USDOT Region V Regional University Transportation Center Final Report

NEXTRANS Project No. 074PIY04

A Decision Support Tool to Locate Shelters in Emergency Logistics

By

Satish V. Ukkusuri, Ph.D.

Professor, Lyles School of Civil Engineering

Purdue University

sukkusur@purdue.edu

Yanfeng Ouyang, Ph.D., M.ASCE

Associate Professor

Department of Civil and Environmental Engineering

University of Illinois at Urbana-Champaign

yfouyang@illinois.edu

DISCLAIMER

Funding for this research was provided by the NEXTRANS Center, Purdue University under Grant No. DTRT07-G-005 of the U.S. Department of Transportation, Research and Innovative Technology Administration (RITA), University Transportation Centers Program. The contents of this report reflect the views of the authors, who are responsible for the facts and the accuracy of the information presented herein. This document is disseminated under the sponsorship of the Department of Transportation, University Transportation Centers Program, in the interest of information exchange. The U.S. Government assumes no liability for the contents or use thereof.

ACKNOWLEDGMENTS

The authors acknowledge the assistance and feedback from the members of the study advisory committee.

TABLE OF CONTENTS

CHAPTER 1. INTRODUCTION

1.1 Study objectives

The objective of this research is to develop a systematic methodology to locate shelters considering both transportation and social factors in the aftermath of disasters. When anticipated demands for hurricane evacuation shelter spaces exceed existing capacity as defined by the preceding standards, there is a need to utilize less preferred facilities. It is critical that shelter selection decisions be made carefully considering both accessibility and facility conditions, and in consultation with local emergency management and public safety officials. While Red Cross and other relief agencies propose strategies to locate shelters, they currently do not consider how evacuees choose facilities based on accessibility to shelters and the in-facility congestion. This was evident in recent disasters such as Hurricane Katrina and Rita where some of the smaller shelters turned out to be inaccessible or unsafe for evacuees to use. The two chapters that follow develop mathematical frameworks to locate shelters in disasters using network optimization techniques and develop new solution approaches to solve them.

1.2 Organization of the research

The remainder of the research is organized as follows. The next chapter provides a mathematical model to consider traveler routing and in facility delays in the location of facilities. Chapter 3 considers the use of social cost functions to optimize the location of facilities in a network. Chapter 4 provides conclusions and directions for future research.

CHAPTER 2. RELIABLE FACILITY LOCATION DESIGN UNDER SERVICE DISRUPTION, EN-ROUTE CONGESTION AND IN-FACILITY QUEUING

This chapter studies the planning of service facility locations under consideration of customers' en-route travel and their in-facility delay, as well as the reliability of service facilities against natural or man-made hazards.

2.1 Introduction

Frequent natural and man-made catastrophic events highlight the need for a reliable and responsive service network to mitigate the adverse impacts and meet non-routine service demand. One of the key components of such designs is to select locations for service facilities (e.g., shelters, emergency medical centers, etc.), which should be properly made to accommodate traffic demand meanwhile mitigating en-route traffic congestion (Bai et al., 2011). Furthermore, when demand exceeds the capacity of service facilities, in-facility congestion inevitably induces significant waiting delays, human suffering and enormous social cost. For instance, in the 2010 Haiti earthquake, thousands of people flocked to a few point-of-distribution centers where food and water became insufficient (Jaller and Holguín-Veras, 2013). Due to the unexpected disasters, facility service disruption often occurs, and customers may sometimes be reassigned to more distant facilities that increases system operational costs and worsens both en-route traffic congestion and in-facility delay. As such, strategic design of service facility locations should not only ensure customer service with minimal en-route and in-facility congestion under normal scenarios, but also reduce human sufferings and negative social-economic consequences under disruption scenarios.

In recent years there is a large body of literature on facility location problems with in-facility congestion, most of which incorporated queuing theory models, e.g., M/M/1 (Zhang et al., 2009), M/G/1 (Berman et al., 1985), and M/M/n (Larson, 1974, 1975; Zhang et al., 2010). Earlier models attempted to capture the capability that customer demand can be covered by the facility. They achieve a certain level of service by either providing redundant coverage (Daskin, 1982; 1983, ReVelle and Hogan, 1989; Ball and Lin, 1993) or explicitly considering the queuing aspect of the problem (Larson 1974, 1975; Berman et al., 1985; Marianov and ReVelle, 1996).

Recent studies further integrated the impact of in-facility congestion in location-allocation problems to determine the optimal number and location of facilities, their service capacity, and the assignment of customers (Marianov and Serra, 2002; Aboolian et al., 2008; Syam, 2008; Castillo et al., 2009; Corrêa et al., 2009; Aboolian et al., 2012). These models were commonly used for the design of service networks in normal operational contexts (e.g., healthcare, banking, or ticket services), and emergency services in disaster contexts (e.g., ambulances, fire stations, shelters and point-of-distribution). Zhang et al. (2009) formulated a preventive healthcare facility location problem and incorporated customers' in-facility congestion as an M/M/1 queue in a nonlinear optimization framework. Later, the same authors extended their model to an M/M/n queue and proposed a bi-level non-linear mathematical program under user equilibrium to improve accessibility of preventive healthcare centers (Zhang et al., 2010). Considering walking, waiting and security risk effect, Jaller and Holguín-Veras (2013) proposed an approximate optimization formulation to minimize the total social cost of human suffering in the configuration of PODs network. Minimizing queuing effects was also frequently considered in competitive facility location problems (Marianov et al., 2008; Zhang and Rushton, 2008). Kwasnica and Stavrulaki (2008) explored competitive facility location and capacity decisions by considering queuing delay in a two-stage game model and conducted a comparative analysis on system equilibrium under three different monopoly conditions. In general, these models have a similar assumption that transportation cost can be estimated by simple shortest travel distances and hence no traffic congestion exists.

The impact of traffic congestion on customer access cost and facility location design is gaining more attention in recent years. For example, Bai et al. (2011) and Hajibabai and Ouyang (2013) examined traffic congestion impact and incorporated shipment routing decisions endogenously in supply chain design problems. Hajibabai et al. (2013) later extended this idea to incorporating the impact of freight traffic on deterioration of highway pavement infrastructure. Traffic routing under congestion was also considered in shelter location problems (Sherali and Carter, 1991; Li et al., 2012). Konur and Geunes (2010; 2011) analyzed a two-stage game to characterize the qualitative effects of traffic congestion costs on supply chain activities in a competitive environment. These existing studies are very relevant, but they did not explicitly capture the occasional unavailability of the involved facilities; i.e., a built facility may become unavailable to customers either due to limited capacity or due to the impact of disasters. Since site-specific facility availability and reliability directly affects customer allocation and traffic assignment, they should be incorporated into facility location design. Omitting the service disruption impact may result in excessive transportation cost, low service quality, and even high socio-economic penalty (e.g., loss of life and properties in an evacuation system).

A related group of literature deals with such reliability issues; i.e., built facilities may be unable to provide service due to facility disruptions (Bundschuh, Klabjan and Thurston, 2003; Snyder and Daskin, 2005) or link failures (Nel and Colbourn, 1990; Eiselt, Gendreau and Laporte, 1996). A number of reliable location models have also been proposed to address possible facility disruption risks in various contexts (e.g., Berman et al., 2007; Snyder et al., 2007; Shen et al., 2011; Qi et al., 2010; Friesz, 2011; Friesz et al., 2011; Peng et al., 2011). For example, Qi et al. (2010) and Chen et al. (2011) integrated inventory decisions into the reliable location design framework; An et al. (2013) examined the impact of service disruption on the transit-based evacuation pick-up location design; Li and Ouyang (2010b, 2012) explored facility location models for several types of network surveillance sensors under various malfunction risks. In recent year, exploring different facility disruption patterns was also of particular interest from site-independent (Snyder and Daskin, 2005; Chen et al., 2011), site-dependent (Cui et al., 2010), to spatially-correlated and cascading (Li and Ouyang, 2010a; Li et al., 2013) facility failures. Despite all the efforts in respective areas, to the best of the authors' knowledge, no previous study has addressed the interrelationship among facility location design, en-route traffic congestion, in-facility waiting delay and probabilistic service disruption in a systematic service network design framework.

To fill these gaps, this project first presents a scenario-based mathematical model that integrates service disruption risks, en-route traffic congestion and in-facility delay across the normal and all probabilistic service disruption scenarios into the facility location problem. Specifically, the model determines the optimal facility location, traffic routing, and system operation strategies including customer-to-facility assignment and server/resource allocation. In this project, we use queuing theory to model the in-facility congestion and the impact of service delay cost on the system operation. The integration of facility location, service allocation, traffic routing, queuing delay and service disruption uncertainty in the same modeling framework makes the problem highly challenging, due to the fact that (i) there is an exponential number of disruption scenarios, (ii) in each scenario facility location and customer allocation involves a large number of integer variables, and (iii) traffic routing and queuing delay functions are highly non-linear. To tackle such challenges, we propose a series of approximation treatments that can yield a lower bound (LB) and an upper bound (UB) to the original model. The main idea is to approximate the various customer arrival rates and en-route travel times respectively by their expectations. This treatment enables us to develop a more tractable MINLP formulation that significantly reduces the difficulties associated with model analysis and computation, and the approximation is found to represent the original model rather precisely. A customized Lagrangian relaxation (LR) solution framework is further developed to decompose the approximation model into two sub-problems. One of the sub-problems determines both facility location and service allocation decisions, and it is reformulated into a conic program. The other sub-problem becomes a standard traffic equilibrium problem that can be solved easily. A series of numerical experiments are performed on both a hypothetical and an empirical case to illustrate the efficiency and validity of our proposed approximation model and solution algorithm. Sensitivity analyses are also conducted to draw insights. The proposed modeling framework is generic and can be applied to a wide range of service facility location planning problems under probabilistic service disruption, e.g., bank, hospital, and shelter or emergency health center under disasters.

The remainder of this project is organized as follows. Section 5.2 introduces the cost components and the scenario-based MINLP model. Section 5.3 presents the approximation method which yields the lower and upper bounds. Section 5.4 develops the LR based algorithm. Section 5.5 shows the results from a series of numerical experiments. Section 5.6 provides conclusions and briefly discusses future research directions.

2.2 Reliable Service Facility Location Model

This section presents a scenario-based MINLP model that simultaneously addresses facility location, service capacity, service allocation, and traffic routing decisions under en-route congestion and in-facility delay across the normal and all probabilistic service disruption scenarios.

We assume that a set of customer groups, $\mathrm{I} = \{1, 2, \cdots, |\mathrm{I}|\}$, are discretely located in a given region, and each customer group $i \in \mathrm{I}$ has some demand to be satisfied in service facilities, which follows a Poisson process with rate ω_i per unit time. Facilities can be opened among a set of candidate locations, $\mathrm{J}=\{1,2,\cdots,|\mathrm{J}|\}$, to serve these customers. For an opened facility at $j\in\mathrm{J}$, regardless of its service capacity, requires a fixed infrastructure investment d_j including construction, basic operation and resource supply. For computational convenience, it is prorated to an equivalent hourly cost. In the planning stage, the agency will decide where to build service facilities among the candidate locations J. If we represent these decisions by a set of binary variables $\mathbf{y}=\{y_j \mid j\in\mathrm{J}\}$, where

$$y_j=\begin{cases}1, & \text{if location } j \text{ is selected to open a service facility}\\ 0, & \text{otherwise}\end{cases},$$

then the total service facility set-up cost is $S(\mathbf{y})=\sum_{j\in\mathrm{J}} d_j y_j$.

Considering the impacts of manmade and natural disasters, we further assume that every opened facility is subject to independent and identically distributed Bernoulli probabilistic disruption with probability $0\leq q<1$. Since each facility is either disrupted or functioning, there are $2^{|\mathrm{J}|}$ possible facility disruption scenarios, which we denote by set S, and we let δ_j^s be a binary parameter that equals to 1 if facility j is functioning in scenario $s\in\mathrm{S}$ and 0 otherwise. Let P^s denote the probability for scenario $s\in\mathrm{S}$ to occur, which depends on the number of disrupted facilities in J and can be expressed as $P^s=q^{|\mathrm{J}|-\sum_{j\in\mathrm{J}}\delta_j^s}(1-q)^{\sum_{j\in\mathrm{J}}\delta_j^s}$.

