XX Models for Reliable Supply Chain Network Design

Lawrence V. Snyder, Mark S. Daskin

[1]Department of Industrial and Systems Engineering, Lehigh University, Bethlehem, PA; *email:* larry.snyder@lehigh.edu
[2]Department of Industrial Engineering and Management Sciences, Northwestern University, Evanston, IL; *email:* m-daskin@northwestern.edu

XX.1 Introduction

Recent examples of disruptions in the news suggest a strong geographical dimension to supply chain disruptions, and to their effects. For example:

- The west-coast port lockout in 2002 strangled U.S. retailers' supply lines while east-coast ports were essentially unaffected (Greenhouse 2002)
- The foot-and-mouth disease scare in the U.K. in 2001 caused the U.S. to ban imports of British meat (Marquis and McNeil 2001).
- The suspension of the license of the Chiron plant in Liverpool, England reduced the U.S. supply of the influenza vaccine by nearly 50% during the 2004/5 flu season (Pollack 2004).
- In the U.S. Gulf Coast region in 2005, Hurricane Katrina idled facilities situated at all levels of the supply chain, including production (e.g., coffee; Barrionuevo and Deutsch 2005), processing (oil refining; Mouawad 2005), warehousing (lumber storage; Reuters 2005), transit (banana imports; Barrionuevo and Deutsch 2005), and retail (groceries and home-repair; Fox 2005, Leonard 2005). These facilities were located in or near New Orleans but were integral parts of global supply chains.

These examples highlight the need for supply chain design models that account for the spatial nature of both supply chains and their operation.

In this chapter, we present several models for reliable facility location in a supply chain that is vulnerable to disruptions. Since facility location decisions are costly to implement and difficult to reverse, these strategic decisions permit very little recourse once a disruption occurs, other than re-assignment of customers to non-disrupted facilities. Our goal, therefore, is to choose facility locations proactively so that the system performs well even if disruptions occur.[1]

Consider the following example. Fig. XX.1 depicts the optimal solution to the uncapacitated fixed-charge location problem (UFLP) for a 49-node data set consisting of the capitals of the 48 continental U.S. states and Washington, DC. All nodes serve as both potential facility location sites and demand points, with demands proportional to state populations. This data set is modified from Daskin (1995). The optimal UFLP solution entails a fixed cost of $386,900 per year to operate the five opened facilities and a transportation cost of $470,228 per year.

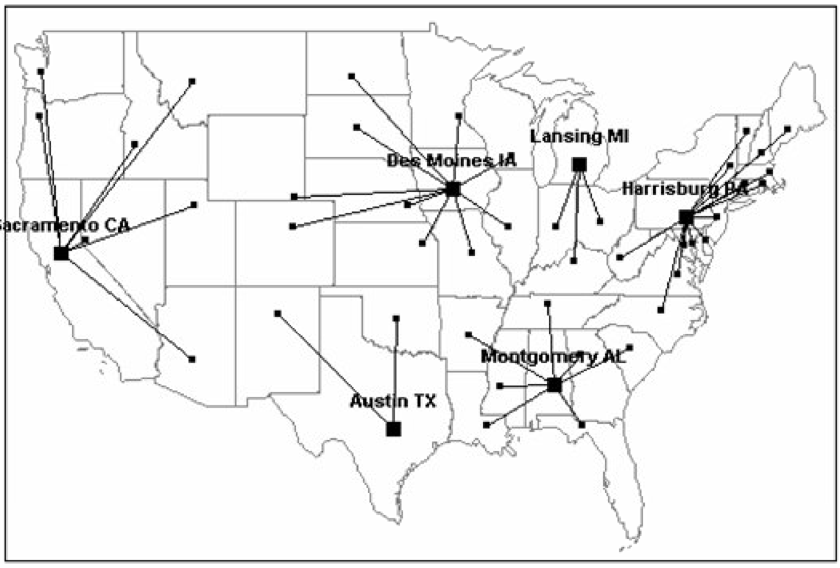

Fig. XX.1. UFLP solution for 49-node dataset

Now suppose that the facility in Sacramento, California becomes unavailable—say, because of a strike or extended power outage. In this case, the west-coast customers served by that facility must instead be served by facilities in Des Moines, Iowa and Austin, Texas (Fig. XX.2), resulting in a transportation cost of $1,019,065, an increase of 117% from the baseline solution.

Table XX.1 lists the "failure costs" (the transportation costs that result after the failure of a facility) for each of the five facilities in the optimal solution, as well as their assigned demands and the transportation cost when no facilities fail. Note that Sacramento serves only 19% of the total demand but generates the largest failure cost because its customers are geographically disparate and the next-closest

[1] In this chapter, we use the terms "failure" and "disruption" interchangeably.

facility is quite distant. The Harrisburg facility serves customers that are tightly clustered, and good "backup" facilities are fairly close by, but its failure cost is still quite large (a 52% increase in transportation cost) because of the volume of demand that it serves. In contrast, Montgomery serves nearly as much demand as Sacramento, but because it is centrally located, close to backup facilities, its failure cost is smaller than that of Sacramento or Harrisburg. Therefore, the reliability of a facility depends on both the demand served by the facility and the distance of those demands from other facilities.

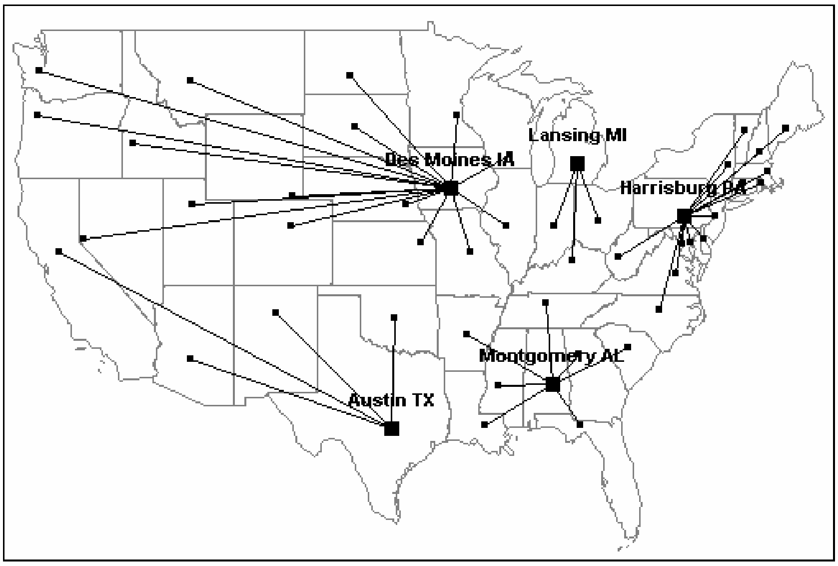

Fig. XX.2. UFLP solution for 49-node dataset, after failure of facility in Sacramento

Table XX.1. Failure costs and assigned demands for UFLP solution

Location	% Demand Served	Failure Cost	% Increase
Sacramento, CA	19	1,019,065	117
Harrisburg, PA	29	713,482	52
Montgomery, AL	17	634,473	35
Austin, TX	9	593,904	26
Des Moines, IA	16	546,599	16
Lansing, MI	12	537,347	14
Transportation cost w/o failures		470,228	0

A more reliable solution locates facilities in the capitals of Alabama, California, Iowa, New York, Ohio, Oregon, Pennsylvania, and Texas. The maximum failure cost occurs when the Austin, TX fails, but this cost is only $476,374, a mere 35 percent increase over the transportation cost of $352,698 when all 8 facilities are working. On the other hand, this solution also requires two additional facilities and is suboptimal for the UFLP. This solution is 7 percent more expensive according to the classical measure of cost (the UFLP cost) but is less expensive when failures are accounted for.

We argue that this latter measure (accounting for failures) is a more accurate measure of cost and that the second solution may be preferable to the first because of its superiority in this measure. Indeed, one of the key aims in this chapter is to demonstrate that large improvements in reliability can often be attained with only small increases in the classical cost.

Although we believe strongly that the "correct" objective functions in facility location problems should account for failures, we also believe strongly that it is important to examine the tradeoff between this objective and the classical ones—that is, the tradeoff between the cost if no disruptions occur and the cost if disruptions do occur. This tradeoff allows us to determine how significant a cost increase is required to add reliability to a system. For example, normal operating cost (sum of the fixed plus transportation costs) had to be twice as large as the optimal UFLP cost to attain a reasonable level of reliability, the additional cost may be unwarranted (unless facility failures are very likely). If, on the other hand, the tradeoff curve is "steep," then firms do not need huge investments in redundant infrastructure to improve the system's reliability. We believe that developing such tradeoff curves is an important step in convincing firms to change their optimization objectives to include disruptions.

Indeed, we generally find that the tradeoff curve is steep in this way. One explanation for this fortuitous finding is that, like many combinatorial optimization problems, facility location problems tend to have many near-optimal solutions. Some of these solutions may, by chance, have desirable properties like reliability. If we can find these solutions, we may find that the their attractive properties outweigh their slight suboptimality.

