows. This is clearly seen with the $p = 4$ and 5 instances, where the minimum inter-hub flows are 11-14% of the maximum inter-hub flows. Interestingly, we see that in cases where solutions to UHMP with HMP differ (highlighted in bold), UHMP yields a somewhat more equitable distribution of inter-hub flows, as indicated by the higher minimum flows and generally lower maximum flows along hub arcs.

Another observation that can be gleaned from Table 6 is that economies of scale are not being fully captured in UHMP. In particular, a significant proportion of spoke arcs, up to 53%, have flows exceeding hub arc flows. This does not adhere to one of the core premises of a hub-and-spoke architecture, namely that discounting is justified by the concentration of flows between hubs. This is a common problem, which a number of studies have documented (O'Kelly and Bryan, 1998; Campbell et al., 2005), when hubs are assumed to be fully connected. It is important to point out, however, that in comparison to a standard hub median model, UHMP tends to mitigate this issue to some extent. In the 4 cases where UHMP produced a different solution from HMP, the mean percentage of spoke arcs with flow greater than hub arc was strictly less. Results for CAB $n = 20$ instances (results not shown) were qualitatively similar, indicating that our findings regarding UHMP hub and spoke flows may be generalizable.

4.5 Failure Probability Sensitivity Analysis

In this section, we analyze the robustness of the UHMP solutions to variations in the failure probabilities and evaluate the possible impacts on solution quality of using inaccurate probability values. For example purposes, the analysis is carried out on the CAB $n = 10$ instance with equal hub failure probabilities q in the range $[0.0, 0.05, \ldots, 0.3]$.

Let X_A^* (X_T^*) be the optimal hub locations obtained by solving UHMP2 with an *assumed* (*true*) failure probability q_A (q_T). We denote by $Z(\cdot)$ the objective function value for a particular solution given that the true failure probability is q_T . We define the relative percentage error in objective value when the assumed probability is q_A and the true probability is q_T as follows:

$$Error = \frac{Z(X_A^*) - Z(X_T^*)}{Z(X_T^*)} \times 100 \tag{38}$$

Computational results of solution robustness for UHMP are presented in Table 7 for each combination of assumed and true hub failure probabilities. The last two columns report the average and maximum percentage error for each assumed probability value. The results suggest that for small values of p, it is better to assume a small probability of failure. Conversely, for large values of p, it is better to assume a large probability of failure. This can be explained by the fact that when resources are limited (small p), there is little room for recourse and the best option is to plan for normal operating conditions. On the other hand, when resources are greater (large p), there is more room to accommodate for potential failure. The maximum percentage error found (10.34%) was for $p = 2$ when the probability of failure is assumed to be high ($q_A \geq 0.15$) but in reality no failure occurs ($q_T = 0$). Percentage error is quite significant as well for the instances $p = 4$ and 5 when the hub network is planned assuming a low failure probability ($q_A \leq 0.1$) but failures actually occur with high probability ($q_A \geq 0.2$).

Table 7: Solution robustness of CAB $n = 10$ instances.

p	Assumed Probability	True Probability 0.00	0.05	0.10	0.15	0.20	0.25	0.30	Avg Error	Max Error
2	0.00	-	0.00	0.00	1.73	2.58	2.79	2.69	**1.63**	**2.79**
	0.05	0.00	-	0.00	1.73	2.58	2.79	2.69	**1.63**	**2.79**
	0.10	0.00	0.00	-	1.73	2.58	2.79	2.69	**1.63**	**2.79**
	0.15	10.34	4.47	0.38	-	0.00	0.00	0.00	2.53	10.34
	0.20	10.34	4.47	0.38	0.00	-	0.00	0.00	2.53	10.34
	0.25	10.34	4.47	0.38	0.00	0.00	-	0.00	2.53	10.34
	0.30	10.34	4.47	0.38	0.00	0.00	0.00	-	2.53	10.34
3	0.00	-	0.00	0.00	0.00	0.02	0.09	0.12	**0.04**	**0.12**
	0.05	0.00	-	0.00	0.00	0.02	0.09	0.12	**0.04**	**0.12**
	0.10	0.00	0.00	-	0.00	0.02	0.09	0.12	**0.04**	**0.12**
	0.15	0.00	0.00	0.00	-	0.02	0.09	0.12	**0.04**	**0.12**
	0.20	0.72	0.49	0.28	0.10	-	0.00	0.00	0.27	0.72
	0.25	0.72	0.49	0.28	0.10	0.00	-	0.00	0.27	0.72
	0.30	0.72	0.49	0.28	0.10	0.00	0.00	-	0.27	0.72
4	0.00	-	0.00	0.20	1.14	2.79	4.26	5.25	2.27	5.25
	0.05	0.00	-	0.20	1.14	2.79	4.26	5.25	2.27	5.25
	0.10	0.58	0.36	-	0.16	0.91	1.54	2.22	0.96	**2.22**
	0.15	2.40	1.64	0.60	-	0.00	0.00	0.00	**0.77**	2.40
	0.20	2.40	1.64	0.60	0.00	-	0.00	0.00	**0.77**	2.40
	0.25	2.40	1.64	0.60	0.00	0.00	-	0.00	**0.77**	2.40
	0.30	2.40	1.64	0.60	0.00	0.00	0.00	-	**0.77**	2.40
5	0.00	-	0.00	0.00	0.09	0.88	2.50	3.35	1.14	3.35
	0.05	0.00	-	0.00	0.09	0.88	2.50	3.35	1.14	3.35
	0.10	0.00	0.00	-	0.09	0.88	2.50	3.35	1.14	3.35
	0.15	1.92	1.41	0.74	-	0.00	0.00	0.72	**0.80**	**1.92**
	0.20	1.92	1.41	0.74	0.00	-	0.00	0.72	**0.80**	**1.92**
	0.25	1.92	1.41	0.74	0.00	0.00	-	0.72	**0.80**	**1.92**
	0.30	8.72	6.81	4.90	2.93	1.69	0.33	-	4.23	8.72