Another key decision in the strategic service system planning is to pre-assign each customer group to R ($R\leq|\mathrm{J}|$) built facilities with priorities. We denote this decision by binary variable $\mathbf{x}=\{x_{ijr} \mid \forall i\in\mathrm{I}, j\in\mathrm{J}\}$ for $r=1,\ldots,R$, where

$$x_{ijr}=\begin{cases}1, & \text{if } j \text{ is customer group } i\text{'s } r^{th} \text{ service facility location choice}\\ 0, & \text{otherwise}\end{cases}.$$

The customer assignment strategy depends on the availability of pre-assigned facilities for group i; i.e., in scenario $s\in\mathrm{S}$, group i will check the functioning facilities and choose the first functioning facility according to the pre-assigned priority order. If all built facilities in the priority list are disrupted in some extreme cases, all customers in group $i \in \mathrm{I}$ has to lose service and suffer a penalty cost. Following the idea in Cui et al. (2010), we assign the unsatisfied demand to a dummy "emergency facility", indexed by $j = |\mathrm{J}| + 1$, which incurs a fixed cost $f_{|\mathrm{J}|+1} = 0$, failure probability $q_{|\mathrm{J}|+1} = 0$ and a unit transportation cost (as penalty cost) η for all customer groups $i \in \mathrm{I}$. Furthermore, we denote the actual customer-to-facility assignment for customer group i to non-emergency facility j in scenario s as $\mathbf{x}^{\mathbf{s}} = \left\{ x_{ij}^{s} \,\middle|\, \forall i \in \mathrm{I}, j \in \mathrm{J}, \mathrm{s} \in S \right\}$, where

$$x_{ij}^{s} = \begin{cases} 1, & \text{if facility } j \text{ is customer group } i\text{'}s \text{ service facility choice under scenario } s \\ 0, & \text{otherwise} \end{cases}.$$

The arrival of customers' demand at a functioning facility j also follows a Poisson process with rate $\lambda_j^s = \sum_{i \in \mathrm{I}} \omega_i x_{ij}^s$. Besides opening facilities, we also determine the service capacity denoted by c_j^s; i.e., the number of identical servers for each facility at location j in scenario s. For simplicity, we assume that the service time of each server is exponentially distributed with rate μ_0, and there is a site-dependent service cost m_j per server per unit time. For model simplicity, we consolidate all c_j^s servers in facility j into one with an equivalent service rate $\mu_j^s = c_j^s \cdot \mu_0$. Now the queuing system in each facility j follows an M/M/1 FCFS (first-come, first-served) mode, and the expected total time that customers spend on waiting and receiving service per unit time in facility j in scenario s is $\frac{\lambda_j^s}{\mu_j^s - \lambda_j^s}, \;\; \forall j \in \mathrm{J}, s \in \mathrm{S}$ if $\lambda_j^s < \mu_j^s$. Recall that the service cost in facility j is proportional to the number of servers, $\frac{\mu_j^s}{\mu_0} m_j$, and hence the total in-facility congestion cost $W(\boldsymbol{\lambda}^{\mathbf{s}}, \boldsymbol{\mu}^{\mathbf{s}})$ per unit time in scenario s is:

$$W(\boldsymbol{\lambda}^s, \boldsymbol{\mu}^s) = \sum_{j \in \mathrm{J}} \left(\frac{\beta \lambda_j^s}{\mu_j^s - \lambda_j^s} + \frac{\mu_j^s}{\mu_0} m_j \right), \ \lambda_j^s < \mu_j^s, \ \forall j \in \mathrm{J}, \ s \in \mathrm{S},$$

where parameter β denote the in-facility waiting time cost, and it also captures the relative weight of customer waiting cost against other cost components. $W(\boldsymbol{\lambda}^s, \boldsymbol{\mu}^s)$ can be further simplified by eliminating the decision variable μ_j^s, $\forall j \in \mathrm{J}, \ s \in \mathrm{S}$; namely, if (1) is minimized over μ_j^s, the optimal service rate is concluded to be $\lambda_j^s + \sqrt{\frac{\mu_0 \beta}{m_j}} \cdot \sqrt{\lambda_j^s}$ in each scenario $s \in \mathrm{S}$. Then equation (1) is simplified as

$$W(\boldsymbol{\lambda}^s) = \sum_{j \in \mathrm{J}} \left(2\sqrt{\frac{\beta m_j}{\mu_0}} \cdot \sqrt{\lambda_j^s} + \frac{m_j}{\mu_0} \lambda_j^s \right), \ \forall j \in \mathrm{J}, \ s \in \mathrm{S}.$$

Customers travel to the assigned functioning service facilities via a transportation network $\mathrm{M} = (\mathrm{N}, \ \mathrm{A})$, where N is the set of nodes and A is the set of directed links, and $\mathrm{I}, \mathrm{J} \subset \mathrm{N}$. For each node $n \in \mathrm{N}$, we define A_n^+ as the set of outbound flow links starting from node n, and A_n^- is the set of inbound flow links ending in n. In an arbitrary scenario s, variable $k_{i,j}^s$ denotes the fraction of customer demand from customer group $i \in \mathrm{I}$ to facility location $j \in \mathrm{J}$. On each link $a \in \mathrm{A}$, the flow associated with origin $i \in \mathrm{I}$ to destination $j \in \mathrm{J}$ in scenario s is $f_a^{i,j,s}$, and hence the corresponding total link flow z_a^s satisfies

$$z_a^s = \sum_{i \in \mathrm{I}} \sum_{j \in \mathrm{J}} f_a^{i,j,s}, \ \forall a \in \mathrm{A}.$$

The link travel time $t_a^s(z_a^s)$ is assumed to be a function of the link flow based on the BPR function (U.S. Bureau of Public Roads); i.e.,

$$t_a^s(z_a^s) = t_a^0\left(1 + 0.15(z_a^s / C_a)^4\right), \text{ if } z_a^s \le C_a,$$

where t_a^0 and C_a are respectively the free flow travel time and traffic capacity of link $a \in \mathrm{A}$. In some extreme cases, e.g., there are very few functioning facilities left (which is very rare, but possible), some link flows may exceed their capacities, i.e. $z_a^s > C_a$. To accommodate such possibilities, we further assume that the link travel time $t_a^s(z_a^s)$ for $z_a^s > C_a$ is a simple linear projection of the BPR function that is continuous and smooth at point $\left(C_a, 1.15t_a^0\right)$, as follows:

$$t_a^s(z_a^s) = 1.75t_a^0 - 0.6t_a^0(C_a / z_a^s), \text{ if } z_a^s > C_a.$$

We also use α to denote the monetary value of travel time which captures the relative weight of total travel cost against the facility setup and service costs, and thus the total expected travel cost $T(\mathbf{z^s})$ per unit time in scenario s can be written as

$$T(\mathbf{z^s}) = \alpha \sum_{a \in \mathrm{A}} z_a^s t_a^s(z_a^s).$$

With the parameters and decision variables described above, we formulate the total expected system cost across all probabilistic service disruption scenarios in S, as follows,

$$\Phi\left(\mathbf{x}, \mathbf{x^s}, \mathbf{y}, \mathbf{z^s}, \mathbf{f^s}, \mathbf{k^s}\right) := \sum_{j \in \mathrm{J}} d_j y_j + \sum_{s \in \mathrm{S}} P^s \left[\alpha \sum_{a \in \mathrm{A}} z_a^s t_a(z_a^s) + \sum_{j \in \mathrm{J}} \left(2\sqrt{\frac{\beta m_j}{\mu_0} \lambda_j^s} + \frac{m_j}{\mu_0} \lambda_j^s \right) + \eta \sum_{i \in \mathrm{I}} \omega_i x_{i,|\mathrm{J}|+1}^s \right].$$

The scenario-based reliable facility location model is formulated as a mixed integer non-linear program (8) as follows. It determines the optimal service facility locations $\{y_j\}$, facility priority decision of each customer group $\{x_{ijr}\}$, customer-to-facility assignments $\{x_{ij}^s\}$ in each scenario, and link flow related variables $\left\{z_a^s\right\}$, $\left\{f_a^{i,j,s}\right\}$ and $\left\{k_{i,j}^s\right\}$ as a result of optimal traffic assignment to minimize the total expected system cost $\Phi\left(\mathbf{x}, \mathbf{x^s}, \mathbf{y}, \mathbf{z^s}, \mathbf{f^s}, \mathbf{k^s}\right)$, which includes facility setup cost $S(\mathbf{y})$, en-route travel cost $T(\mathbf{z}^s)$ and in-facility delay cost $W(\boldsymbol{\lambda}^s)$ across all possible facility disruption scenarios.

$$\underset{\mathbf{x},\mathbf{x}^s,\mathbf{y},\mathbf{z}^s,\mathbf{f}^s,\mathbf{k}^s}{\text{Min}} \sum_{j\in \mathrm{J}} d_j y_j + \sum_{s\in \mathrm{S}} P^s \left[\alpha \sum_{a\in \mathrm{A}} z_a^s t_a(z_a^s) + \sum_{j\in \mathrm{J}} \left(2\sqrt{\sum_{i\in \mathrm{I}} \frac{\beta m_j}{\mu_0} \omega_i x_{ij}^s} + \frac{m_j}{\mu_0} \sum_{i\in \mathrm{I}} \omega_i x_{ij}^s \right) + \eta \sum_{i\in \mathrm{I}} \omega_i x_{i,|\mathrm{J}|+1}^s \right] \quad \text{(1a)}$$

subject to (3), (4), (5), and

$$k_{i,j}^s = \omega_i x_{ij}^s \qquad \forall i\in \mathrm{I}, j\in \mathrm{J}, s\in \mathrm{S} \quad \text{(1b)}$$

$$\sum_{a\in \mathrm{A}_n^+} f_a^{i,j,s} - \sum_{a\in \mathrm{A}_n^-} f_a^{i,j,s} = \begin{cases} k_{i,j}^s, & \forall n\in \mathrm{I} \\ 0, & \forall n\in \mathrm{N}\backslash(\mathrm{I}\bigcup \mathrm{J}) \\ -k_{i,j}^s, & \forall n\in \mathrm{J} \end{cases} \quad \text{(1c)}$$

$$\sum_{r=1}^{R} x_{ijr} \le y_j \qquad \forall i\in \mathrm{I}, j\in \mathrm{J} \quad \text{(1d)}$$

$$\sum_{j\in \mathrm{J}} x_{ijr} + \sum_{l=1}^{r} x_{i,|\mathrm{J}|+1,l} = 1 \qquad \forall i\in \mathrm{I}, 1\le r\le R+1 \quad \text{(1e)}$$

$$\sum_{r=1}^{R+1} x_{i,|\mathrm{J}|+1,r} = 1 \qquad \forall i\in \mathrm{I} \quad \text{(1f)}$$

$$\sum_{j\in \mathrm{J}} x_{ij}^s + x_{i,|\mathrm{J}|+1}^s = 1 \qquad \forall i\in \mathrm{I}, s\in \mathrm{S} \quad \text{(1g)}$$

$$x_{ij}^s \ge x_{ijr}\delta_j^s - \sum_{l=1,\ldots,r-1} \sum_{j'\in \mathrm{J}\bigcup\{|\mathrm{J}|+1\}\backslash\{j\}} x_{ij'l}\delta_{j'}^s \qquad \forall i\in \mathrm{I}, j\in \mathrm{J}\bigcup\{|\mathrm{J}|+1\}, r=1,\ldots,R+1 \quad \text{(1h)}$$

$$y_j, x_{ij}^s, x_{ijr} = \{0,1\} \qquad \forall i\in \mathrm{I}, 1\le j\le |\mathrm{J}|+1, s\in \mathrm{S}, r=1,\ldots,R+1 \quad \text{(1i)}$$

$$z_a^s, k_{i,j}^s, f_a^{i,j,s} \ge 0 \qquad \forall a\in \mathrm{A}, i\in \mathrm{I}, j\in \mathrm{J}, s\in \mathrm{S}$$

The objective function (8a) minimizes the total expected system cost. Constraints (8b) define the traffic flow assigned for each OD pair (i, j) in scenario s. Constraints (8c) enforce traffic flow conservation at all origins, destinations and other network nodes. Constraints (8d) guarantee that a facility should be built before assigned to a customer group. Constraints (8e) state that for each customer group $i\in \mathrm{I}$ at lever r, either they are assigned to a service facility $j\in \mathrm{J}$ or an emergency facility $|\mathrm{J}|+1$, but for levels $l<r$, no pre-assignment to emergency facility

occurs. Constraints (8f) require each customer to be assigned to the emergency facility at a certain level. Constraints (8g) assure that in scenario s each customer group $i \in \mathrm{I}$ will actually be assigned either to one of its pre-assigned functioning facilities or the "emergency facility". Constraints (8h) implement the actual customer assignment strategy based on the disruption scenario s and pre-assigned priority orders. Constraints (8i) and (8j) specify all binary and nonnegative variables.