Of course, there are a number of possible ways to formulate objectives that consider disruptions. For example, one might try to minimize the expected failure cost (by weighting the failure costs in Table XX.1 by the probability of each facility's disruption), minimize the maximum failure cost (among all rows in Table XX.1), or find a solution whose cost stays within a given threshold with some probability.

In this chapter, we consider optimization models for the design of reliable facility location systems under a variety of risk measures and operating strategies, including those discussed in the previous paragraph and others. Our focus is on the formulation of these models and the insights that can be gained from comparing solutions obtained from different objectives. We briefly discuss algorithmic techniques for solving some of these models, but generally we refer to other sources for such discussions.

The remainder of this chapter is organized as follows. We present a brief literature review in Sect. 2. In Sect. 3, we introduce a base model that will be used as a foundation for the other models to follow. We discuss two ways to formulate this model, as well as a capacitated extension. In Sect. 4, we formulate several models using a range of risk measures. We summarize our findings and discuss opportunities for future research on Sect. 5.

XX.2 Literature Review

In this section, we present a brief overview of the literature on reliable supply chain network design problems. A more formal review of this body of literature is presented by Snyder et al. (2006). We refer the reader to the textbooks by Daskin (1995), Drezner, (1995), or Drezner and Hamacher (2002) for an introduction to facility location. Owen and Daskin (1998), Daskin, Snyder, and Berger (2005), and Snyder (2006) all provide reviews of stochastic location models (generally considering uncertainty in demand, rather than disruptions to facilities). See Birge and Louveaux (1997) or Higle (2005) for an introduction to general stochastic programming techniques.

Snyder and Daskin (2005) introduce several models, based on classical facility location problems, in which facilities may fail with a given probability. They minimize a weighted sum of two objectives, one of which is a classical objective (ignoring disruptions) and the other of which is the expected cost after accounting for disruptions. Customers are assigned to several facilities, one of which is the "primary" facility that serves it under normal circumstances, one of which serves it if the primary facility fails, and so on. One of their models is discussed below in Sect. XX.3.2. Snyder and Ülker (2005) present a capacitated version of their model (Sect. XX.3.1) and Jeon, Snyder, and Shen (2006) present a version that incorporates inventory costs into the location decision.

Berman, Krass, and Menezes (2005a) consider structural properties of a model that is less computationally tractable than Snyder and Daskin's but more general. A subsequent paper (Berman, Krass, and Menezes 2005b) assumes that customers do not know in advance which facilities are operational and must travel from facility to facility in search of a working site.

Church and Scaparra (2005) and Scaparra and Church (2005, 2006) consider the fortification, rather than design, of facilities—that is, the network is assumed to exist and the firm has resources to prevent disruptions at some of them, thus partially fortifying the network. Their model finds the best facilities to fortify assuming that an interdictor will attempt to cause worst-case losses for the firm by disrupting a fixed number of the un-fortified facilities. Similarly, Daskin et al. (2005) allow the firm to choose whether each facility opened is reliable or unreliable; reliable facilities come at a higher cost. (See Sect. XX.4.2 below).

Reliable facility location models are related to network reliability theory (Coburn 1987, Shier 1991, Shooman 2002), which attempts to calculate or maximize the probability that a network remains connected after random link failures. It is also related to the literature on facility location with congestion, in which facilities are sometimes unavailable due to excess demand (rather than to facility disruptions). (See Berman and LeBlanc (1984), Berman et al. (1985), Daskin (1982, 1983), Larson (1974), ReVelle and Hogan (1989).)

XX.3 Base Model

In this section, we present a base model that will be used as a foundation for most of the models to come. We formulate this base model in two ways. The first method uses scenarios to represent uncertain events and resembles the formulation of other stochastic facility location problems. This formulation is quite flexible and can be used to model the variations discussed throughout this chapter. However, the number of scenarios may be exponentially large: If there are N facilities and each can fail independently, there are 2^N failure scenarios. This type of formulation was used previously for a capacitated facility location problems with disruptions (Snyder and Ülker 2005). We present an uncapacitated version first, and then the capacitated version.

The second method captures the uncertain events implicitly, without explicit enumeration of all failure scenarios, and can be solved more efficiently than the scenario-based formulation. Unfortunately, it requires a restrictive assumption (that all facilities have the same probability of disruption) and cannot be extended with the same flexibility as the scenario-based formulation. This formulation was first introduced by Snyder and Daskin (2005a).

All of our models are based on the uncapacitated fixed-charge location problem (UFLP; Balinski 1965, Daskin 1995). We are given a set I of customer locations, each of which has an annual demand h_i for a single product. In addition, we have a set J of potential facility sites, each with an annual fixed operating cost f_j. If we choose to open facility j, then f_j is incurred at all times, regardless of whether the facility is operational. The cost to transport one unit of demand from facility j to customer i is denoted d_{ij}.

In the classical UFLP, there are two sets of decision variables, location variables and assignment variables. The location variables are denoted by X_j, which equals 1 if we open a facility at site j. The formulation of the assignment variables is different for different models below; we defer further discussion until we formulate those models.

Associated with each customer is a per-unit penalty cost θ_j that represents the cost of not serving the customer. This cost is incurred if all open facilities have failed, or if the facilities close to i (with respect to the transportation cost d_{ij}) have failed so that it is cheaper to pay the penalty than to serve the customer. θ_j may represent a lost-sales cost, or the cost to pay a competitor to serve the customer temporarily. Rather than modeling this cost explicitly, we add a dummy "emergency facility," denoted u, to the set J. Facility u is always open, has no fixed cost, and has a transportation cost of θ_j to customer i—that is, $X_u = 1$, $f_u = 0$, and $d_{iu} = \theta_j$ for all i. Moreover, facility u can never fail. Henceforth, we assume that the facility set J has been augmented in this way, and we ignore the penalty cost θ_j.

XX.3.1 Scenario-Based Formulation

Model

Let S be a set of scenarios, each of which specifies the failure state of all facilities in J. In particular, let A_s be the set of facilities that fails in scenario s. For convenience, we also define $a_{js} = 1$ if facility j fails in scenario s and 0 otherwise. Scenario s occurs with probability q_s. These scenarios may have been identified *a priori* by managers as likely possibilities that are worth planning against. Alternately, they may represent *all* possible combinations of facility failures. For example, if each facility j fails with probability p_j and failures are independent, then scenario s occurs with probability

$$q_s = \prod_{j \in A_s} p_j \prod_{j \in J \setminus A_s} (1 - p_j) . \tag{1}$$

We can modify these probabilities accordingly if failures are dependent. (Failures may be dependent because of geographic proximity, supplier commonality, etc.) To model the emergency facility, we require $a_{us} = 0$ for all s, or, equivalently, $q_s = 0$ if $a_{us} = 1$.

The scenario probability q_s is interpreted as the long-run fraction of time that the precise set of facilities A_s is disrupted. Put another way, the fraction of time in which facility j is disrupted is given by $p_s = \sum_{s \in S : j \in A_s} q_s$. In some cases, the q_s may be estimated from historical data, while in others it must be estimated subjectively. Our models are most easily interpreted as infinite-horizon models in which the facilities in A_s are disrupted for q_s fraction of the time. However, if the modeler has in mind a particular finite time horizon T, then q_s may be used to capture probabilistic information about the timing of the disruptions.

For example, suppose scenario s represents the situation in which exactly one facility, j, fails. Further, suppose that facility j will fail with probability 0.1, and if it does, it will fail in all periods from 1 through 5 with probability 0.3 and in all periods from 3 through T with probability 0.7. (Note that this means that if j fails at all, it will surely be non-operational during periods 3 through 5.) Then q_s is given by

$$q_s = \frac{0.9 \times 0 + 0.1 \times [0.3 \times 5 + 0.7 \times (T - 2)]}{T} . \tag{2}$$

For simplicity, we assume that scenarios specify only facility failures. However, it is simple to extend this formulation so that demands and transportation costs are also scenario dependent.

In each scenario, we need to assign customers to facilities. The decision variable for these doing so is given by Y_{ijs}, which equals the fraction of customer i's demand that is assigned to facility j in scenario s. As in the classical UFLP, single sourcing is optimal; that is, there exists an optimal solution for which $Y_{ijs} \in \{0,1\}$ for all i, j, and s.