5 Discussion

The design of reliable hub networks is an important issue that has received relatively limited attention in the literature. In this paper, we present a model for locating unreliable hubs, assuming that multiple hubs can fail simultaneously and that disrupted flows must be rerouted through remaining operational hubs. Besides presenting a nonlinear formulation of the problem, we devise an exact linear version based on the use of a specialized flow network referred to as a probability lattice. A probability lattice extends the concept of a probability chains by linking multiple probability chains together into a backbone and spur structure in order to evaluate joint, compound probability terms. Computational experiments demonstrate the efficiency of our linear model on small-sized instances ($\leq$20 nodes). For large-sized instances, heuristic methods are required. We therefore propose a Tabu search procedure employing a parallel computing strategy to find optimal or near optimal hub locations with reasonable computational effort. A sensitivity analysis was carried out to gain a preliminary understanding of how expected flows on hub and spoke arcs are affected by discounting and how variability in hub failure probability affects solution robustness.

It is worth pointing out some of the caveats and limitations of our study. Firstly, a bit of caution is warranted about over-interpreting the performance of our heuristic solution method. In particular, the results showed that using simple 1-opt hub facility moves managed to produce low optimality gaps. This finding, however, is based on tests involving relatively small problem size instances. With larger problem instances, a 1-opt strategy may not perform as well, in which case the use of larger search neighborhoods may prove beneficial.

A key issue we discovered is that our model suffers from many of the same problems as a standard single allocation p-hub median problem due to the assumption of having a fully connected hub network. As the computational results clearly show, solutions to UHMP can have imbalanced inter-hub flows and, more importantly, spoke-to-hub flows that often exceed hub-to-hub flows. This latter characteristic contradicts a key premise of hub-and-spoke networks, namely that discounting is justified by the concentration of flows between hubs. That being said, we are encouraged by the fact our model tends to restrict the magnitude of both imbalanced flows and excessively high spoke arc flows when compared to a standard hub median problem, suggesting that a reliability framework has additional advantages beyond mitigating against the impacts of hub disruption. Moving forward, an interesting line of research would be to incorporate more realistic cost structures into our hub reliability modeling framework, for example flow discounting (O'Kelly and Bryan, 1998; Bryan, 1998) or hub arc location (Campbell et al., 2005; Alumur et al., 2009), which better account for economies of scale.

Another limitation of our model relates to the manner in which spoke nodes are reassigned to hubs. We employ a level-set approach, whereby spokes assign to hubs in a specified order based on the operational status of hubs. Under this assumption, it would be challenging to incorporate, in a straightforward way, certain model extension such as multiple hub allocation (Ernst and Krishnamoorthy, 1999) or hub capacity constraints (Marín et al., 2006). Formulating a model that incorporates these types of problem considerations would likely involve moving away from a level-set approach to a scenario based approach in which every combination of hub failures is enumerated and explicitly modeled. The trick, of course, would be to devise ways of dealing with the combinatorial explosion in the number of scenarios as the number of hubs increases and attempting to linearizing such a model. Other interesting lines of research would be to consider correlated hub failures and the use of other hub service protocols such as a maximum covering objective. These are areas we are actively investigating.