Model (8) is a non-convex, nonlinear, mixed-integer, and scenario-based stochastic program. The facility location part of the problem is by itself NP-hard, while the cost functions of in-facility waiting and en-route travel are highly non-linear. Most importantly, the stochasticity associated with facility disruptions (i.e., the exponential number of the total potential facility disruption scenarios $O(|\mathrm{J}|^R)$) adds a huge layer of complexity. Even if the optimal facility location design and the customer-to-facility assignment are given, it is still difficult to evaluate the total expected en-route travel and in-facility waiting costs. In other words, each of the exponential number of possible disruption scenarios involves a network assignment problem with a different set of highly-nonlinear cost components. In the next section, we propose an approximation model that can yield tight lower and upper bounds to the original problem (8). This new model can be solved efficiently to near-optimum with acceptable optimality gaps.

2.3 <u>Model Approximation</u>

2.3.1 Lower Bound (LB) Formulation

Recall that the in-facility waiting cost in each scenario, $W(\boldsymbol{\lambda}^s) = \sum_{j \in \mathrm{J}} \left(2\sqrt{\frac{\beta m_j}{\mu_0}} \cdot \sqrt{\lambda_j^s} + \frac{m_j}{\mu_0} \lambda_j^s \right)$, is a concave function of $\boldsymbol{\lambda}^s$, and the expected total in-facility delay cost across all possible scenarios is:

$$E[W(\boldsymbol{\lambda}^s)] = \sum_{j \in \mathrm{J}} \sum_{s \in \mathrm{S}} P^s \left(2\sqrt{\frac{\beta m_j}{\mu_0}} \cdot \sqrt{\lambda_j^s} + \frac{m_j}{\mu_0} \lambda_j^s \right).$$

The concave function $2\sqrt{\frac{\beta m_j}{\mu_0}} \cdot \sqrt{\lambda_j^s} + \frac{m_j}{\mu_0}\lambda_j^s, \forall j \in \mathrm{J}, s \in \mathrm{S}$ can be bounded from below by a linear function in the range of 0 and the upper bound of λ_j^s (denote as $\lambda_j^{s,up}$), i.e., $a_j + b_j\lambda_j^s \le 2\sqrt{\frac{\beta m_j}{\mu_0}} \cdot \sqrt{\lambda_j^s} + \frac{m_j}{\mu_0}\lambda_j^s, \forall \lambda_j^s \in [0, \lambda_j^{s,up}]$, where a_j and b_j are parameters to be calibrated. Now we obtain a lower bound of $E[W(\boldsymbol{\lambda}^s)]$ with appropriately chosen $\{a_j\}$ and $\{b_j\}$:

$$E[W(\boldsymbol{\lambda}^s)] \ge \sum_{j\in\mathrm{J}}\sum_{s\in\mathrm{S}} P^s \cdot (a_j + b_j\lambda_j^s) = \sum_{j\in\mathrm{J}}\left(a_j + b_j\sum_{s\in\mathrm{S}} P^s\lambda_j^s\right), \forall \lambda_j^s \in [0, \lambda_j^{s,up}].$$

Along the secant line, it can be found that when λ_j^s equals 0, which means no customer gathered in facility *j* in scenario s and thus no in-facility congestion, then parameter $a_j = 0$; while if only facility j is functioning in a certain scenario, the maximal possible number of customers gathered in facility *j* is $\lambda_j^{s,up} = \sum_{i\in\mathrm{I}}\omega_i$. Thus, by setting $a_j + b_j\lambda_j^{s,up} = 2\sqrt{\frac{\beta m_j}{\mu_0}} \cdot \sqrt{\lambda_j^{s,up}} + \frac{m_j}{\mu_0}\lambda_j^{s,up}$, we can obtain the value of parameter b_j, i.e., $b_j = 2\sqrt{\frac{\beta m_j}{\mu_0\sum_{i\in\mathrm{I}}\omega_i}} + \frac{m_j}{\mu_0}$.

Recall that in any scenario, customer group i always checks the availability of built facilities in its pre-assignment list until the first functioning facility is selected or the demand is lost and suffers a penalty cost when there is no functioning facility. Therefore, the following equation holds:

$$E[\lambda_j^s] = \sum_{i\in\mathrm{I}}\sum_{s\in\mathrm{S}} P^s\omega_i x_{ij}^s = \sum_{i\in\mathrm{I}}\sum_{r=1}^{R}\omega_i(1-q)q^{r-1}x_{ijr}.$$

According to inequality (12), $E[W(\boldsymbol{\lambda}^s)]$ can be bounded from below as follows

$$E[W(\boldsymbol{\lambda}^s)] \geq \sum_{j\in J}\sum_{i\in I}\sum_{r=1}^{R}\left(2\sqrt{\frac{\beta m_j}{\mu_0\sum_{i\in I}\omega_i}}+\frac{m_j}{\mu_0}\right)\omega_i(1-q)q^{r-1}x_{ijr}.$$

On the other hand, the travel cost in each scenario s, $T(\mathbf{z}^s)=\alpha\sum_{a\in A} z_a^s t_a(z_a^s)$ is a convex function of link flow $\mathbf{z}^s$. According to Jensen's Inequality, the total expected travel cost always satisfies $\alpha\sum_{a\in A}\sum_{s\in S}P^s T(z_a^s) \geq \alpha\sum_{a\in A} T\left(\sum_{s\in S}P^s z_a^s\right)$.

As such, we arrive at the following model which can be shown to be a lower bound of original model (8),

$$\underset{\mathbf{x}',\mathbf{y}',\mathbf{z},\mathbf{f},\mathbf{k}}{\text{Min}} \sum_{j\in J} d_j y_j + \alpha\sum_{a\in A} z_a t_a(z_a) + \sum_{j\in J}\sum_{i\in I}\sum_{r=1}^{R}\left(2\sqrt{\frac{\beta m_j}{\mu_0\sum_{i\in I}\omega_i}}+\frac{m_j}{\mu_0}\right)\omega_i(1-q)q^{r-1}x_{ijr} + \sum_{i\in I}\sum_{r=1}^{R+1}\omega_i\eta q^{r-1}x_{i,|J|+1,r}$$

Subject to 8(d)-8(f), and

$$t_a(z_a)=\begin{cases} t_a^0\left(1+0.15(z_a/C_a)^4\right), & z_a \leq C_a \\ 1.75t_a^0-0.6t_a^0(C_a/z_a), & z_a > C_a \end{cases}$$

$$z_a=\sum_{i\in I}\sum_{j\in J} f_a^{i,j} \qquad \forall a\in A$$

$$k_{i,j}=\sum_{r=1}^{R}\omega_i(1-q)q^{r-1}x_{ijr} \qquad \forall i\in I, j\in J$$

$$\sum_{a\in A_n^+} f_a^{i,j} - \sum_{a\in A_n^-} f_a^{i,j} = \begin{cases} k_{i,j}, & \forall n\in I \\ 0, & \forall n\in N\backslash(I\cup J) \\ -k_{i,j}, & \forall n\in J \end{cases}$$

$$y_j, x_{ijr} = \{0,1\} \qquad \forall i \in \mathrm{I}, 1 \le j \le |\mathrm{J}|+1, 1 \le r \le R+1$$

$$z_a, k_{i,j}, f_a^{i,j} \ge 0 \qquad \forall a \in \mathrm{A}, i \in \mathrm{I}, j \in \mathrm{J}$$

Especially, constraints (13d) ensure that flow assigned for each OD pair (i, j) is equal to the expected flow from origin $i \in \mathrm{I}$ to destination $j \in \mathrm{J}$ across all assignment levels. The other constraints have similar explanations as those of model (8) but not limited to a certain scenario. The following proposition shows the optimal solution of model (13) indeed yields a lower bound for model (8).

Proposition 1. *The optimal objective function of model (8) is always bounded from below by that of model (13).*

Proof. Since we have already shown that for any feasible solution, the objective of (13) is smaller than that of (8), we only need to prove that the optimal solutions of original model (8) yields a feasible solution to model (13) with a smaller or equal objective value. Let $\{\mathbf{x}^*, \mathbf{x}^{s*}, \mathbf{y}^*, \mathbf{z}^{s*}, \mathbf{f}^{s*}, \mathbf{k}^{s*}\}$ be the optimal solution of original model (8), then based on the definition of $\mathbf{x}^*$, each customer group corresponds to a facility visiting sequence, which means $\mathbf{x}^*$ and $\mathbf{y}^*$ can be transformed into a feasible solution of (13) denoted as $\mathbf{x}' = \mathbf{x}^*$ and $\mathbf{y}' = \mathbf{y}^*$, respectively. Clearly, $\mathbf{x}'$ and $\mathbf{y}'$ satisfy constraints (8d), (8e), and (8f), and (13f) always hold for these values. In addition, since in any scenario $s \in \mathrm{S}$, the constraints (13b), (13c), (13e) and (13g) hold for the optimal $\mathbf{z}^{s*}$, $\mathbf{f}^{s*}$, and $\mathbf{k}^{s*}$, thus the expected values $z_a = \sum_{s \in \mathrm{S}} P^s z_a^{s*}$, $f_a^{i,j} = \sum_{s \in \mathrm{S}} P^s f_a^{i,j,s*}$, and $k_{i,j} = \sum_{s \in \mathrm{S}} P^s k_{i,j}^{s*}$, also satisfy these constraints as well as constraints (13d). This implies that we can transform the optimal solution of original model (8) into a feasible solution of model (13), with a lower objective value which embodies in the deviation of total expected en-route travel cost and in-facility waiting cost above. This completes the proof. □

2.3.2 Upper Bound (UB) Formulation

For the in-facility delay cost formula, since $\sum_{i\in \mathrm{I}} \frac{4\beta m_j}{\mu_0} \omega_i x_{ij}^s \ge 0$, according to the basic inequality $\chi \le \left(\frac{\chi+1}{2}\right)^2, \chi \ge 0$, it is easy to obtain the inequality

$$\left(\sum_{i\in \mathrm{I}} \frac{4\beta m_j}{\mu_0} \omega_i x_{ij}^s\right)^{\frac{1}{2}} \le \left(\frac{4\beta m_j \omega_{\min}}{\mu_0}\right)^{\frac{1}{2}} \cdot \left(\frac{1}{2}\sum_{i\in \mathrm{I}} \frac{\omega_i}{\omega_{\min}} x_{ij}^s + \frac{1}{2}\right),$$

where $\omega_{\min} = \min_{i\in \mathrm{I}}\{\omega_i\}$. As such, the in-facility delay cost can be bounded from above as follows,

$$\sum_{j\in \mathrm{J}}\sum_{s\in \mathrm{S}} P^s \left[\left(\sum_{i\in \mathrm{I}} \frac{4\beta m_j}{\mu_0} \omega_i x_{ij}^s\right)^{\frac{1}{2}} + \sum_{i\in \mathrm{I}} \frac{m_j}{\mu_0} \omega_i x_{ij}^s\right] \le \sum_{i\in \mathrm{I}}\sum_{j\in \mathrm{J}}\sum_{s\in \mathrm{S}} \left(\sqrt{\frac{\beta m_j}{\mu_0 \omega_{\min}}} + \frac{m_j}{\mu_0}\right) P^s \omega_i x_{ij}^s + \sum_{j\in \mathrm{J}} \left(\frac{\beta m_j \omega_{\min}}{\mu_0}\right)^{\frac{1}{2}}$$

The following inequality also holds for the expected total travel cost,

$$\alpha \sum_{a\in \mathrm{A}}\sum_{s\in \mathrm{S}} P^s z_a^s t_a^0 (z_a^s) \le \begin{cases} 1.15\alpha \sum_{a\in \mathrm{A}}\sum_{s\in \mathrm{S}} P^s z_a^s t_a^0, & z_a^s \le C_a \\ \alpha \sum_{a\in \mathrm{A}}\sum_{s\in \mathrm{S}} P^s \left(1.75 z_a^s t_a^0 - 0.6 t_a^0 C_a\right), & z_a^s > C_a \end{cases} \le 1.75\alpha \sum_{a\in \mathrm{A}}\sum_{s\in \mathrm{S}} P^s z_a^s t_a^0,$$

Using (14), (15) and (11) to replace their counterparts in the original model, we have the following,

$$\underset{\mathbf{x}',\mathbf{y}',\mathbf{z},\mathbf{f},\mathbf{k}}{\mathrm{Min}} \sum_{j\in \mathrm{J}} d_j y_j + 1.75 \sum_{a\in \mathrm{A}} z_a t_a^0 + \sum_{i\in \mathrm{I}}\sum_{j\in \mathrm{J}}\sum_{r=1}^{R} \left(\sqrt{\frac{\beta m_j}{\mu_0 \omega_{\min}}} + \frac{m_j}{\mu_0}\right) \omega_i (1-q) q^{r-1} x_{ijr} + \sum_{j\in \mathrm{J}} \left(\frac{\beta m_j \omega_{\min}}{\mu_0}\right)^{\frac{1}{2}}$$
$$+ \sum_{i\in \mathrm{I}}\sum_{r=1}^{R+1} \omega_i \eta q^{r-1} x_{i,|\mathrm{J}|+1,r}$$
.