We formulate our base model with the objective of minimizing the expected cost, though in future sections we will consider alternate risk measures. The scenario-based formulation of the reliability fixed-charge location problem (RFLP1) is formulated as follows:

$$\text{(RFLP1)} \quad \text{minimize} \sum_{j\in J} f_j X_j + \sum_{s\in S}\sum_{i\in I}\sum_{j\in J} q_s h_i d_{ij} Y_{ijs} \tag{3}$$

$$\text{subject to} \sum_{j\in J} Y_{ijs} = 1 \quad \forall i\in I, s\in S \tag{4}$$

$$Y_{ijs} \le (1-a_{js}) X_j \quad \forall i\in I, j\in J, s\in S \tag{5}$$

$$X_j \in \{0,1\} \quad \forall j\in J \tag{6}$$

$$Y_{ijs} \ge 0 \quad \forall i\in I, j\in J, s\in S \tag{7}$$

The objective function (3) minimizes the fixed cost plus the expected transportation cost across all scenarios. Constraints (4) require each customer to be assigned to some facility in every scenario. Constraints (5) prohibit a customer from being assigned to a facility that has not been opened, or to a facility that has failed in a given scenario. Constraints (6) require the location variables to be binary, and constraints (7) require the assignment variables to be non-negative (though, as stated above, an optimal solution always exists in which they are binary). Note that, although we do not explicitly require $X_u = 1$, any optimal solution will open the emergency facility if it is needed for some scenario since it has no fixed cost.

Note that, if there is a single scenario, and no facilities fail in this scenario, this model reduces to the classical UFLP. Since the UFLP is NP-hard (Garey and Johnson 1979), so is the RFLP.

(RFLP1) can be solved using standard IP solvers like CPLEX. However, if the scenarios represent all possible combinations of failures, then S is exponentially large. In this case, sampling techniques such as sample average approximation (SAA; ; Kleywegt, Shapiro and Homem-de-Mello 2001; Linderoth, Shapiro, and Wright 2002) may be used to solve the problem with a reduced set of scenarios and obtain statistical bounds on the quality of the solutions.

Capacitated Model

The formulation above assumes that facilities have infinite capacity or that they can serve any number of demands. In many cases, this might not be true. We can define k_{js} to be the capacity of a facility at candidate site j in scenario s. This notation and the following formulation, allow a facility to incur impaired capacity in a scenario without completely failing. We let the capacity of the dummy facility *u* be $k_{us} = \infty$ for all scenarios *s*, indicating that this facility can accommodate all demands if necessary in each scenario. With this notation, we replace constraint (5) by its more traditional version

$$Y_{ijs} \le X_j \qquad \forall i \in I, j \in J, s \in S \tag{8}$$

In addition, we add the following capacity constraint, where the demand placed on a facility's capacity is measured in terms of the demand units h_i

$$\sum_{i \in I} h_i Y_{ijs} \le k_{js} X_j \qquad \forall j \in J, s \in S \tag{9}$$

This formulation, denoted CRFLP, was first suggested by Snyder and Ülker (2005).

Two observations are worth making about the CRFLP. First, constraints (8) are implied by (9) and are therefore not technically needed. However, in most cases, the addition of (8) will strengthen any relaxation of the model. Hence, we suggest including constraints (8) explicitly in any model or algorithm. Second, constraints (9) allow demands at a node to be split between multiple facilities since the assignment variables can be fractional by constraints (7). However, the extent of multiple sourcing or fractional assignment of demands to facilities is bounded in each scenario. In particular, the maximum number of demand nodes that can be fractionally assigned to facilities is less than or equal to $\sum_{j \in J} X_j - 1$ in each scenario. Multiple sourcing may not be overly problematic, if this number is small relative to the total number of demand nodes, $|I|$. In such cases, an approximate solution to the single sourcing problem can often be found for each scenario using the approach suggested by Daskin and Jones (1993). When single sourcing is required and strict optimality is also needed, constraints (7) should be replaced by the obvious integrality constraints

$$Y_{ijs} \in \{0,1\} \qquad \forall i \in I, j \in J, s \in S \tag{10}$$

The imposition of these constraints is likely to increase the difficulty associated with solving the problem considerably.

XX.3.2 Implicit Formulation

Model

We next present a formulation of the RFLP in which the random disruptions are modeled implicitly, rather than using explicit scenarios. This formulation is based on the model presented by Snyder and Daskin (2005a). It requires us to make the (rather strong) assumption that the facilities are divided into two sets; the facilities in the first set never fail, while all of the facilities in the second set fail independently with the same probability, q. The first set is called NF (for "non-failable"), while the second is called F (for "failable"). Since the emergency facility never fails, we have $u \in NF$. Note that F and NF constitute a partition of J.

In the implicit formulation of the RFLP, denoted (RFLP2), assignments are made not based on scenarios but based on "assignment levels." In particular, an assignment of customer i to facility j is said to be a "level-r assignment" if there are r open, failable facilities that are closer to i than j is. If $r = 0$, then j is i's "primary" facility—the facility that serves it under normal circumstances—while if r>0, j is a "backup" facility. A given customer must be assigned to some facility at every level r from 0 to the number of open facilities, unless it is assigned to some non-failable facility at level $s < r$. We define $Y_{ijr} = 1$ if customer i is assigned to facility j as a level-r assignment.

Since each facility fails with the same probability, we can compute the probability that customer i is served by facility j knowing only the level of i's assignment to j—that is, knowing how many facilities are closer to i but not knowing which facilities those are. This allows a compact formulation of the expected cost. In particular, (RFLP2) is formulated as follows:

$$\text{(RFLP2)} \quad \text{minimize} \sum_{j \in J} f_j X_j + \sum_{i \in I} \sum_{r=0}^{|J|-1} h_i d_{ij} \left[\sum_{j \in NF} q^r Y_{ijr} + \sum_{j \in F} q^r (1-q) Y_{ijr} \right] \tag{11}$$

$$\text{subject to} \sum_{j \in J} Y_{ijr} + \sum_{j \in NF} \sum_{s=0}^{r-1} Y_{ijs} = 1 \quad \forall i \in I, r = 0, \ldots, |J|-1 \tag{12}$$

$$Y_{ijr} \le X_j \quad \forall i \in I, j \in J, r = 0, \ldots, |J|-1 \tag{13}$$

$$\sum_{r=0}^{|J|-1} Y_{ijr} \le 1 \quad \forall i \in I, j \in J \tag{14}$$

$$X_j \in \{0,1\} \quad \forall j \in J \tag{15}$$

$$Y_{ijr} \ge 0 \quad \forall i \in I, j \in J, r = 0, \ldots, |J|-1 \tag{16}$$

The objective function (11) minimizes the fixed cost plus the expected transportation cost. The transportation cost term reflects the fact that if customer i is assigned to facility j at level r, then it will be served by j if the r closer facilities fail (which happens with probability q^r) and if j itself does not fail (which happens with probability q if j is failable and with probability 1 if j is non-failable). Constraints (12) stipulate that each customer must be assigned to some facility at each level r, unless the facility is assigned to a non-failable facility at level $s < r$. (By convention, we take $\sum_{s=0}^{r-1} Y_{ijs} = 0$ if $r = 0$.) Constraints (13) prevent an assignment to a facility that has not been opened, while constraints (14) prevent a customer from being assigned to a given facility at more than one level. Constraints (15) and (16) require integrality and non-negativity of the location and assignment variables, respectively. As in the uncapacitated version of (RFLP1), this formulation has an optimal solution in which the assignment variables are binary even though we only require them to be non-negative. Also as in (RFLP1), there exists an optimal facility in which the emergency facility u is open even though we do not explicitly require it. Although assignment levels cannot exceed the number of open facilities, which is not known *a priori*, it is safe to extend the index r to $|J|-1$ in the formulation since each customer is assigned to *some* non-failable facility (possibly u) at some level less than $|J|-1$.

Once the location variables are fixed, it is optimal to assign a customer to its closest open facility at level 0, its second-closest at level 1, and so on, until it is assigned to some non-failable facility (possibly u).

Snyder and Daskin (2005a) propose a Lagrangian relaxation algorithm to solve (RFLP2). They relax constraints (12) to obtain a subproblem that can be solved efficiently to obtain a lower bound for a fixed set of Lagrange multipliers. Upper bounds are obtained by converting the X vector from the lower-bound solution into a feasible solution by assigning customers as described in the previous paragraph. The Lagrange multipliers are updated using subgradient optimization, and the algorithm can be embedded into a branch-and-bound procedure if the bounds produced are not sufficiently tight.

Tradeoff Curve

As discussed above, it is interesting to examine the tradeoff between the UFLP objective and the objective that accounts for failures. Snyder and Daskin (2005a) construct this tradeoff by formulating a multi-objective programming problem with two objectives based on (RFLP2):

$$w_1 = \sum_{j\in J} f_j X_j + \sum_{i\in I}\sum_{j\in J} h_i d_{ij} Y_{ij0} \tag{17}$$

$$w_2 = \sum_{i\in I}\sum_{r=0}^{|J|-1} h_i d_{ij}\left[\sum_{j\in NF} q^r Y_{ijr} + \sum_{j\in F} q^r (1-q) Y_{ijr}\right] \tag{18}$$

Objective w_1 is the classical UFLP objective, while objective w_2 is the objective function from (RFLP2) without the fixed-cost term. We replace the objective function in (RFLP2) with a weighted sum of these two objectives:

$$\text{minimize} \quad \alpha w_1 + (1-\alpha) w_2, \tag{19}$$

where $0 \leq \alpha \leq 1$. By solving the problem for varying values of α using the weighting method of multi-objective programming (Cohon 1978), we can generate a tradeoff curve consisting entirely of non-dominated solutions. (A solution is *non-dominated* if every other solution is worse than it in at least one of the two objectives.)