Finally, our current model neglects some practical considerations with regard to reassigning disrupted flows. In particular, we do not take into account how far the rerouting process can continue. This stands in contrast to protocols often used in the transportation sector, airlines being a good example, where a journey (flight) is canceled if the added travel travel distance/time exceeds some threshold. This could, in part, be dealt with by only considering assignments up to some level $r < p$ (i.e., assignment levels s and t could range from $1, \ldots, r$ instead of $1, \ldots, p$). This would have the added advantage of reducing the size of model by eliminating variables and constraints associated with assignment levels $> r$. Another possibility would be to introduce additional variables and constraints to enforce a strict limit on how far rerouting is allowed to occur in the event of hub disruption. The obvious downside is that a very large number of additional variables and constraints may be required, which would almost certainly make an already complex model yet more complex and difficult to solve.

Acknowledgments

We are grateful to three anonymous referees and the associate editor for very helpful comments on earlier drafts of this paper.

References

Abdinnour-Helm, S., 1998. A hybrid heuristic for the uncapacitated hub location problem. European Journal of Operational Research 106 (2-3), 489–499.

Aksen, D., Aras, N., 2012. A bilevel fixed charge location model for facilities under imminent attack. Computers & Operations Research 39 (7), 1364–1381.

Alumur, S., Kara, B., 2008. Network hub location problems: The state of the art. European Journal of Operational Research 190 (1), 1–21.

Alumur, S., Kara, B., Karasan, O., 2009. The design of single allocation incomplete hub networks. Transportation Research Part B 43 (10), 936–951.

An, Y., Zhang, Y., Zeng, B., 2011. The reliable hub-and-spoke design problem: Models and algorithms. `http://www.optimization-online.org/DB_FILE/2011/05/3043.pdf`, accessed 01/07/2013.

Azizi, N., Chauhan, S., Salhi, S., Vidyarthi, N., 2014. The impact of hub failure in hub-and-spoke networks: Mathematical formulations and solution techniques. Computers & Operations Research, DOI: 10.1016/j.cor.2014.05.012.

Ball, M., Barnhart, C., Nemhauser, G., Odoni, A., 2007. Air transportation: Irregular operations and control. In: Barnhart, C., Laporte, G. (Eds.), Handbook of Operations Research and Management Science: Transportation. Elsevier, pp. 1–67.

Beasley, J., 1990. Or-library. `http://people.brunel.ac.uk/~mastjjb/jeb/orlib/phubinfo.html`, accessed 27/07/2013.

Berman, O., Krass, D., Menezes, M., 2007. Facility reliability issues in network p-median problems: Strategic centralization and co-location effects. Operations Research 55 (2), 332–350.

Bryan, D., 1998. Extensions to the hub location problem: Formulations and numerical examples. Geographical Analysis 30 (4), 315–330.

Camm, J., Norman, S., Polasky, S., Solow, A., 2002. Nature reserve site selection to maximize expected species covered. Operations Research 50 (6), 946–955.

Campbell, J., 1994. Integer programming formulations of discrete hub location problems. European Journal of Operational Research 72, 387–405.

Campbell, J., Ernst, A., Krishnamoorthy, M., 2005. Hub arc location problems: Part I - introduction and results. Management Science 51, 1540–1555.

Campbell, J., O'Kelly, M., 2012. Twenty-five years of hub location research. Transportation Science 46 (2), 153–169.

Chen, J., 2007. A hybrid heuristic for the uncapacitated single allocation hub location problem. Omega 35 (2), 211 – 220.

Cui, T., Ouyang, Y., Shen, Z., 2010. Reliable facility location design under the risk of disruptions. Operations Research 58 (4), 998–1011.

Daskin, M., 1983. A maximum expected covering location model: Formulation, properties and heuristic solution. Transportation Science 17 (1), 48–70.

Drezner, Z., 1987. Heuristic solution methods for two location problem with unreliable facilities. Journal of the Operational Research Society 38 (6), 509–514.

Ebery, J., 2001. Solving large single allocation p-hub problems with two or three hubs. European Journal of Operational Research 128 (2), 447–458.

Ernst, A., Krishnamoorthy, M., 1996. Efficient algorithms for the uncapacitated single allocation p-hub median problem. Location Science 4 (3), 139 – 154.

Ernst, A., Krishnamoorthy, M., 1999. Solution algorithms for the capacitated single allocation hub location problem. Annals of Operations Research 86, 141–159.

Farahani, R., Hekmatfar, M., Arabani, A., Nikbakhsh, E., 2013. Hub location problems: A review of models, classification, solution techniques, and applications. Computers & Industrial Engineering 64 (4), 1096 – 1109.