Subject to (8d) - (8f), and (13b) - (13g).

The optimal solution to model (16) yields an upper bound to that of model (8), as stated below.

Proposition 2. *The optimal objective of model (16) is an upper bound to that of model* (8).

Proof. In order to demonstrate that the original scenario-based formulation can be bounded by model (16), it is sufficient to show that the optimal solution of model (16) yields a feasible solution to model (8). Let $\left\{\mathbf{x}'^{*},\mathbf{y}'^{*},\mathbf{z}^{*},\mathbf{f}^{*},\mathbf{k}^{*}\right\}$ be the optimal solution of UB model (16). As we know, $\mathbf{x}'=\left\{x_{ijr}\middle| i\in \mathrm{I}, j\in \mathrm{J}, r=1,...,R\right\}$ provide ordered service allocation decisions for customer groups, through constraints (8d), (8e), (8f), and (13f), they guarantee there is a unique visiting strategy for each customer group in case possible facility disruption happens. Based on the optimal facility locations $\mathbf{y}'^{*}$ and facility binary functioning indicator δ_j^s in each scenario, we can get $\mathbf{y}=\mathbf{y}'^{*}$ and decompose $\mathbf{x}'$ to a series of scenario-based service allocation variables $\mathbf{x}^{\mathbf{s}}=\left\{x_{ij}^s\middle|\sum_s P^s x_{ij}^s=\sum_{r=1}^{R}(1-q)q^{r-1}x_{ijr}, \forall i\in \mathrm{I}, j\in \mathrm{J}, s\in \mathrm{S}\right\}$ for all possible disruption scenarios. Let $\left\{f_a^{i,j,r},\forall a\right\}$ denote the portion of optimal link flow $\left\{\left(f_a^{i,j}\right)^{*},\forall a\right\}$, which is originated from the demand $\omega_i\left(1-q\right)q^r\left(x'_{ijr}\right)^{*}$ and could be zero if j is not customer i's rth choice. Clearly, $\left\{f_a^{i,j,r},\forall a\right\}$ satisfy constraints (13e) with $\left(k_{i,j}\right)^{*}=\omega_i\left(1-q\right)q^r\left(x'_{ijr}\right)^{*}$. In addition, in each scenario $s\in \mathrm{S}$, for each customer $i\in \mathrm{I}$, suppose $j_i^s\in \mathrm{J}$ is i's assigned facility in this scenario and ranks $\hat{r}$ of the visiting sequence, then $\left\{f_a^{i,j,s}\right\}$ can be defined as follows:

$$f_a^{i,j,s}=\frac{\delta_j^s f_a^{i,j,\hat{r}}\left(x'_{ij\hat{r}}\right)^{*}}{\left(1-q\right)q^{\hat{r}}}=\begin{cases}\dfrac{f_a^{i,j,\hat{r}}}{\left(1-q\right)q^{\hat{r}}}, & \text{if } j=j_i^s, \forall i\in \mathrm{I}, \forall s\in \mathrm{S}, \forall a\in \mathrm{A}\\ 0, & \text{otherwise}\end{cases}$$

which clearly satisfies (8c). Now we have

$$\sum_{s\in S} P^s z_a^s = \sum_{s\in S}\sum_{i\in I}\sum_{j\in J} P^s f_a^{i,j,s} = \sum_{s\in S}\sum_{i\in I}\sum_{j\in J} P^s \frac{\delta_j^s f_a^{i,j,\hat{r}}\left(x'_{ij\hat{r}}\right)^*}{(1-q)q^{\hat{r}}} = \sum_{i\in I}\sum_{j\in J}\sum_{s\in S} P^s \frac{\delta_j^s f_a^{i,j,r_i}\left(x'_{ij\hat{r}}\right)^*}{(1-q)q^{\hat{r}}}$$
$$= \sum_{i\in I}\sum_{j\in J} f_a^{i,j,\hat{r}}\left(x'_{ij\hat{r}}\right)^* = \sum_{i\in I}\sum_{j\in J} f_a^{i,j} = \left(z'_a\right)^*, \forall a \in A$$

From the above construction, we will have a set of feasible solution $\{z_a^s\}$, $\{f_a^{i,j,s}\}$ and $\{k_{i,j}^s\}$ for the original model (8), which has a lower objective value than the optimal value of model (16). □

2.3.3 Approximation Model (AM) Formulation

Based on the derivation of lower and upper bounds above, now we propose an approximate MINLP model that is much more computationally tractable. The expected in-facility delay cost and travel cost are approximated by using the expected values of arrival rate in facilities and traffic flow. For example, considering the expected arrival rate $\bar{\lambda}_j = \sum_{i\in I}\sum_{r=1}^{R} \omega_i (1-q) q^{r-1} x_{ijr}$, the approximated total in-facility delay cost per unit time can be shown as follows

$$W(\bar{\boldsymbol{\lambda}}) = \sum_{j\in J} \left(2\sqrt{\frac{\beta m_j}{\mu_0}} \cdot \sqrt{\bar{\lambda}_j} + \frac{m_j}{\mu_0} \bar{\lambda}_j\right).$$

Similarly, the total travel cost can be approximated by plugging in the expected flow $\mathbf{z}$ which is linearly dependent on $\bar{\boldsymbol{\lambda}}$. Therefore, the following mixed integer non-linear program is presented to approximate the true value of total expected system cost (8). The following proposition will show that the optimal value of model (17) is always within the range of the lower bound from (13) and upper bound from (16). In later sections, we will show numerically that (17) provides near optimal solutions, and is much easier to solve for large scale problems.

$$\underset{\mathbf{x},\mathbf{y},\mathbf{z},\mathbf{f},\mathbf{k}}{\text{Min}} \sum_{j\in J} d_j y_j + \alpha \sum_{a\in A} z_a t_a(z_a) + \sum_{j\in J}\left[\left(\sum_{i\in I}\sum_{r=1}^{R}\frac{4\beta m_j}{\mu_0}\omega_i(1-q)q^{r-1}x_{ijr}\right)^{\frac{1}{2}} + \sum_{i\in I}\sum_{r=1}^{R}\frac{m_j}{\mu_0}\omega_i(1-q)q^{r-1}x_{ijr}\right]$$
$$+\sum_{i\in I}\sum_{r=1}^{R+1}\omega_i \eta q^{r-1} x_{i,|J|+1,r} \tag{10}$$

Subject to (8d) - (8f), and (13b) - (13g).

Proposition 3. *The optimal objective of AM model (18) is bounded from below by that of model (13), and from above by that of model (16).*

Proof. The AM model has the similar structure with LB model (13) and UB model (16). In order to prove the optimal value of model (18) is smaller than (16) and greater than (13), according to the composition of objective functions of (13), (16) and (18), the key point is to demonstrate the establishment of following inequalities, (i) the cost term

$$\sum_{j\in J}\left(\sum_{i\in I}\sum_{r=1}^{R}\frac{4\beta m_j}{\mu_0}\omega_i(1-q)q^{r-1}x_{ijr}\right)^{\frac{1}{2}} \ge \sum_{j\in J}\sum_{i\in I}\sum_{r=1}^{R} 2\sqrt{\frac{\beta m_j}{\mu_0 \sum_{i\in I}\omega_i}}\omega_i(1-q)q^{r-1}x_{ijr} \tag{11}$$

and (ii)

$$\sum_{j\in J}\left(\sum_{i\in I}\sum_{r=1}^{R}\frac{4\beta m_j}{\mu_0}\omega_i(1-q)q^{r-1}x_{ijr}\right)^{\frac{1}{2}} \le \sum_{j\in J}\left(\frac{\beta m_j \omega_{\min}}{\mu_0}\right)^{\frac{1}{2}} + \sum_{i\in I}\sum_{j\in J}\sum_{r=1}^{R}\sqrt{\frac{\beta m_j}{\mu_0 \omega_{\min}}}\omega_i(1-q)q^{r-1}x_{ijr}$$

For inequality (19), when decomposing both sides by j and remove the constant terms $2\sqrt{\frac{\beta m_j}{\mu_0}}$, we can obtain

$$\left(\sum_{i\in I}\sum_{r=1}^{R}\omega_i(1-q)q^{r-1}x_{ijr}\right)^{\frac{1}{2}} \ge \sum_{i\in I}\sum_{r=1}^{R}\frac{\omega_i(1-q)q^{r-1}x_{ijr}}{\sqrt{\sum_{i\in I}\omega_i}}.$$

Since $\sum_{i\in I}\omega_i \ge \sum_{i\in I}\sum_{r=1}^{R}\omega_i(1-q)q^{r-1}x_{ijr}$ always hold when $0 \le q < 1$, the inequality (19) holds.

In the other side, we conduct the same decomposition for (20) by j and remove the constant term $\sqrt{\frac{\beta m_j}{\mu_0}}$ for both sides, inequality (20) becomes

$$\left(\sum_{i\in \mathrm{I}}\sum_{r=1}^{R}4\omega_i(1-q)q^{r-1}x_{ijr}\right)^{\frac{1}{2}} \le \sqrt{\omega_{\min}}+\frac{1}{\sqrt{\omega_{\min}}}\sum_{i\in \mathrm{I}}\sum_{r=1}^{R}\omega_i(1-q)q^{r-1}x_{ijr}$$
.

Obviously it always holds, and it leads to the establishment of inequality (20). □

2.4 Solution Approach

2.4.1 Lagrangian relaxation

To solve the LB model (13), UB model (16) and integrated AM model (18), we propose a customized Lagrangian relaxation algorithm to find near-optimum solutions. Due to the similar model structure, we only introduce the LR-based algorithm design for the integrated AM model which immediately applies to all other models. We relax constraints (13d) with a set of nonnegative multipliers $\boldsymbol{\phi}=\{\phi_{ij}\in \mathrm{R}\}$ and add them to objective (18), and the relaxed problem becomes:

$$\Phi(\boldsymbol{\phi}):=\min_{\mathbf{x},\mathbf{y},\mathbf{z},\mathbf{f},\mathbf{k}}\ \sum_{j\in \mathrm{J}}d_j y_j+\sum_{j\in \mathrm{J}}\left[\left(\sum_{i\in \mathrm{I}}\sum_{r=1}^{R}\frac{4\beta m_j}{\mu_0}\omega_i(1-q)q^{r-1}x_{ijr}\right)^{\frac{1}{2}}+\sum_{i\in \mathrm{I}}\sum_{r=1}^{R}\frac{m_j}{\mu_0}\omega_i(1-q)q^{r-1}x_{ijr}\right]$$
$$+\alpha\sum_{a\in \mathrm{A}}z_a t_a(z_a)+\sum_{i\in \mathrm{I}}\sum_{r=1}^{R+1}\omega_i\eta q^{r-1}x_{i,|\mathrm{J}|+1,r}+\sum_{i\in \mathrm{I}}\sum_{j\in \mathrm{J}}\phi_{ij}\left(\sum_{r=1}^{R}\omega_i(1-q)q^{r-1}x_{ijr}-k_{i,j}\right)$$

Subject to (8d) - (8f), and (13b), (13c), (13e)-(13g).

For any given $\boldsymbol{\phi}$, $\Phi(\boldsymbol{\phi})$ is a lower bound of approximation model (18). Note that the relaxed problem (21) can be further decomposed into two sub-problems.

Sub-problem 1: (traffic assignment)

$$\operatorname*{Min}_{\mathbf{z},\mathbf{f},\mathbf{k}}\ \alpha\sum_{a\in \mathrm{A}}z_a t_a^0\left(1+0.15(z_a/C_a)^4\right)-\sum_{i\in \mathrm{I}}\sum_{j\in \mathrm{J}}\phi_{ij}k_{i,j} \qquad (12)$$

Subject to (13b), (13c), (13e) and (13g).

Sub-problem 2: (facility location and service allocation)

$$\underset{\mathbf{x},\mathbf{y}}{\text{Min}} \quad \sum_{j\in J} d_j y_j + \sum_{j\in J}\left[\left(\sum_{i\in I}\sum_{r=1}^{R}\frac{4\beta m_j}{\mu_0}\omega_i(1-q)q^{r-1}x_{ijr}\right)^{\frac{1}{2}} + \sum_{i\in I}\sum_{r=1}^{R}(\frac{m_j}{\mu_0}+\phi_{ij})\omega_i(1-q)q^{r-1}x_{ijr}\right] \tag{13}$$
$$+\sum_{i\in I}\sum_{r=1}^{R+1}\omega_i\eta q^{r-1}x_{i,|J|+1,r}$$

Subject to (8d) - (8f) and (13f).