The resulting tradeoff curves for the 49-node data set described earlier are depicted in Fig. XX.3 for q = 0.01, 0.05, and 0.10. All facilities are assumed to be failable. The UFLP cost (w_1) is plotted on the x-axis and the failure cost (w_2) is plotted on the y-axis. Each point on a curve represents a different value of α and a different solution.

The solution that is optimal for the classical UFLP (found by solving (RFLP2) with $\alpha = 1$) is the left-most point on each curve. These points are equal on the horizontal axis (since they represent the same solution and hence have the same UFLP cost) but unequal on the vertical axis since they have different failure probabilities and hence different expected failure costs.

Fig. XX.3 suggests that as q decreases, the tradeoff curve shifts. That is, if the firm can somehow reduce the failure probability at its facilities, it can attain a higher level of reliability with the same UFLP cost—or, equivalently, it can attain the same level of reliability with a lower UFLP cost.

The steepness of the left part of each curve suggests that there are solutions that are much better than the UFLP solution in terms of reliability but not much worse in terms of cost. For example, consider the bottom curve, corresponding to q = 0.01. The third point from the left of this curve represents a solution that is 25% better than the UFLP solution in the reliability objective (w_2) but only 7% worse in the UFLP objective (w_1). Similarly, the fifth point is 38% better in w_2 but only 15% worse in w_1. These solutions are depicted in Figs. XX.4 and XX.5.

The number of facilities open in each solution tends to increase as we move rightwards in the curve, since more reliable solutions tend to have more facilities open. The right-most portion of the curve is quite flat, but this portion of the curve is not of much interest because nearly all of the facilities are open in these solutions; they are very reliable but excessively expensive.

We find tradeoff curves with this shape for a wide range of models and data sets, suggesting that large improvements in reliability can often be attained with only small increases in cost.

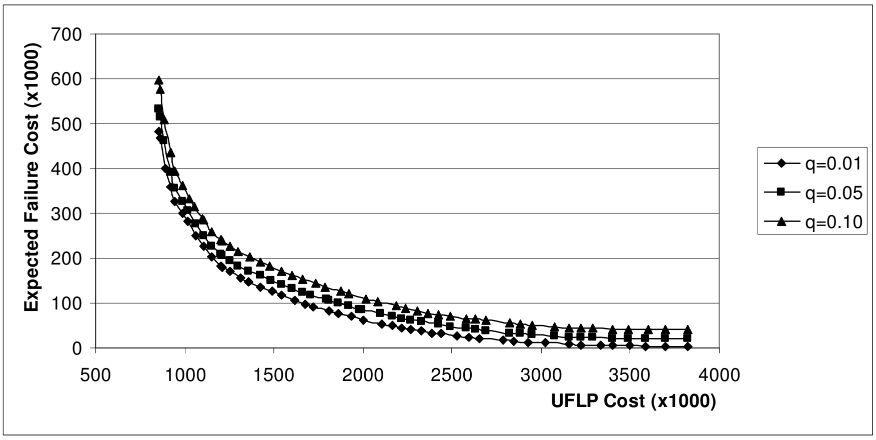

Fig. XX.3. Tradeoff curve for 49-node dataset

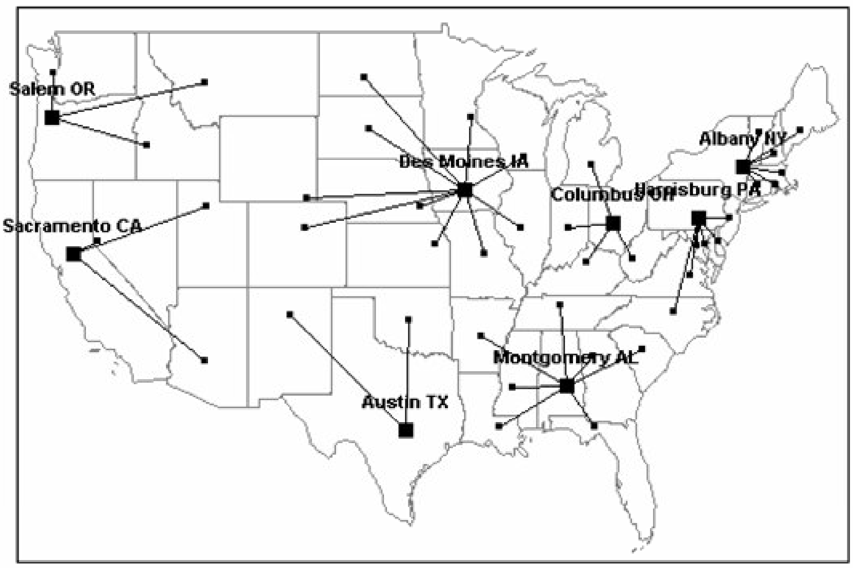

Fig. XX.4. Solution corresponding to third point on $q = 0.01$ tradeoff curve in Fig. 3.

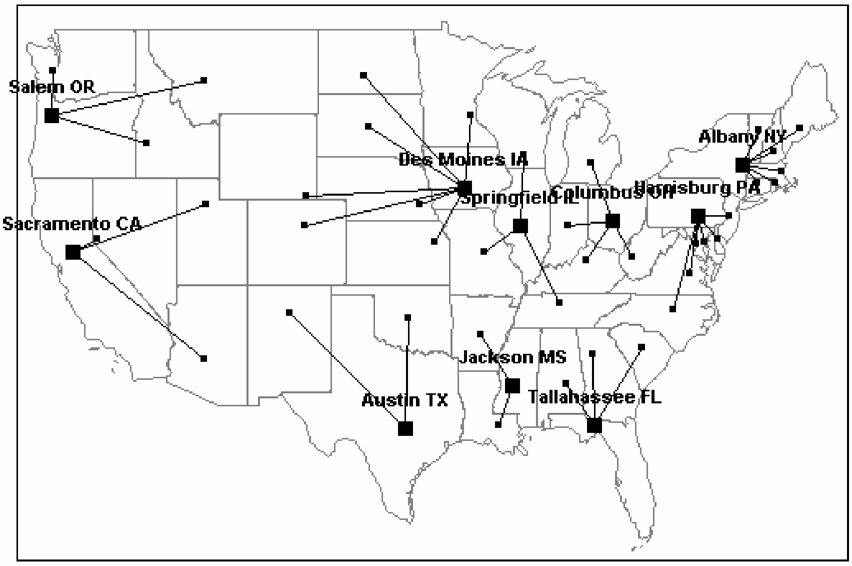

Fig. XX.5. Solution corresponding to fifth point on q = 0.01 tradeoff curve in Fig. XX.3.

XX.4 Alternate Operating Characteristics and Risk Measures

XX.4.1 Introduction

In this section, we outline a number of extensions to the base models defined above. We begin with a variant of the models that allows us to locate two different types of facilities: facilities that are perfectly reliable, or completely hardened against any and all attacks, and facilities that are subject to failure or that are unreliable in some way. The model will determine how many of each type of facility to locate and where they should be. In the second portion of this section, we explore alternative risk measures that also extend the formulations identified above.

XX.4.2 Reliable and Unreliable Facilities

In recent years, much attention has focused on the need to harden facilities against attacks. The attacks can be intentional, as in the case of terrorist attacks, or random or unintentional, as in the case of natural disasters. Scaparra and Church (2005a,b) outline defender/interdictor extensions to the traditional *P*-median problem in which *P* facilities *already exist* in a network. A defender can fortify q of these facilities against an attack by an interdictor against r of the remaining undefended facilities. The objective of the interdictor is to maximize the demand-weighted total distance with demands assigned to the closest non-interdicted facilities, while the defender attempts to minimize this worst-case cost by defending a subset of the facilities. Brown et al. (2005) provide an excellent tutorial on this class of defender/attacker problems.

We adopt a somewhat different approach, first suggested by Daskin (2005) and Daskin et al. (2006). First, we assume that facilities fail randomly. As such, we do not need to model the behavior of an interdictor whose objective is to maximize the damage that he or she inflicts on a network. Second, we do not assume that any facilities exist in the network; rather we formulate the model below based on *de novo* planning with no pre-existing facilities. The model can readily be adapted to the case in which some facilities already exist, through appropriate changes in the fixed costs.

One of two types of facilities can be established at each candidate site j. A reliable facility will never fail. Such a facility costs f_j^R at candidate site j. Alternatively, we may elect to construct an unreliable facility which can fail with probability q but which costs f_j^U. Clearly we require $f_j^U < f_j^R$ for there to be an incentive to locate any unreliable facilities. We define location decision variables X_j^R (and X_j^U) to be 1 if we locate a reliable (or unreliable) facility at candidate site j and 0 otherwise.