Glover, F., 1989. Tabu search - part i. ORSA Journal on Computing 1 (3), 190–206.

Ilić, A., Urosevic, D., Brimberg, J., Mladenovic, N., 2010. A general variable neighborhood search for solving the uncapacitated single allocation p-hub median problem. European Journal of Operational Research 206 (2), 289 – 300.

Janić, M., 2005. Modeling the large scale disruptions of an airline network. Journal of Transportation Engineering 131 (4), 249–260.

Kim, H., 2012. P-hub protection models for survivable hub network design. Journal of Geographical Systems 14 (4), 437–461.

Kim, H., O'Kelly, M., 2009. Reliable p-hub location problems in telecommunication networks. Geographical Analysis 41 (3), 283–306.

Klincewicz, J., 1992. Avoiding local optima in the p-hub location problem using tabu search and grasp. Annals of Operations Research 40, 283–302.

Klincewicz, J., 1998. Hub location in backbone or tributary network design: A review. Location Science 6 (1-4), 307 – 335.

Liberatore, F., Scaparra, M., Daskin, M., 2012. Hedging against disruptions with ripple effects in location analysis. Omega 40 (1), 21–30.

Losada, C., Scaparra, M., Church, R., MS, D., 2012b. The stochastic interdiction median problem with disruption intensity levels. Annals of Operations Research 201 (1), 345–365.

Losada, C., Scaparra, M., O'Hanley, J., 2012a. Optimizing system resilience: A facility protection model with recovery time. European Journal of Operational Research 217 (3), 519–530.

Marín, A., Cánovas, L., Landete, M., 2006. New formulations for the uncapacitated multiple allocation hub location problem. European Journal of Operational Research 172 (1), 274 – 292.

Morton, D., Pan, F., Saeger, K., 2007. Models for nuclear smuggling interdiction. IIE Transactions 39, 3–14.

O'Hanley, J., Church, R., 2013. Designing robust coverage to hedge against worstcase facility losses. European Journal of Operational Research 209 (1), 23 – 36.

O'Hanley, J., Scaparra, M., García, S., 2013a. Probability chains: A general linearization technique for modeling reliability in facility location and related problems. European Journal of Operational Research 230 (1), 63 – 75.

O'Hanley, J., Wright, J., Diebel, M., Fedora, M., Soucy, C., 2013b. Restoring stream habitat connectivity: A proposed method for prioritizing the removal of resident fish passage barriers. Journal of Environmental Management 125, 19 – 27.

O'Kelly, M., 1987. A quadratic integer program for the location of interacting hub facilities. European Journal of Operational Research 32 (3), 393 – 404.

O'Kelly, M., Bryan, D., 1998. Hub location with flow economies of scale. Transportation Research Part B 32 (8), 605 – 616.

O'Kelly, M., Bryan, D., Skorin-Kapov, D., Skorin-Kapov, J., 1996. Hub network design with single and multiple allocation: A computational study. Location Science 4 (3), 125 – 138.

O'Kelly, M., Kim, H., Kim, C., 2006. Internet reliability with realistic peering. Environment and Planning Part B 33 (3), 325 – 343.

Parvaresh, F., Moattar Husseini, S., Hashemi Golpayegany, S., Karimi, B., 2012. Hub network design problem in the presence of disruptions. Journal of Intelligent Manufacturing 25 (4), 755 – 774.

Peng, P., Snyder, L., Lim, A., Liu, Z., 2011. Reliable logistics networks design with facility disruptions. Transportation Research Part B 45 (8), 1190–1211.

Qi, L., Shen, Z., LV, S., 2010. The effect of supply disruptions on supply chain design decisions. Transportation Science 44 (2), 274–289.

Scaparra, M., Church, R., 2008. An exact solution approach for the interdiction median problem with fortification. European Journal of Operational Research 189 (1), 76–92.

Silva, M., Cunha, C., 2009. New simple and efficient heuristics for the uncapacitated single allocation hub location problem. Computers & Operations Research 36 (12), 3152 – 3165.

Skorin-Kapov, D., Skorin-Kapov, J., 1994. On tabu search for the location of interacting hub facilities. European Journal of Operational Research 73 (3), 502 – 509.

Skorin-Kapov, D., Skorin-Kapov, J., O'Kelly, M., 1996. Tight linear programming relaxations of uncapacitated p-hub median problems. European Journal of Operational Research 94 (3), 582–593.

Snyder, L., 2006. Facility location under uncertainty: A review. IIE Transactions 38 (7), 547–564.

Snyder, L., Daskin, M., 2005. Reliability models for facility location: The expected failure cost case. Transportation Science 39 (3), 400–416.