Note that both LB and UB models are mixed integer linear programs, so the resulting decomposed sub-problems for these two models can be solved by existing solvers. However, Sub-problem 1 is a convex optimization problem, which can be easily solved by commercial solvers. Sub-problem 2 contains only binary variables, but it is very difficult to solve because of the square root term in the objective function. In the next subsection, we manage to reformulate it into an equivalent conic quadratic MIP that is solvable.

2.4.2 Conic quadratic MIP reformulation of sub-problem 2

To solve the type of mixed integer programs with square root terms in the objective e.g., problem (23), an LR based polynomial-time exact algorithm was developed by Chen et al. (2011). In this project, we adopt a conic quadratic MIP reformulation in Atamtürk et al. (2012), by which means Sub-problem 2 can be solved by commercial solvers (e.g., CPLEX).

We introduce an auxiliary variables $\mathbf{h}=\{h_j\}\in R^{|J|}$ to represent the square root terms in the objective of (23), i.e., $h_j=\left(\sum_{i\in I}\sum_{r=1}^{R}\omega_i(1-q)q^{r-1}x_{ijr}\right)^{\frac{1}{2}}$, $\forall j\in J$ and based on the fact that $x_{ijr}=x_{ijr}^2$ for any binary variable x_{ijr}, we reformulate (23) as

$$\underset{\mathbf{x},\mathbf{y},\mathbf{h}}{\text{Min}} \quad \sum_{j\in J}\left[d_j y_j + 2\sqrt{\frac{\beta m_j}{\mu_0}}h_j + \frac{m_j}{\mu_0}h_j^2 + \sum_{i\in I}\sum_{r=1}^{R}\phi_{ij}\omega_i(1-q)q^{r-1}x_{ijr}\right] + \sum_{i\in I}\sum_{r=1}^{R+1}\omega_i\eta q^{r-1}x_{i,|J|+1,r} \tag{14a}$$

Subject to

$$\sum_{i\in I}\sum_{r=1}^{R}\omega_i(1-q)q^{r-1}x_{ijr}^2 \le h_j^2 \qquad \forall j\in J \tag{14b}$$

$$h_j \ge 0 \qquad \forall j\in J \tag{14c}$$

(8d) - (8g) and (13f).

Now the objective function (24a) is quadratic, and the constraints are conic or linear, so the problem can be solved directly by existing solvers such as CPLEX.

2.4.3 Algorithm framework and feasible solutions

Solutions to the relaxed sub-problems can be used to generate feasible solutions to (16) which provide upper bounds to the optimal objective values of AM model (18). The general algorithm framework for the AM model is as follows and similar procedure applies to the LB and UB models..

Step 1: Initialize $\boldsymbol{\phi}^0$, $n = 0$.

Step 2: In iteration *n*, solve Sub-problem 1 (e.g., by NLP solvers) and Sub-problem 2 (e.g., by CPLEX) under $\boldsymbol{\phi}^n$ and obtain the corresponding solutions $(\mathbf{x}, \mathbf{y})^{sub1}$ and $(\mathbf{z}, \mathbf{f}, \mathbf{k})^{sub2}$.

Step 3: Find feasible solutions to the original problem and update upper bound ($\hat{U}$).

Step 3.1: Fix the values of location decision variables *y* and allocation decision variables *x* from the relaxed problem, and solve the rest variables in problem (18); i.e., set $(\mathbf{x}, \mathbf{y}) = (\mathbf{x}, \mathbf{y})^{sub1}$ in the original problem and solve for $(\mathbf{z}, \mathbf{f}, \mathbf{k})$.

Step 3.2: Fix the values of continuous variables from Sub-problem 2, i.e., set $(\mathbf{z}, \mathbf{f}, \mathbf{k}) = (\mathbf{z}, \mathbf{f}, \mathbf{k})^{sub2}$, for $\forall i \in \mathrm{I}$, sort $\{k_{i,j}\}$ in a descending order and let l_{ir} record the corresponding r^{th} assigned facility index for customer group *i*. Set $x_{ijr} = \begin{cases} 1, & \text{if } j = l_{ir}, \forall i \in \mathrm{I}, \forall r = 1, 2, \ldots, R, \\ 0, & \text{otherwise.} \end{cases}$, and $y_j = \begin{cases} 1, & \text{if } \sum_{i \in \mathrm{I}} \sum_{r=1}^{R} x_{ijr} > 0, \forall j \in \mathrm{J} \\ 0, & \text{otherwise.} \end{cases}$. Then, solve the problem (18) by fixing $(\mathbf{x}, \mathbf{y})$ to obtain feasible values $(\mathbf{z}, \mathbf{f}, \mathbf{k})$.

Step 3.3: Compare the feasible solutions obtained from Step 3.1 and Step 3.2 with the current best feasible solutions, and update the best upper bound if a better solution is found.

Step 4: Compare the solution of the relaxed problem with the best lower bound ($\hat{L}$), and update the lower bound if a larger bound is found. Terminate the algorithm if the optimality gap is smaller than a specified tolerance or the computation time exceeds the limit, otherwise update the multipliers $\boldsymbol{\phi}^{n+1}$ (see section 4.4) and set $n = n + 1$, and go to step 2.

2.4.4 Multiplier updates

Through the framework of LR algorithm, dual multipliers ϕ can be updated iteratively to find the best lower bound and search for the optimal value of problem (18). Initial values of multipliers are set as $\phi^0 = \{\phi_{ij}^0 = 0\}$. Then in iteration n, the multiplier can be updated by applying the conventional sub-gradient method (Fisher, 1981) for $\forall i \in \mathrm{I}, j \in \mathrm{J}$ as follows:

$$\phi_{ij}^{n+1} = \phi_{ij}^{n} + \varphi^{n}(k_{i,j} - \sum_{r=1}^{R} \omega_i (1-q) q^{r-1} x_{ijr}).$$

The step size φ^{n+1} in the nth iteration can be updated as

$$\varphi^{n+1} = \frac{\tau^{n}(\hat{U} - \Phi(\phi))}{\sum_{i \in I} \sum_{j \in J} (k_{i,j} - \sum_{r=1}^{R} \omega_i (1-q) q^{r-1} x_{ijr})^2},$$

where $\hat{U}$ is the current best upper bound. If $\hat{U}$ does not improve in K predefined consecutive iterations, we set

$$\varphi^{n+1} = \frac{\tau^{n} * \text{abs}(\Phi(\phi)) * 0.4}{\sum_{i \in I} \sum_{j \in J} (k_{i,j} - \sum_{r=1}^{R} \omega_i (1-q) q^{r-1} x_{ijr})^2}, \quad (15)$$

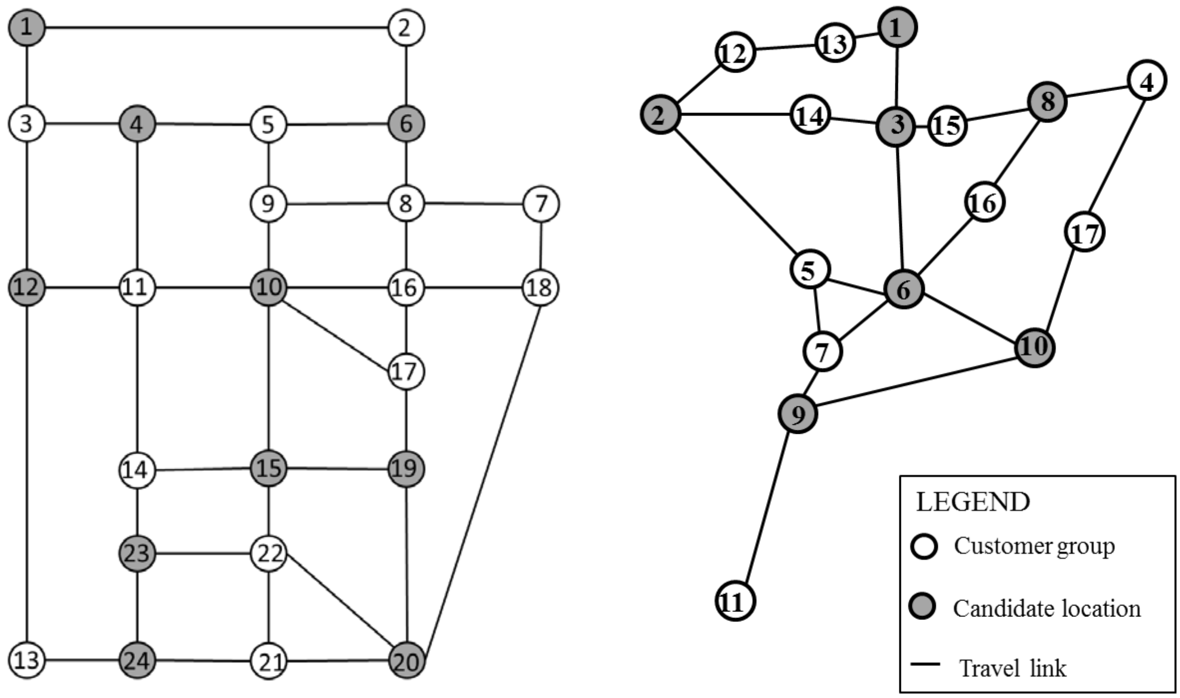
where τ^n is a step size parameter. If the lower bound does not improve in K predefined consecutive iterations, we set $\tau^n = \frac{\tau^0}{n^{0.5}}$.

2.5 Numerical Example

In this section, a set of numerical experiments are conducted under different parameter settings to measure the gaps between the upper and lower bounds of the scenario-based reliable service facility location model. The computational efficiency of the proposed LR based solution algorithm is also tested through these experiments to numerically illustrate the accuracy of the proposed approximation model compared to the original scenario-based model. Besides, the approximation model with and without considering in-facility delay are compared to show the effect of in-facility congestion on facility location design and the total system cost. Insights are drawn from a series of sensitivity analyses with regard to a range of key parameters values. The proposed models and algorithm are coded in GAMS scripts, and run on a personal computer with 2.67 GHz CPU and 4 GB of RAM. We use off-the-shelf solvers KNITRO and CPLEX as benchmark solution methods, and use them to solve the LR sub-problems (or their reformulations).

2.5.1 Computational experiments

The proposed LB model, UB model and AM model are applied first to the Sioux-Falls network (Bar-Gera, 2009), as shown in Figure 1(a). In this example, candidate facility nodes are shown by dark shades and others are customer demand nodes. The majority of parameter values are taken or derived from Castillo et al. (2009), Bai et al. (2011) and An et al. (2013). The hourly demand rate at each demand node (ω_i) is generated randomly in uniform distribution between [1800, 3000]; the prorated hourly fixed cost of per server at candidate locations (m_j) is randomly drawn from interval [20, 50]; the unit penalty cost for each customer, η = $10000; and a series of sensitivity analyses are conducted for parameters shown in Table 1, including the unit value of travel time α, unit value of in-facility waiting time β, average service rate μ_0, service disruption probability q, and prorated hourly facility setup investment d_j. The following parameter values related to the LR algorithm are used in all experiments. The initial values of Lagrangian multipliers and step size parameter are ${\phi^0}_{ij} = 0, \forall i \in \mathrm{I}, j \in \mathrm{J}$ and τ^0 = 2; the maximum iteration numbers for all runs is 100; the optimal gap tolerance is 0.1%; and the computation time limits for all sub-problems for the LR procedure and directly using the solver are 3000s and 6000s respectively, and the allowed number of consecutive iterations when $\hat{U}$ does not improve is K=3.



(A) (B)

Figure 1 (A) The Sioux-Falls network. (B) 17-node hypothetical network.

Table 1 presents the computational results for a range of problem instances for the Sioux-Falls network. It can be observed that the gaps between lower bounds and upper bounds are mostly less than 10%, indicating that the difference between the proposed approximation model and the original scenario-based reliable facility location model should be very small. Also, in most tested parameter settings (only except for case (a) and (e)), the total number of optimal facilities and their locations from the UB, AM and LB models are precisely the same, which means that our proposed approximation model indeed provides good estimations of the true optimal reliable service facility location design.