Similarly, every demand node i must be assigned to both a primary facility and a backup facility. The primary assignment will be used if the closest facility has not failed. The backup assignment will be to the closest reliable facility and will be used when the primary facility has failed. Thus, if the primary facility to which a demand node is assigned has failed, the demands at that node are served by the nearest reliable facility, not the nearest facility which has not failed. In this way, the model is a simplification of the base model outlined above. This assignment scheme is chosen primarily for computational reasons. However, during a disruption, real-time information is often limited, and it may be quite reasonable to assume that firms re-assign customers to their nearest reliable facility rather than trying to ascertain whether a closer unreliable facility is operational. We use decision variables Y_{ij}^P and Y_{ij}^B for the primary and backup assignments, respectively.

With this notation, the model becomes

$$\text{(RFLP3)} \quad \text{minimize} \tag{20}$$
$$\sum_{j\in J} f_j^U X_j^U + \sum_{j\in J} f_j^R X_j^R + (1-q)\sum_{i\in I}\sum_{j\in J} h_i d_{ij} Y_{ij}^P + q\sum_{i\in I}\sum_{j\in J} h_i d_{ij} Y_{ij}^B$$

$$\text{subject to} \quad \sum_{j \in J} Y_{ij}^P = 1 \qquad \forall i \in I \tag{21}$$

$$\sum_{j \in J} Y_{ij}^B = 1 \qquad \forall i \in I \tag{22}$$

$$Y_{ij}^P \le X_j^U + X_j^R \qquad \forall i \in I, j \in J \tag{23}$$

$$Y_{ij}^B \le X_j^R \qquad \forall i \in I, j \in J \tag{24}$$

$$X_j^R + X_j^U \le 1 \qquad \forall j \in J \tag{25}$$

$$\sum_{j \in J} X_j^R \ge 1 \tag{26}$$

$$X_j^R \in \{0,1\} \qquad \forall j \in J \tag{27}$$

$$X_j^U \in \{0,1\} \qquad \forall j \in J \tag{28}$$

$$Y_{ij}^P \ge 0 \qquad \forall i \in I, j \in J \tag{29}$$

$$Y_{ij}^B \ge 0 \qquad \forall i \in I, j \in J \tag{30}$$

The objective function (20) minimizes the total fixed cost for reliable and unreliable facilities as well as the transportation cost for primary and backup assignments. Primary assignments occur with probability $1-q$ for each demand node and backup assignments occur with probability q. If a customer's primary facility is reliable, then its backup assignment will be to the same facility, and the objective function computes the transportation cost to this facility with probability 1. Constraints (21) and (22) require that each demand node be assigned to a primary and backup facility. Constraints (23) state that the primary assignment can only be made to an open (reliable or unreliable) facility, while constraints (24) state that the backup assignment can only be to a reliable facility. Constraints (25) state that at any candidate site either a reliable or an unreliable facility can be located, but not both. Constraint (26) requires the model to locate at least one reliable facility. Constraints (27) and (28) are standard integrality constraints for the location variables, while constraints (29) and (30) are non-negativity constraints for the primary and backup assignment variables respectively.

Constraints (25) and (26) are not strictly needed. In the formulation as stated, there is no incentive to locate both a reliable and an unreliable facility at any candidate site; hence constraints (25) are not needed. Similarly, constraints (26) are implied by the need to provide a backup assignment to a reliable facility for every demand node (constraints 22 and 24). However, in many solution algorithms which relax one or more of the remaining constraints, these constraints are valuable additions as they tighten the relaxed formulation. For example, Daskin (2005) and Daskin et al. (2006) outline an extension of this model that allows the backup distance or cost to differ from the primary distance or cost even for the same demand node/facility pair. This extension requires the incorporation of additional decision variables, additional terms in the objective function and additional constraints to correct for the case in which a demand node is assigned to a reliable facility as both its primary and backup facility. (This correction is not needed when the primary and backup distances for each demand node/facility pair are the same as is the case in the formulation above.) They outline a Lagrangian solution approach that relaxes constraints (21) and (22) above. Constraints (25) and (26) significantly strengthen the bounds that result from this relaxation.

Table XX.2 shows the results associated with applying the model to the 49-node dataset. In these results, we increased the demand by a factor of 3 compared to the earlier results so that more facilities would be justified in the base case when no facilities are subject to failure. (All costs are in units of \$1000.) For all of these runs, the cost of a reliable facility was set to twice the cost of an unreliable facility at each candidate site.

Table XX.2. Results from RFLP3 Model for the 49-node dataset

Failure Prob	# Reliable	# Unrel.	Total Cost (x\$1,000)	Reliable Sites	Unreliable Sites
0.000	0	13	1,544		CA CO FL IA IL MI MS NY OH OR PA TX VA
0.010	1	12	1,643	PA	CA CO FL IA IL MI MS NY OH OR TX VA
0.030	2	11	1,742	IA PA	CA CO FL IL MI MS NY OH OR TX VA
0.050	3	10	1,805	MS OR PA	CA CO FL IA IL MI NY OH TX VA
0.100	3	9	1,910	IL OR PA	CA CO FL IA MI MS NY OH TX
0.150	4	7	1,992	IL MS OR PA	CA FL IA MI NY OH TX
0.200	4	6	2,046	CA IL MS PA	FL IA NY OH OR TX
0.250	5	4	2,079	AL CA IL PA TX	IA NY OH OR
0.300	5	4	2,107	AL CA IL PA TX	IA NY OH OR
0.350	5	4	2,135	AL CA IL PA TX	IA NY OH OR
0.360	6	3	2,139	AL CA IA OH PA TX	IL NY OR
0.400	6	2	2,153	AL CA IA OH PA TX	NY OR
0.450	6	2	2,168	AL CA IA OH PA TX	NY OR
0.475	6	1	2,174	AL CA IA OH PA TX	NY
0.500	6	0	2,177	AL CA IA OH PA TX	

Fig. XX.6 shows the solution when the facility failure probability is 0.05. Fig. XX.7 shows the results for a failure probability of 0.15, while Fig. XX.8 shows the results for a failure probability of 0.25. In all figures, the unreliable sites are shown in *italics.* Some demand nodes are shown with one assignment while others – those whose primary assignment is to an unreliable facility (dashed lines) – are shown with two assignments.

Several observations are worth noting. First, as the probability of a facility failing increases, the number of reliable facilities increases, the number of unreliable facilities decreases and the total cost increases. Second, for moderate values of the failure probability (under 0.05 in this case), the total number of sites does not change from the optimal number found when facilities are not subject to failure, but some facilities are hardened to insure that they do not fail. For larger failure probabilities, the total number of facilities decreases. Third, as the failure probability increases, some facilities will be eliminated completely (e.g., the facility at Richmond, VA which is eliminated once the failure probability gets to 0.10). Some facilities will be converted to reliable facilities as the failure probability increases (e.g., the facility at Harrisburg, PA, which becomes a reliable facility and remains a reliable facility for any failure probability). Other facilities change from unreliable, to reliable, back to unreliable and then back to reliable facilities again as the failure probability increases (e.g., the facility in Des Moines, IA, or the facility in Springfield, IL, which goes from an unreliable site, to a reliable facility and then back to an unreliable site). Finally, some facilities are introduced into the solution as the probability of failure increases (e.g., the facility at Montgomery, AL which enters the solution when the facility failure probability reaches 0.25).

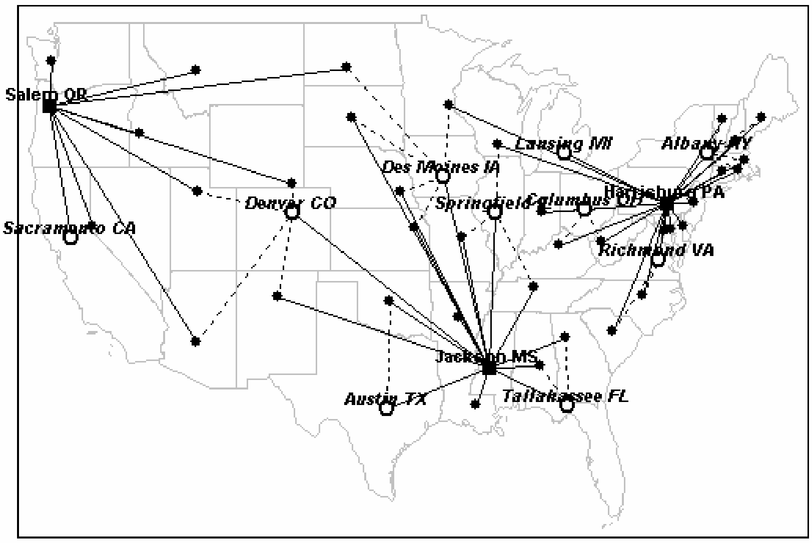

Fig. XX.6. Optimal locations of 3 reliable sites and 10 unreliable sites when failure probability is 0.05

In addition, as the failure probability increases, the expenditure on reliable facilities increases, while the contribution of the fixed facility costs for unreliable facilities decreases. Also, as the failure probability increases, the primary transportation cost increases (as there tend to be fewer facilities overall) but the backup transportation cost decreases (since the number of reliable sites increases with the failure probability). Finally, for failure probabilities exceeding 0.5 in this case, it is not cost-effective to utilize unreliable sites. In fact, an extension of a simple analytic model to incorporate both reliable and unreliable facilities indicates that, under the idealized assumptions of the analytic model (including equal reliable-facility costs of f^R across facilities, and similarly for f^U, and a uniform distribution of demand), unreliable facilities are not employed when the failure probability exceeds $\left(f^R - f^U\right)/f^R$ (Daskin, 2005; Daskin et al., 2006). While the discrete model whose results are shown above does not require all facility sites to cost the same amount of money, at any candidate site a reliable facility will be twice the cost of an unreliable facility. Thus, loosely speaking, the ratio above will be 0.5 even for the discrete results. As shown in Tab. 1, when the failure probability exceeds 0.5, no unreliable facilities are used.

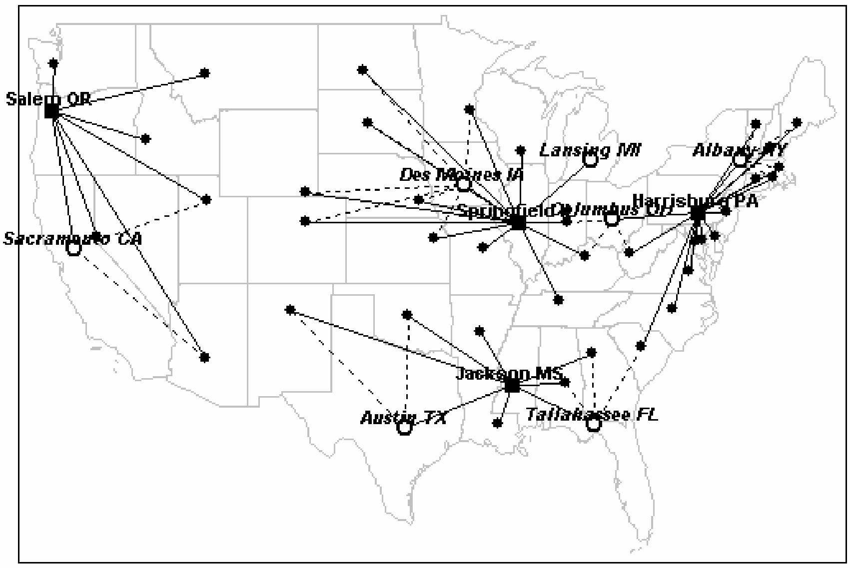

Fig. XX.7. Optimal locations of 4 reliable sites and 7 unreliable sites when failure probability is 0.15

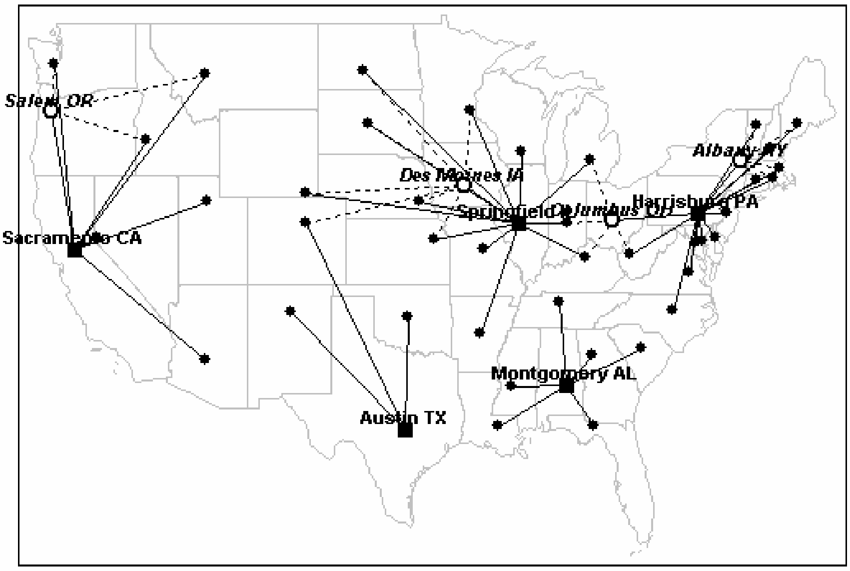

Fig. XX.8. Optimal locations of 5 reliable sites and 4 unreliable sites when failure probability is 0.25

XX.4.3 Other Risk Measures

The risk measures discussed so far focus on the average performance of the system when facilities fail. Such models assume that decision makers are risk neutral. In many contexts, decision makers are risk averse: they are concerned not only with the expected performance, but with the potential deviation from it. This may be particularly true when managers are faced with the prospects of losing facilities to natural or man-made disasters. Therefore, in this sub-section, we briefly formulate a number of extensions to the base model that allow decision makers to explore alternate risk measures. In general, these risk measures have all appeared in the literature on facility location under demand uncertainty but have not previously been used for disruption problems.

Minimax Cost Model

The first extension to the base model entails minimizing the worst-case cost in the event of a failure. To do so, we define a new decision variable, U, which is equal to the worst-case fixed plus transportation cost over all scenarios. Objective (31) below minimizes this cost subject to constraint (32), which defines the cost in terms of the total fixed plus demand-weighted transportation cost in each scenario.

$$\text{minimize } U \tag{31}$$

$$\text{subject to} \quad \sum_{j\in J} f_j X_j + \sum_{i\in I} \sum_{j\in J} h_i d_{ij} Y_{ijs} \le U \quad \forall s \in S \tag{32}$$

$$(4) – (7)$$

This formulation has the advantage of not requiring scenario probabilities as inputs. However, while the expected cost measure defined in (3) is risk neutral, the minimax objective of (31) is extremely risk averse. In fact, the location plan is frequently defined by one (possibly low-probability) scenario, as is often the case in minimax objectives (including the *P*-center model, for example). Such a strong aversion to the worst case often leads to solutions that are quite costly in the non-worst cases. As such, the minimax approach, which places undue emphasis on the worst case, is difficult to justify, just as is the expected value objective of (3), which allows very bad worst-case results. Additional approaches are outlined below.

Mean-Variance

One of the first and most famous objectives considered for optimization under uncertainty is the mean-variance approach. In this model, we minimize a weighted sum of mean cost and the variance of the cost. To define this model, let $z_s(Y)$ be the transportation cost in scenario s if the allocation variables are given by Y. Then the mean-variance model may be formulated as follows

$$\text{minimize} \quad \sum_{j\in J} f_j X_j + \sum_{s\in S} q_s z_s(Y) + \lambda \left[\sum_{s\in S} q_s \left(z_s(Y)\right)^2 - \left(\sum_{s\in S} q_s z_s(Y) \right)^2 \right] \tag{33}$$

$$\text{subject to} \quad (4)\text{-}(7)$$

where λ is a weight that is placed on the variance of the transportation costs. The variance places a higher implicit penalty on transportation costs that are significantly larger (and smaller) than the average.

The key problem with this model is that the objective function is highly nonlinear. Also, equally penalizing transportation costs that are lower than the average and higher than the average seems somewhat illogical as decision makers are most likely to be concerned with costs that exceed the mean.

Bounding the Cost

One approach to balancing the average cost and the worst-case cost is to minimize one cost while bounding the other. For example, we can minimize the expected cost over all scenarios – objective (3) – while bounding the cost in each scenario. This formulation is shown below.

$$\text{minimize} \quad \sum_{j\in J} f_j X_j + \sum_{s\in S}\sum_{i\in I}\sum_{j\in J} q_s h_i d_{ij} Y_{ijs} \tag{3}$$

$$\text{subject to} \quad \sum_{j\in J} f_j X_j + \sum_{i\in I}\sum_{j\in J} h_i d_{ij} Y_{ijs} \le r \quad \forall s \in S \tag{34}$$

(4) – (7)

Constraint (34) limits the cost in each scenario, including the fixed facility costs which are common across all scenarios, to a value r. Alternatively, we can simply minimize the uncapacitated fixed charge location problem (UFLP) objective subject to (34) as well as (4)-(7). The UFLP objective is simply:

$$\text{minimize} \quad \sum_{j\in J} f_j X_j + \sum_{i\in I}\sum_{j\in J} h_i d_{ij} Y_{ij} \tag{35}$$

This is equivalent to minimizing the cost in the scenario in which no facilities fail subject to a constraint on the costs incurred when facilities do fail. This approach was proposed by Snyder (2003).

One problem with this approach is that the costs incurred when facilities fail may differ significantly from one scenario to another. Thus, it may make more sense to constrain the costs in scenario s relative to the best we could do in scenario s, had we known that scenario s would occur, rather than relative to some absolute limit r. To do so, we define z_s to be the optimal objective function value in scenario s. We can then modify (34) to constrain the total cost in scenario s to be (1+r) times the optimal cost in scenario s as shown in constraint (36).

$$\sum_{j\in J} f_j X_j + \sum_{i\in I}\sum_{j\in J} h_i d_{ij} Y_{ijs} \le (1+r) z_s \quad \forall s \in S \tag{36}$$

Let us define $R_s = \sum_{j\in J} f_j X_j + \sum_{i\in I}\sum_{j\in J} h_i d_{ij} Y_{ijs} - z_s$. R_s is the *absolute regret* in scenario s: the absolute difference in total cost between the best we can do in scenario s and the best we could have done in scenario s had we known that scenario s would occur. Similarly, R_s / z_s is the *relative regret*, which represents the percentage difference.