Table 1. Comparison of UB, LB and AM models

Case	α ($/h)	β ($/h)	q (%)	μ_0 (cus./h)	d_j ($/h)	Model	CPU time	Optimality gap	Total num.	Optimal locations	Total cost ($\times 10^5$ $)	Gap between

							(s)	(%)				UB and LB (%)
(a)	1	15	5	12	20000	UB	1531	0.2	4	6,12,15,19	1.79	7.8
						AM	3611	0.09	4	1,6,12,15,	1.66	
						LB	1415	0.2	4	6,12,15, 24	1.65	
(b)	1	15	5	120	5000	UB	2052	1.0	5	6,10,12,15,19	0.52	17.7
						AM	3604	0.3	5	6,10,12,15,19	0.44	
						LB	1618	0.3	5	6,10,12,15,19	0.43	
(c)	1	15	5	120	20000	UB	1338	0.5	4	6,12,15,19	1.24	8.1
						AM	4312	0.09	4	6,12,15,19	1.15	
						LB	648	0.09	4	6,12,15,19	1.14	
(d)	1	5	5	120	20000	UB	1745	0.5	4	6,12,15,19	1.23	7.5
						AM	756	0.07	4	6,12,15,19	1.14	
						LB	803	0.07	4	6,12,15,19	1.14	
(e)	1	15	10	12	20000	UB	1606	0.2	6	1,4,6,12, 15,24	2.05	6.8
						AM	1437	0.2	6	1,6,12, 15,19,20	1.92	
						LB	1626	0.2	6	1,6,12, 15,19,20	1.91	
(f)	1	15	10	120	20000	UB	1525	0.3	6	6,10,12, 15,19,24	1.49	6.1
						AM	4247	0.1	6	6,10,12 15,19,24	1.41	
						LB	1576	0.1	6	6,10,12 15,19,24	1.40	

We also apply the proposed approximation model to both a 17-node hypothetical network from Hajibabai et al. (2013), as shown in Fig. 1(b) and the Sioux-Falls network to further test the impact of in-facility congestion on facility location design and the computational performance of the proposed LR based solution algorithm with α =10 \$/h, β =15 \$/h, μ_0 =12 customers/h, q =0.05, and d_j =5000 \$/h. For each network, we consider two optimization objectives (with or without considering in-facility delay) and two solution methods (solver KNITRO or LR based algorithm). The total system costs of all solutions are evaluated using the approximated total cost formula (i.e., the objective function (18)). The computational results are summarized in Table 2. For all of the test cases, the LR based algorithm efficiently converges to near-optimum solutions with a reasonably small gap (i.e., less than 1.5%). In contrast, the solver finds optimal solutions only for the case of the 17-node network without considering in-facility delay, which takes much longer CPU time than the LR algorithm; the solver fails to find any feasible solutions for all other cases within 6000 second time limit and 1% gap tolerance. As such, the proposed LR based algorithm not only improves computational efficiency by providing near-optimum solutions within a shorter computation time, but also can be applied to relatively large-scale instances.

Table 2. Comparison of solver KNITRO and LR based algorithm on two objectives

Solution Approach	Network	Cases	CPU time (s)	Gap (%)	Total num.	Optimal locations	Cost components (%)					Total cost ($\times 10^5$ $)
							Travel cost	Waiting cost	Service cost	Setup cost	Penalty cost	
Solver KNITRO	17-node network	Considering in-facility delay	6000	-	-	-	-	-	-	-	-	-
		No in-facility delay	397	-	5	3,6,8,9,10	53..5	1.94	29.68	14.62	0.27	1.71
	Sioux-Falls network	Considering in-facility delay	6000	-	-	-	-	-	-	-	-	-
		No in-facility delay	6000	-	-	-	-	-	-	-	-	-
LR based Algorithm	17-node network	Considering in-facility delay	986	0.6	5	3,6,8,9,10	54.68	1.9	28.35	14.9	0.28	1.69
		No in-facility delay	797	0.9	5	3,6,8,9,10	53.58	1.92	29.6	14.62	0.27	1.71
	Sioux-Falls network	Considering in-facility delay	1565	0.6	5	6,10,12,15,19	53.19	2.19	32.02	12.2	0.41	2.05
		No in-facility delay	1774	1.2	5	4,6,10,15,24	46.31	2.34	39.53	11.44	0.38	2.19

We can also see that failure to consider in-facility delay leads to different facility location designs in the Sioux Falls network and higher total system cost. When the objective aims only at minimizing the travel cost and facility setup cost, i.e., without considering in-facility delay, more dispersedly distributed service facilities may be selected (e.g., the spatial distribution of 4, 6, 10, 15, 24 is more scattered than that of 6, 10, 12, 15, 19 when considering in-facility delay in optimal facility location design). In this case, travel cost is more dominant so customers are assigned to more decentralized facilities to reduce customers' average travel distance and mitigate traffic congestion. On the other hand, the in-facility waiting and service costs and thus the total system cost may increase greatly. This observation implies that ignoring the in-facility delay in service facility location design may result in a quite suboptimal location deployment as well as erroneous system cost estimations. The in-facility waiting and service costs increases with more decentralized facilities due to a "risk pooling effect", and we will explain in more details in the next section.

2.5.2 Sensitivity analysis with AM model on the Sioux-Falls network

Numerical experiments in the previous section implies that the values of average service rate μ_0, prorated facility setup investment d_j, service disruption probability q, and time value α and β have major impacts on the optimal solutions. We extract more results from sensitivity analyses with the Sioux-Falls network to provide more insights of these parameters. A total of 18 cases of the AM model are solved by the LR based algorithm: 3 levels of prorated hourly facility setup investment d_j = {\$2000, \$5000, \$20000}, $\forall j \in J$, 2 levels of average service rate μ_0 = {12, 120} customers/hour, 4 levels of service disruption probability q = {0, 0.05, 0.1, 0.3}, and 2 different α / β ratios {10/15, 15/10}. Table 3 summarizes the results.

Table 3. Sensitivity analysis on the Sioux-Falls network

Case	α (\$/h)	β (\$/h)	q (%)	μ_0 (cus./h)	d_j (\$/h)	CPU time (s)	Gap (%)	Total num.	Optimal locations	Cost components (%)					Total cost (×10^5 \$)
										Travel cost	Waiting cost	Service cost	Setup cost	Penalty cost	
1	10	15	5	12	2000	1910	0.7	6	6,10,12,15,19,24	56.37	1.27	35.86	6.48	0.02	1.85
2	10	15	5	12	5000	4565	0.7	5	6,10,12,15,19	53.20	2.19	32.02	12.18	0.41	2.05
3	10	15	5	12	20000	1894	0.3	4	6,12,15,19	41.90	0.67	22.04	29.28	6.11	2.73
4	10	15	5	120	2000	2352	0.8	7	4,6,10,12,15,19,24	79.25	1.33	7.78	11.64	0.00	1.20
5	10	15	5	120	5000	2513	1.5	5	4,6,10,15,24	73.11	1.17	7.05	18.07	0.60	1.38
6	10	15	5	120	20000	1467	1	4	4,6,10,15	50.14	0.30	4.00	37.68	7.86	2.12
7	15	10	5	12	20000	1664	0.7	5	6,10,12,	48.38	0.53	19.63	30.05	0.25	3.33

									15,19						
8	15	10	5	120	20000	1774	0.8	5	4,6,10, 15,24	57.64	0.50	3.58	37.96	0.32	2.63
9	10	15	0	12	5000	2198	0.2	3	6,12,15	60.50	1.73	30.15	7.62	0.00	1.97
10	10	15	10	12	20000	1808	0.1	6	6,10,12, 15,19,24	35.14	1.66	22.71	39.61	0.88	3.03
11	10	15	20	12	20000	12274	0.8	8	1,4,6,10,12 ,15,19,24	31.06	0.77	19.85	46.34	1.98	3.45
12	10	15	0	120	5000	2473	2.4	4	4,6,10,15	77.72	0.93	6.74	14.60	0.00	1.37
13	10	15	10	120	20000	1721	0.8	6	4,6,10, 12,15,19	42.58	0.75	4.08	51.45	1.15	2.33

We see that most cases are solved to near optimality (less than 1% gap) in less than an hour. In the first 6 cases, as d_j increases, fewer service facilities are built, which saves facility setup cost but also results in longer travel time and severer traffic congestion. An interesting finding is that deploying fewer service facilities is sometimes beneficial in terms of decreasing the total in-facility waiting and service cost. This is consistent with the well-known "resource pooling effect" for general queuing systems, as stated in Theorem 1 of Benjaafar (1995).

The "resource pooling effect" also subverts our intuition when the average service rate becomes very small. For example, when μ_0 changes from 120 (i.e., serving 2 customers per minute) to 12 (i.e., serving 0.2 customers per minute), it is intuitive to select more facilities to leverage customer's seemingly increased waiting time, however, the system tends to select fewer facilities (with higher service capacity in each), e.g., by comparing cases 1 and 4, or facility locations tend to be much closer, e.g., by comparing cases 2 and 5, 3 and 6. Although this arrangement may slightly increase the customers' travel cost, due to the pooling effect, customers suffer less waiting due to more pooled servers (as evidenced by the decrease of average waiting and service time in the facilities).

We find that the service disruption probability also has a significant impact on the system design. For example, by comparing cases 9 and 2, 3, 10 and 11, 12 and 5, 6 and 13, we find that the total service facility number greatly increases with q, suggesting that more facilities are needed to provide sufficient back-up options when facilities become less reliable. The increase of q also leads to the lower of total expected number of customers gathered in each service facility and this mainly accounts for the rise of customer waiting and service costs due to the risk pooling effect. As every customer has a greater probability to resort to backup service facilities or lose service, the expected customer travel cost and system penalty cost both increase. The sharp increase of penalty cost is also highlighting a real concern over significant loss of life and property in extreme disaster events that lead to high disruption probability.

To study the impact of en-route travel and in-facility waiting delay, we vary the weights of travel and waiting costs, i.e., α and β. Through comparing cases 3 and 7, 6 and 8, we observe that the optimal number of service facilities is greater when travel cost has larger weights (e.g., cases 7 and 8). This indicates that when en-route traveling is dominating, more facilities shall be deployed to reduce the average customer travel distance. On the other hand, when in-facility waiting is dominating (e.g., cases 3 and 6), customer demand tends to be pooled to fewer facilities to reduce average waiting and service time (as well as the total number of servers).

Figure 2 shows the optimal facility deployments, the corresponding server numbers and level-1 customer assignments for six cases. We consider (I) case 3 in Table 2 as the benchmark; (II) case 1, to show the impact of prorated facility setup investment d_j; (III) case 6, to show the impact of average service rate μ_0 in the system; (IV) case 7, to show the impact of different α / β ratios; (V) case 11, to show the impact of service disruption probability; and (VI) without considering in-facility delay.

We see that in each scenario the total number of servers is around 2250 across all facilities when μ_0 =12 customers/hour. This number seems a little large, which mainly results from (i) the demand rate in each customer group is large; and (ii) the in-facility waiting cost is reduced at the expense of providing more servers (as stated in the derivation of the optimal service rate μ_j).

In Figure 2, the benchmark scenario (I) has four facilities built at nodes 6, 12, 15 and 19. With the optimal level-1 customer assignments and server arrangements, the average waiting and service time in each service facility is around 16.4s, the unit service cost is around $26.2, and the total server number is 2294. When in-facility delay is not considered in the objective, e.g., as in scenario (VI), there is a 13.3% increase in the average waiting and service time and 39.6% in the average service cost in each service facility as compared with scenario (I) due to the more scattered distribution of optimal facility locations, which also roughly explains the rise of total expected system cost in the facility location design, even though the total travel cost drops by 7.61%. When en-route travel has a larger weight, e.g., as in scenario (IV), and when service facilities become less reliable, e.g. q changing from 0.05 to 0.2 as in scenario (V), the resource pooling effect dominantly accounts for a similar rising trend of average waiting and service time and unit service cost.

Decreasing the prorated facility setup cost d_j as in scenario (II) and increasing the average service rate μ_0 as in scenario (III) are both beneficial to lower the total expected system cost, however, if putting emphasis on reducing the en-route travel, in-facility waiting and penalty for potentially losing service related costs, the latter is undoubtedly a better measure as evidenced by the sharp decrease of these three cost components from \$1.93×105 in benchmark scenario (I) to \$1.32×105 in scenario (III), compared to the amount of \$1.73×105 in scenario (II).

Another interesting finding is that as traffic congestion and in-facility waiting delay are integrated into the objective, a customer may no longer be assigned to its nearest functioning service facility. For example in scenario (III), node 6 is much closer to nodes 7, 16 and 18, however, all these three customer groups are assigned to node 4 as their level-1 service facility. This is not surprising, as distance no longer plays a significant role in determining costs.