Effectively, (36) constrains the relative regret in each scenario to be no more than r. This approach is similar to the "stochastic p-robust optimization" approach introduced by Snyder and Daskin (2005b), which minimizes the expected cost in a facility location problem with uncertain demands and costs, subject to a constraint requiring the regret in any scenario to be no more than p. Snyder and Daskin argue that stochastic p-robustness combines the attractive elements of the min-expected-cost and minimax-cost approaches by optimizing the expected performance while ensuring adequate performance in every scenario. They show that large improvements in robustness (i.e., decreases in worst-case cost) are possible with only small increases in expected cost.

A similar phenomenon is evident in the results of the model in which we minimize the expected cost (3) subject to (36) and (4)–(7). Table XX.4 reports the solutions of this model for various values of r for the 49-node data set. For computational reasons, these tests only include scenarios in which zero or one facilities fail. The first column lists r, the maximum allowable relative regret. The second column gives the expected cost of the resulting solution (×1000), while the third gives the maximum relative regret of this solution (which must be no greater than r). The fourth column lists the states in which facilities are opened in the solution.

Notice that substantial reductions in regret are possible with only minor increases in expected cost. For example, the second solution has a maximum regret that is 29% smaller than the baseline solution ($r = \infty$) but has only 4% greater expected cost. Similarly, the last solution ($r = 0.25$) has 68% smaller maximum regret but only 6% greater expected cost.

The last row of Table XX.4 corresponds to the optimal solution for the scenario in which the PA facility fails. This scenario is the one that attains the maximum regret for all values of r except ∞. As r decreases, this is the critical scenario, and the solution adjusts to reduce the regret in it. When $r = 0.25$ (corresponding to 25% regret), the solution is quite similar to the optimal solution for that scenario: the two solutions have four facilities in common, two neighboring pairs of facilities (OH / MI and PA / NJ), and only one outlier facility. If we reduce r below 0.209, a second scenario becomes critical, and it is impossible to reduce the regret of both scenarios simultaneously; therefore, the problem becomes infeasible.

This last point highlights one of the main difficulties with models that bound the cost in each scenario. In the other models we have discussed, it is trivial to find a feasible solution. In contrast, as r decreases, it can become quite difficult to find a solution that is feasible with respect to (36). In fact, Snyder and Daskin (2005b) prove that, if the number of scenarios is at least 2, then determining whether a given problem instance is *feasible* is NP-complete. Their result applies to a problem with uncertain demands and costs, but a similar result can be proven for problems with facility failures.

Table XX.4. Solutions to problems with bounded costs

R	Exp. Cost (×1000)	Max Regret	Locations
∞	737	0.649	AL IL NV PA TX
0.5	768	0.462	AL CA IL OR PA TX
0.4	774	0.341	AL CA IN OR PA TX
0.3	776	0.274	AL CA IA MI OR PA TX
0.25	782	0.209	AL CA IA OH OR PA TX
[PA fails]	—	—	AL CA IA MI NJ TX

α-Reliability

In many personal, private sector and public-sector decision contexts, it makes sense to plan not just for the average or expected-value case, or for the worst case, but rather for some eventuality in between these extremes. For example, the expected outcome of the more than 2 million cosmetic surgical procedures performed in the U.S. in 2004 (ASAPS 2004) was an improvement in the patient's appearance. The worst-case outcome undoubtedly was death in a small percentage of the cases. While most people who have such elective surgery expect the best possible outcome, it is prudent to plan for adverse results as well. For example, it is wise to have an up-to-date will as well as a living will and health care proxy before undergoing any surgical procedure. Similarly, in the design of public facilities such as airports, we clearly do not plan just for the average volume, but we also do not size airports for the peak demands associated with the Thanksgiving weekend. Airport capacity is based on values that are intermediate between the average and maximum daily demand levels.

In a similar manner, we can plan against an endogenously determined subset of the scenarios whose combined probability is at least α. One variant of this approach would minimize the maximum regret over all such scenarios, ignoring the regret in the remaining scenarios. To do this, we define a new variable W_s to be 1 if scenario s is in the "reliability set" against which we are planning and 0 otherwise. (Note that the term *reliability* is used in a different context in this model and refers to an endogenously determined set of scenarios against which the model is planning.) We also define R to be the maximum regret over all scenarios in the reliability set. Finally, we let M be a large number, larger than any possible scenario regret. With this notation, the problem can be formulated as:

$$\text{minimize } R \tag{39}$$

$$\text{subject to} \quad \sum_{s \in S} q_s W_s \geq \alpha \tag{40}$$

$$\sum_{j \in J} f_j X_j + \sum_{i \in I} \sum_{j \in J} h_i d_{ij} Y_{ijs} - z_s - M(1 - W_s) \leq R \quad \forall s \in S \tag{41}$$

(4)-(7)

The objective function (39) minimizes the maximum regret R over the scenarios in the reliability set. Constraint (40) requires the reliability set over which the minimization is performed to have a probability of at least α. Constraint (41) defines the maximum regret in terms of the scenario-based regrets, but excludes scenarios that are not part of the reliability set. This model was first proposed by Daskin et al. (1997), though scenarios in the original model referred to uncertainty in demand rather than in supply.

One problem with the α-reliable minimax regret model above is that it ignores the regret associated with scenarios that are not part of the reliability set. Chen et al. (2005) have proposed the α-reliable mean excess regret model, which, when applied to the problems at hand results in the following formulation:

$$\text{minimize } \zeta + \frac{1}{1-\alpha} \sum_{s \in S} q_s U_s \tag{42}$$

$$\text{subject to } U_s \geq R_s - \zeta \quad \forall s \in S \tag{43}$$

$$\sum_{j \in J} f_j X_j + \sum_{i \in I} \sum_{j \in J} h_i d_{ij} Y_{ijs} - z_s \leq R_s \quad \forall s \in S \tag{44}$$

(4)-(7)

In this model, we can think of ζ as the regret contribution of every scenario. A fraction of the scenarios (approximately equal to α) will have regret values that exceed this endogenously determined value. The objective function (42) minimizes the sum of ζ and the expected regret in excess of this value. Constraint (43) defines the excess regret in scenario s as the amount by which the regret in scenario s exceeds the nominal value ζ. Constraint (44) defines the regret in scenario s in terms of the compromise locations, the scenario-specific demand assignments and the optimal objective function for scenario s, z_s.

Although the α-reliable minimax regret and α-reliable mean excess regret models were originally formulated for problems with demand uncertainty, they can be applied equally well to problems with facility failures.

Chance-Constrained Approach

Finally, we note that the α-reliable minimax regret model is similar to a chance-constrained model. For example, we can minimize the expected cost over all scenarios – objective (3) – subject to a constraint that the sum of the probabilities associated with scenarios in which the cost exceeds some value C_{target} is less than or equal to β, a user-specified value. Let us define the decision variable T_s to be 1 if the cost in scenario s exceeds C_{target} and 0 otherwise. A chance-constrained model can now be formulated as follows:

$$\text{minimize } \sum_{j \in J} f_j X_j + \sum_{s \in S} \sum_{i \in I} \sum_{j \in J} q_s h_i d_{ij} Y_{ijs} \tag{3}$$

$$\text{subject to } \sum_{s \in S} q_s T_s \leq \beta \tag{45}$$

$$\sum_{j \in J} f_j X_j + \sum_{i \in I} \sum_{j \in J} h_i d_{ij} Y_{ijs} - C_{\text{target}} \leq M T_s \quad \forall s \in S \tag{46}$$

(4)-(7)

The objective minimizes the expected cost over all scenarios. Constraint (45) states that the sum of the probabilities of all scenarios with cost greater than C_{target} must be less than or equal to β. Constraint (46) links the location and allocation variables which define the cost in scenario *s* to the variables T_s. If the scenario-specific cost exceeds C_{target}, meaning that the left-hand side of (46) is positive, then T_s must be 1; otherwise T_s may be 0. Again, *M* is a sufficiently large value so that constraint (46) will not be binding whenever $T_s = 1$.

XX.5 Conclusions

Supply chain planners face a significant amount of uncertainty, particularly during the strategic planning phase. Facility location decisions are very expensive to change, so planners must take uncertainty into account when choose facility locations. In this chapter, we have illustrated the broad range of strategies that decision makers might take for approaching risk in facility location models with supply disruptions. A planner may choose one or more of these approaches based on his or her level of risk aversion, the type of disruptions that are of greatest concern, the flexibility of each measure to fine-tune parameters and add side constraints, the computational difficulty with which each model can be solved, and other factors.

One key insight that comes from many of the models we have discussed is that it is often relatively inexpensive to "buy" reliability—that is, if decision makers are willing to sacrifice just a bit in the objectives they are used to considering, they can gain significant improvements in other objectives, including reliability.

The models discussed in this chapter by no means represent an endpoint for research on facility location with disruptions. Several important issues remain to be addressed. One is computational: Many of these models are simply too difficult to solve, for reasonably sized instances, using off-the-shelf IP solvers. Rather, special-purpose algorithms, such as those proposed by Snyder and Daskin (2005a,b) and others, must be developed to solve these problems.

Another important direction for future research involves capturing other types of supply chain decisions in a unified model. A number of models attempt to incorporate tactical decisions, such as inventory and vehicle routing, into the facility location decision. These models tend to offer a substantial improvement over a sequential optimization approach in which facility locations are chosen first, and then tactical decisions are made while keeping the strategic decisions fixed. A natural next step is to consider facility failures in these models. For example, Jeon, Snyder, and Shen (2006) consider facility failures in the context of the joint location-inventory model first proposed by Daskin, Coullard, and Shen (2002) and Shen, Coullard, and Daskin (2003).

A third avenue for future research involves multi-echelon facility location and network design problems with disruptions. Such models might be based on the seminal distribution network design problem of Geoffrion and Graves (1974). In the multi-echelon case, a key question is how to model the "cascading" effect of disruptions, as failures at one echelon lead to failures downstream, either explicitly (because of geographical proximity of the facilities, for example) or implicitly (as downstream facilities become starved for raw materials during a disruption). We hope that this chapter will help to spark future research on these and other related topics.

References

ASAPS (American Society for Aesthetic Plastic Surgery) (2004), Cosmetic Surgery Quick Facts: Highlights of the ASAPS 2004 Statistics on Cosmetic Surgery, http://www.cosmeticplasticsurgerystatistics.com/statistics.html, accessed March 2006

Balinski ML (1965) Integer programming: Methods, uses, computation. Management Science 12:253-313

Barrionuevo A, Deutsch CH (2005) A distribution system brought to its knees. New York Times, Sep. 1, 2005, p. C1

Berman O, Krass D (2002) Facility location problems with stochastic demands and congestion. In: Drezner Z, Hamacher HW (eds) Facility location: Applications and theory. Springer-Verlag, New York, chapter 11.

Berman O, Krass D, Menezes MBC (2005a) Facility reliability issues in network p-median problems: Strategic centralization and co-location effects. Forthcoming in Operations Research

Berman O, Krass D, Menezes MBC (2005b) MiniSum with imperfect information: Trading off quantity for reliability of locations. Working paper, Rotman School of Management, University of Toronto, Toronto, Canada

Berman O, LeBlanc B (1984) Location-relocation of mobile facilities on a stochastic network. Transportation Science 18:315-330

Berman O, Larson RC, Chiu SS (1985) Optimal server location on a network operating as an M/G/1 queue. Operations Research 33:746-771

Birge JR, Louveaux F (1997) Introduction to stochastic programming. Springer-Verlag, New York.

Brown GG, Carlyle WM, Salmerón J, Wood K (2005) Analyzing the vulnerability of critical infrastructure to attach and planning defenses. In: Greenberg HJ (ed) TutORials in operations research. INFORMS, Baltimore, pp. 102-123

Chen G, Daskin MS, Shen ZJ, Uryasev S (2003) The α-reliable mean-excess regret model for stochastic facility location. Working paper, University of Florida, Department of Industrial and Systems Engineering

Church RL, Scaparra MP (2005) Protecting critical assets: The *r*-interdiction median problem with fortification. Working paper 79, Kent Business School, Canterbury, England

Cohon JL (1978) Multiobjective programming and planning. Academic Press, New York

Colbourn C (1987) The combinatorics of network reliability. Oxford University Press, New York

Daskin MS (1982) Application of an expected covering model to EMS system design. Decision Sciences 13:416-439

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

Daskin MS (1995) Network and discrete location: Models, algorithms, and applications. Wiley, New York
Daskin MS, Chopra SL, Lim M (2005) To harden or not to harden. Presentation, University of Kent Business School, Canterbury, England, October 2005
Daskin MS, Chopra SL, Lim M (2006) Integrated models of reliable and unreliable facility location problems. Working paper, Northwestern University, Evanston, IL, Department of Industrial Engineering and Management Sciences
Daskin MS, Coullard CR, Shen ZJM (2002) An inventory-location model: Formulation, solution algorithm and computational results. Annals of Operations Research 110:83-106
Daskin MS, Hesse SM, ReVelle CS (1997) α-reliable *p*-minimax regret: A new model for strategic facility location modeling. Location Science 5:227-246
Daskin MS, Hogan K, ReVelle C (1988) Integration of multiple, excess, backup, and expected covering models. Environment and Planning B 15:15-35
Daskin MS, Jones PC (1993) A new approach to solving applied location/allocation problems. Microcomputers in Civil Engineering 8:409-421
Daskin MS, Snyder LV, Berger RT (2005) Facility location in supply chain design. In: Langevin A, Riopel D (eds) Logistics systems: Design and operation. Springer, New York, pp. 39-66
Drezner Z (ed) (1995) Facility location: A survey of applications and methods. Springer-Verlag, New York
Drezner Z, Hamacher HW (eds) (2002) Facility location: Applications and theory. Springer-Verlag, New York
Fox J (2005) A meditation on risk. Fortune 152:50-62
Garey MR, Johnson DS (1979) Computers and intractability: A guide to the theory of NP-completeness. W. H. Freeman and Company, New York
Geoffrion AM, Graves GW (1974) Multicommodity distribution system design by Benders decomposition. Management Science 20:822-844
Greenhouse S (2002) Both sides see gains in deal to end port labor dispute. New York Times, Nov. 25, 2002, p. A14
Higle JL (2005) Stochastic programming: Optimization when uncertainty matters. In: Greenberg HJ (ed) TutORials in operations research. INFORMS, Baltimore, pp. 30-53
Jeon HM, Snyder LV, Shen ZJM (2006) Joint location-inventory optimization with unreliable facilities. Working paper, Lehigh University, Department of Industrial and Systems Engineering
Kleywegt AJ, Shapiro A, Homem-de Mello T (2001) The sample average approximation method for stochastic discrete optimization. SIAM Journal on Optimization 12:479-502
Larson RC (1974) A hypercube queueing model for facility location and redistricting in urban emergency services. Computers and Operations Research 1:67-95
Leonard D (2005) “The only lifeline was the Wal-Mart”. Fortune 152:74-80
Linderoth J, Shaprio A, Wright S (2002) The empirical behavior of sampling methods fo stochastic programming. Working paper, University of Wisconsin-Madison, Computer Sciences Department
Marquis C, McNeil DG (2001) Meat from Europe is banned by U.S. as illness spreads. New York Times, Mar. 14, 2001, p. A1
Mouawad J (2005) Katrina’s shock to the system. New York Times, Sep. 4, 2005, p. 3.1
Owen SH, Daskin MS (1998) Strategic facility location: A review. European Journal of Operational Research 111:423-447
Pollack A (2004) U.S. will miss half its supply of flu vaccine. New York Times, Oct. 6, 2004
Reuters (2005) Lumber, coffee prices soar in Katrina's wake. Reuters, Sep. 1, 2005

ReVelle C, Hogan K (1989) The maximum availability location problem. Transportation Science 23:192-200

Scaparra MP, Church RL (2005) An optimal approach for the interdiction median problem with fortification. Working paper #78, Kent Business School, Canterbury, England

Scaparra MP, Church RL (2006) A bilevel mixed integer program for critical infrastructure protection planning. Working paper, Kent Business School, Canterbury, England

Shen ZJM, Coullard CR, Daskin MS (2003) A joint location-inventory model. Transportation Science 37:40-55

Shier D (1991) Network reliability and algebraic structures. Clarendon Press, Oxford

Shooman ML (2002) Reliability of computer systems and networks: Fault tolerance, analysis, and design. John Wiley & Sons, New York

Snyder LV (2003) Supply chain robustness and reliability: Models and algorithms. PhD thesis, Northwestern University, Evanston, IL, Dept. of Industrial Engineering and Management Sciences

Snyder LV (2006) Facility location under uncertainty: A review. IIE Transactions 38:537-554

Snyder LV, Daskin MS (2005a) Reliability models for facility location: The expected failure cost case. Transportation Science 39:400-416

Snyder LV, Daskin MS (2005b) Stochastic *p*-robust location problems. Forthcoming in IIE Transactions

Snyder LV, Scaparra MP, Daskin ML, Church RC (2006) Planning for disruptions in supply chain networks. In: Greenberg HJ (ed) TutORials in operations research. INFORMS, Baltimore, forthcoming

Snyder LV, Ülker NŞ (2005) A model for locating capacitated, unreliable facilities. IERC Conference, Atlanta, GA, May 2005