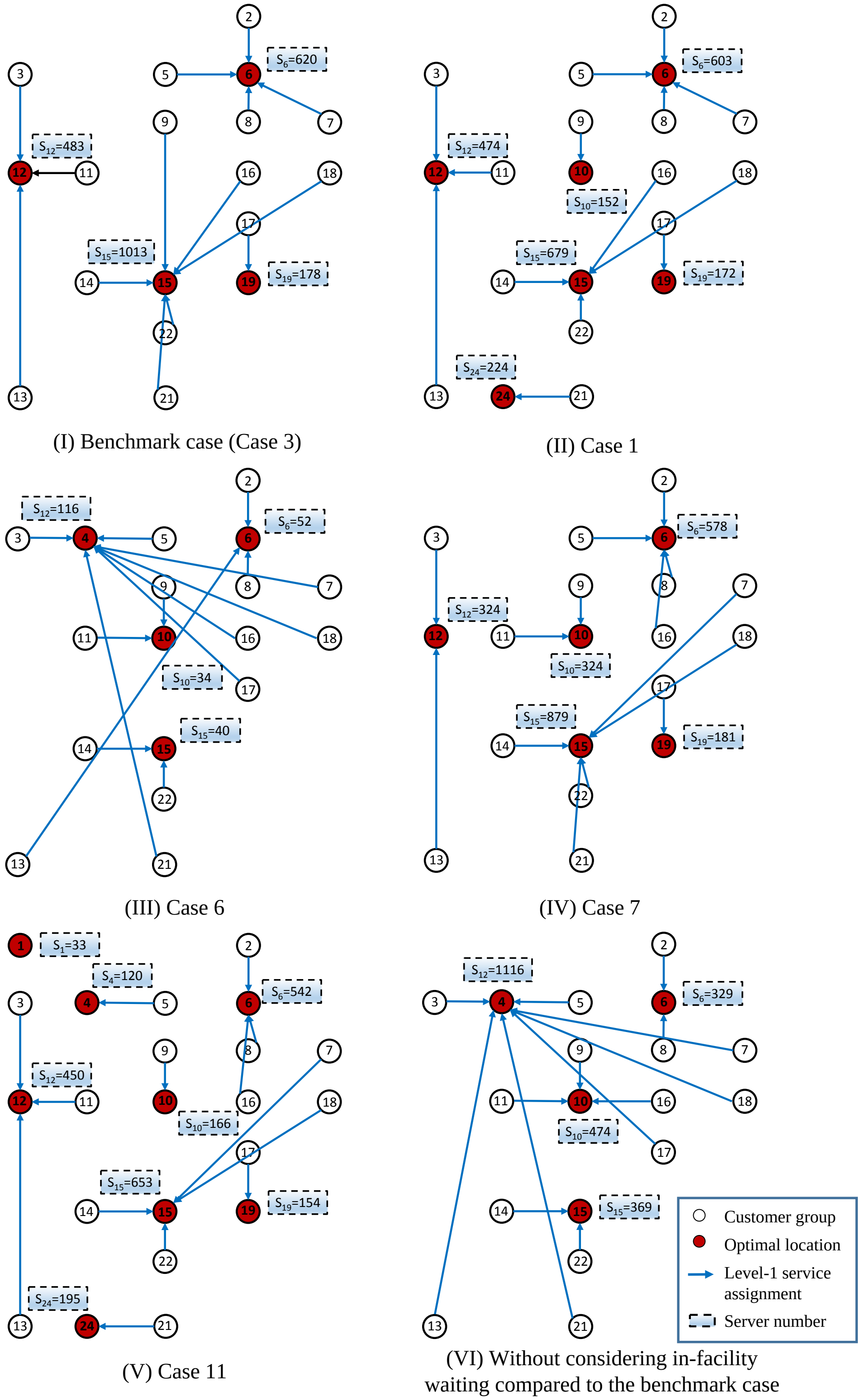


Figure 2. Facility location, leverl-1 assignment and service capacity for the Sioux-Falls network.

2.6 Conclusion

In this study, traffic congestion, in-facility waiting delay and service disruption uncertainty were incorporated into a stochastic facility location model. It determines the optimal facility location, capacity allocation, customer-to-facility assignment, traffic network routing, and in-facility system operation strategies across all normal and possible service disruption scenarios to minimize the overall expected system cost. An approximate MINLP model was proposed to reduce the computational complexity associated with the original model, and an LR based solution approach was developed to solve the approximation model. A conic program transformation method was used to further simplify and solve one of the hard sub-problems (for facility location and service allocation decisions). We further developed lower and upper bounds for the optimal objectives of both the approximation model and the original stochastic problem. Numerical experiments shows that the proposed approximation model indeed provides a good approximation of the optimal reliable service facility location design, and the LR based algorithm is shown to obtain near-optimum solutions within a relative short time. Some useful managerial is also drawn through a series of sensitivity analyses.

Future research can be conducted in a number of directions. This project makes several simplifying assumptions and approximations, e.g., identical and independent facility disruptions and M/M/1 queuing systems. It is worthwhile to explore ways to relax these assumptions in the future. Furthermore, our work can be potentially extended to large-scale instances, so it might be appealing to develop more efficient algorithm or explore alternative models with better computational scalability.

CHAPTER 3. LOCATING FACILITIES IN DISASTERS WITH CONSIDERATION OF SOCIAL COSTS

3.1 Background

There is a significant amount of research in the area of modeling emergency operations for ambulance location, ambulance redeployment, hospitals and fire stations. However, humanitarian supply chain, as a research area, is relatively new (Sheu 2007). One category explores the dynamics of geographical facility location with respect to other factors such as cost, and service. It uses facility location models to identify optimal locations for pre-positioning of supplies, shelters, and other critical resources. This problem has mostly been addressed as a discrete facility location problem or as assignment problem. Under this perspective, Akkihal (Akkihal 2006) has addressed the problem of locating worldwide distribution centers using an algorithm that solves the uncapacitated facility location problem (UFLP). This is one of the initial works in addressing the problem using a demand function based on population needs. Given that this model uses a discrete UFLP each demand point is an aggregation of the demand in the area.

In the same line of locating worldwide distribution centers, Balcik and Beamon (Balcik and Beamon 2008) propose a model for the pre-positioning problem. They assume that the supplies are pre-positioned in a distribution center. The model is a modified version of a traditional location problem called the maximal covering location problem (see Daskin 1995 and the references therein) that includes an inventory model at each distribution center. The demands are based on scenarios, where each scenario has a probability of occurrence. The response time is addressed using a level of service, which is defined by lower and upper limits in response time.

More specifically, depending on the distance, there is a coverage level which depends on the scenario. In a more regional and local perspective, Ukkusuri and Yushimito (Ukkusuri and Yushimito 2008) have addressed the location problem that includes routing decisions. Since routes can be disrupted, they provide a methodology to locate the warehouses in places that assure that the most reliable path between the demand point and the pre-positioning places are chosen. Rawls and Turnquist (Rawls and Turnquist 2009) developed an emergency response planning tool that determines the location and quantities of various types of emergency supplies using a two-stage stochastic mixed integer program (SMIP) taking into consideration the transportation network availability after an event.

A few authors have studied problems related to distributing the critical supplies as well. Tzeng et al. (Tzeng et al 2007) use an approach where commodities need to be collected at distribution centers and delivered to the demand points. The model includes transfer points, which can also be demand points. Horner and Downs (Horner and Downs 2007) developed a model linked to a Geographic Information System (GIS) that identifies accessible locations in which to place intra-urban uncapacitated relief goods distribution sites. The work by Maliszewski and Horner (2010) provides a similar GIS framework for siting critical supply chain infrastructure. Building upon this model the authors (Horner and Downs 2008) later proposed a spatial model for hurricane relief goods distribution, suggesting the location of intermediate distribution stations using a capacitated facility location problem in which they seek to minimize the transportation costs. The following section seeks to provide another perspective in the location problem for disasters.

3. 2 Some Theoretical Aspects of Location for Disasters

One of the early stages of disaster preparedness involves the selection of facilities from a specified set of locations or finding the strategic locations where distribution centers or facilities could be built, that would be used to provide relief in case of a disaster. In simplified geographical terms, the possible locations could be represented by a vector x_i with coordinates (x_{i1}, x_{i2}) in a two dimensional cartesian space S. Our main objective is the coverage of a region satisfying certain criteria that are unique in disaster relief. As mentioned in Balcik and Beamon (Balcik and Beamon 2008), models with coverage objectives are the most appropriate ones when response time is the primary objective. In these models, coverage is defined as the ability to reach a demand point from a facility within a specific response time. This section explains these criteria and how it is related to the Voronoi diagram. These concepts are important because they are the basis for the development of the heuristic in Section 5.

Location theory states that if facilities can provide the same utility with a transportation cost function $t(d_i)$ defined in terms of d_i which is a measure of distance between facility x_i to a demand point x, for instance $d_i = \|x_i - x\| = \left|(x_i - x)^T (x_i - x)\right|^{1/2}$ (Euclidean distance). Then each demand point will use the facility $i \in I$ if it maximizes his/her total utility $u_i - t(d_i)$, that is, whenever $\max_{i \in I} u_i - t(d_i)$. If all facilities provide the same utility, then the previous maximization problem is equivalent to $\min_{i \in I} t(d_i)$. In the discrete version, this is equivalent to minimize the total transportation cost from the node selected as facility to the demand point. Most optimization models consider this. In the continuous case, the problem is more complicated because there is not a finite number of locations. As noted by Akkihal (Akkihal 2006) the treatment of continuous demand as an area or surface might approximate better human settlements, making it more appropriate than discrete models. In this case, given that the demand is assumed to be distributed over an area, the problem tries to divide the region such that the demand is assigned to a region enclosed around the selected facility. For instance, if the distance that we seek to optimize is the average distance, the problem falls under the class of continuous *p-median* problem. As presented by Okabe and Suzuki (1997), this is a class of location problems that can be solved through Voronoi diagrams. Therefore, the partition can be assumed as follows: the total region covered by facility i is given by $V(x_i) = \{x : \|x_i - x\| \leq \|x_j - x\|, j \in I\}$ which is nothing but the Voronoi Diagram[1] or Thiesen Polygon.

Moreover, the socially optimal configuration definition is the one in which the total transportation cost of delivering goods to all customers in the region is minimum. Mathematically, if *F*(*X)* represents the total cost, provided that the facilities are located at *X,* it can be expressed by

$$F(X) = \sum_{i=1}^{n} \int_{x \in V(x_i)} \|x_i - x\| dx \tag{1}$$

and the social optimal configuration is given by

$$F(X^*) = \min_{X \in \chi} F(X) \tag{2}$$

[1]In mathematics, a Voronoi diagram is a special kind of partitioning a space determined by distances to a specific set of points in the same space. In its simplest case, we are given a set of points ***S*** in the plane, which are the Voronoi sites. Each site has a Voronoi cell consisting of all points closer to the Voronoi site *j* than to any other site (see Aurenhammer (*1991*) for a complete discussion).

where χ is the set of all possible locations X on region S.

The previous model falls in the category of geographical optimization problems for commercial supply chains. In the traditional approach, the focus is on maximizing efficiency measures such as delivery cost. However, as mentioned in Section 1, Holguín-Veras et al. (Holguin-Veras et al. 2006) and Beamon and Balcik (Beamon and Balcik 2005) have recognized, the particular characteristics and challenges of humanitarian logistics. In particular Holguin-Veras et al. (Holguin-Veras et al. 2010) have addressed the need to incorporate the social costs of delivering goods to areas affected by large disasters. Rather than defining operational success in terms of travel time, total cost, or even cargo delivered, for humanitarian supply chains, they propose that the primary objective is the minimization of human suffering. These social benefits (or its negative, the social costs) may be estimated with valuation techniques from theoretical economics, such as stated preference experiments, in which participants state how much they would pay for a particular good if they have not had it for a specified period of time. The valuations from a reasonable responder would be sensitive to the type and time interval without the commodity, thus providing decision makers a common economic measure to evaluate distribution alternatives. Based on these assumptions they propose a deprivation cost function of the form:

$$f(\Delta t) = Population \times e^{(\alpha + \beta \Delta t)} \quad (3)$$

where Δt is the time (deprivation time) since the last delivery was received at the node, e is the mathematical constant for the inverse natural logarithm and α, β are parameters calibrated using econometric techniques.

The population can be replaced by a demand density function $D(x)$ of the area affected, and by assuming that greater distances will imply larger deprivation times (due to cycle times) (Holguin-Veras et al. 2010), this difference in time can be put in terms of distance d using a function of urgency that increases with distance, provided that an increase in distance represents an increase in time. This issue has also been discussed by Akkihal, 2007: "…the critical factor, outbound transportation time, is minimized as the stocks are positioned closer to the demand point…", that is inventory positioned near demand points yields to lower lead times. In other words, we obtain a function in terms of distance $f(d)$ and each facility i that is attempting to minimize total suffering would like to $\min_{i \in I} f(d_i)$. As a consequence, the resulting problem that should be solved in order to attain the optimal configuration would be

$$F(X^*) = \min_{X \in \chi} F(X) \tag{4}$$

Where

$$F(X) = \sum_{i=1}^{n} \int_{x \in V(x_i)} D(x) e^{(\alpha + \beta f(d))} dx \tag{5}$$

The next section provides a more formal definition of the problem. We present an approach for the discrete version of this problem that is able to solve a large number of demand points distributed over an area, which can approximate a continuous demand function.

3.3 Problem Statement

The problem can be stated as follows: we focus on identifying a strategic set of locations for a pre-specified number *n* of facilities to be sited in a continuous, two-dimensional region such that (1) all demand points are covered while (2) minimizing the urgency social function shown in Equation 3. Given the limited information available regarding demand distributions for this type of problems, we present the discrete version of the problem. That is, we aggregate the demand into discrete points in the two-dimensional region. However, a continuous demand function can be approximated using a sufficient number of discrete demand points. Another assumption implies that given that we are looking for the strategic locations where facilities are going to be sited, there are no restrictions on the location of the facilities and their capacities.

Mathematically, this problem (P1) can be expressed by $\min F(X)$, where the objective function can be written as:

P1 $$F(X) = \sum_{i=1}^{n} \sum_{j=1}^{m} p_j e^{f_j(d_j)} \qquad (6)$$

and

- I is the set of facilities (1…*n*)
- J is the set of demand points (1…*m*)
- d_{ij} is the maximum distance between the demand area *j* and the prospective facility *I*. For example it can be computed as $\|x_j - x_i\|$ or any other type of distance measure (network distance, Hamiltonian distance.)
- f_j is a function of urgency that depends on distance. For example: $f_j = \alpha + \beta d_{ij}$,

The next section proposes a solution heuristic to find the locations, solving the proposed minimization problem. The reason for using the heuristic is because of the non-convex nature of problem P1. The heuristic is an iterative process that can only guarantee local minima which depends largely on the initial solution. We also address the issue of selecting an initial set of locations later on in Section 6.

3.4 Solution Method

Given that, to approximate a continuous demand function with discrete demand points a significantly large number of demand points is required, and that the problem has a non-convex objective function we propose a heuristic to solve the problem presented in Section 4 bearing a tradeoff between running time and optimality. Additional assumptions for this heuristic are:

- The region S under analysis is a convex polygon in the plane.
- The facilities have infinite capacities and have no probability of failure.
- The urgency function is a direct function of distance. In this case, it is assumed as a function of the Euclidean distance $f(d) = \alpha + \beta \left\| x_j - x_i \right\|$. However, as it was also explained earlier it can be any other type of distance measure such as network distance.

Heuristic Approach for the Discrete Case

In this section we present a solution heuristic based on Voronoi diagrams. The heuristic procedure has to find the best Voronoi diagram following iteratively the 5 steps discussed below. It follows the intuition of the heuristic developed by Suzuki and Drezner (Suzuki and Drezner 1996) with the difference that we solve a sub-problem to improve the location at each iteration. Moreover, we use a method that does not require the computation of derivatives which allows speeding up the heuristic (see Section 5.2 for the solution of the sub-problem):

- *Step* 1 Choose an initial set of *n* random points in the plane and set $Z^{best} = \infty$. The strategy for selecting the initial points is addressed in Section 6.
- *Step* 2 Construct a Voronoi diagram using each random point as a Voronoi site *i* (our initial facility location).
- *Step* 3 For each Voronoi cell $V(x_i)$ assign all demand points *k* located in the Voronoi cell $V(x_i)$ and solve the following sub-problem SP1:

$$\min \sum_{j=1}^{k} \left(p_j e^{\left(\alpha + \beta \left\| x_j - x_i \right\|\right)} \right) \tag{7}$$

s.t.

$$x_i \in V(x_i)$$

Notice that we are solving a relaxation of the problem (P1) for all the points assigned to $V(x_i)$.

- *Step* 4 Compute the objective function (social optimal configuration) replacing the values of the distances in Equation 6. That is, at iteration *t we* compute $Z^t = \sum_{i=1}^{n} \sum_{j=1}^{m} \left(p_j e^{\left(\alpha + \beta \left\| x_j - x_i \right\|\right)} \right)$
- *Step* 5 If the following convergence criteria is satisfied $\max \left| Z^t - Z^{best} \right| < \varepsilon$, then stop. Otherwise, if $Z^t < Z^{best}$ make $Z^{best} = Z^t$ and go to *Step* 2. If $Z^t > Z^{best}$ go to *Step* 2.

Notice that the heuristic first constructs a Voronoi diagram using an initial set of random points. Then, instead of computing the center of the polygon by allocating the demands in the vertices of the Voronoi cells as in Suzuki and Drezner (Suzuki and Drezner 1996) for the continuous case, we solve the sub-problem (SP1) in Step 4 (see Figure 2) by using the demand points and finding the p-median solution of this sub-problem.

Then, we re-compute the objective function (see Equation 6). The process is repeated until a convergence criterion is satisfied. The convergence criterion is based on either the maximum number of iterations (for large problems) or the objective function does not change by a pre-defined factor ε.

Figure 2 represents the process of improving at each iteration for $n = 5$ (number of facilities) and $n=50$ (number of demand points) in planar region [0,1]. The initial 5 locations of the Voronoi sites were chosen randomly. The process starts by setting the initial location of the facilities (red squares) and constructing the Voronoi diagram. Then using the demand points that belong to this Voronoi cell we solve the problem and find the new Voronoi site (blue star) and proceed as indicated in Steps 4 and 5. The next iteration starts with constructing the Voronoi cells using the solution from the previous iteration.

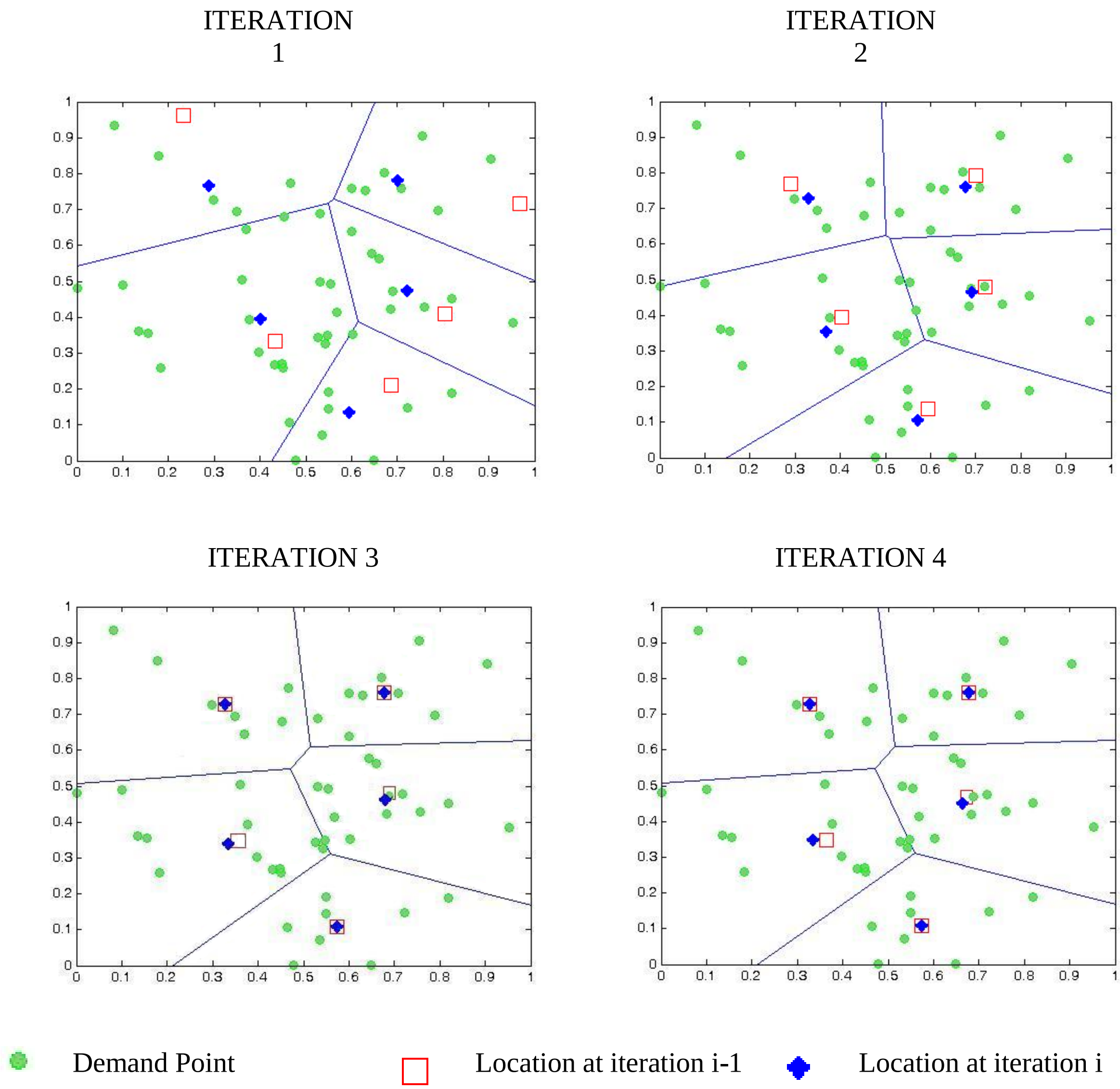


Figure 1. Example of first 4 iterations of the heuristic for a problem with facilities (*m*) = 10, demand points (*n*)= 100

Solution of the Sub-problems

As it can be observed in previous section, the heuristic requires the solution of a nonlinear optimization sub-problem at Step 3. In this subsection we discuss a solution approach for both sub-problems. Both sub-problems are non-linear unconstrained optimization problems of the form $\min_{x \in V(x_i)} F_i(x)$, that is, it is only unconstrained in the region defined by the Voronoi set $V(x_i)$. This complicates the problem in the sense that the problem becomes a constrained non-linear optimization problem. Solving this type of problem is usually computationally expensive. We propose a Nelder-Mead based solution of the sub-problem (Nocedal and Wright 2005).

The Nelder-Mead simplex-reflection method is a popular Derivative Free Optimization method that searches for an improving solution in the convex hull formed by n+1 points of interest in $\Re^n$ (see Figure 3 for the main steps pg the method). This convex hull should form a non-degenerate simplex[2] S with vertices $\{y_1, y_2, ..., y_{n+1}\}$. A single iteration implies the evaluation of these vertices and removes the vertex with the worst function value. This value is replaced by another point with a better value by reflecting, expanding or contracting the simplex along the line joining the worst vertex with the centroid of the remaining vertices. For instance, suppose that we want to minimize a function $f(.)$ and that the initial vertices of the simplex are $\{y_1, y_2, y_3\}$ with $f(y_1) \le f(y_2) \le f(y_3)$. If the centroid is denoted by $\bar{y}$, then the line joining the centroid with the worst vertex is denoted by

$$\bar{y}(t) = \bar{y} + t(y_3 - \bar{y}) \tag{8}$$

The value of the scalar t allows us to get the reflection point and its expansion or contraction. By selecting an appropriate initial simplex we can maintain the search inside of our current Voronoi set. In Section 7 we show that this heuristic provides good solutions when we compare the results with the optimal value.

[2] A simplex is not degenerate if the matrix Q, formed with the distances of the n edges from one of its vertices, is not a singular.

Compute the reflection point $\bar{y}(-1)$ and evaluate $f_{-1} = f(\bar{y}(-1))$
if $f(y_1) \leq f_{-1} < f(y_n)$
  replace y_{n+1} by $\bar{y}(-1)$ and go to next iteration
else if $f_{-1} < f(y_1)$
    compute $\bar{y}(-2)$ and evaluate $f_{-1} = f(\bar{y}(-2))$
    if $f_{-2} < f_{-1}$
      replace y_{n+1} by y_{-2} and go to next iteration
    else
      replace y_{n+1} by y_{-1} and go to next iteration
else if $f_{-1} \geq f(y_n)$
    if $f(y_n) \leq f_{-1} < f(y_{n+1})$
      evaluate $f_{-1/2} = \bar{y}(-1/2)$
      if $f_{-1/2} \leq f_{-1}$
        replace y_{n+1} by $y_{-1/2}$ and go to next iteration
    else
      evaluate $f_{1/2} = \bar{y}(1/2)$
        if $f_{1/2} < f_{n+}$`
        replace y_{n+1} by $y_{1/2}$ and go to next iteration
    Replace $y_i \leftarrow (1/2)(y_1 + y_i)$ for $i = 2,3....,n+1$

Figure 2 One step of the Nelder-Mead Procedure (Nocedal and Wright 2005)

3.4 Selection of the Initial Set Points

As explained in Section 4, this is a non-convex nonlinear problem. That is, it can have multiple local solutions. Our experimental results have confirmed the insights from Suzuki and Okabe (Suzuki and Okabe 1995) that the initial set of solutions for constructing the Voronoi diagrams has an important influence on the quality of the final solution. They suggest a Monte Carlo sampling (we will refer this method as simply random sampling), i.e. the coordinates of the starting points were randomly generated following a Uniform distribution (0,1). We have evaluated this approach and, alternatively, we tested two additional strategies for selecting initial feasible facility locations. The strategies are based on two other well-known sampling methods: (i) proportional sampling based approach and (ii) latin hypercube sampling (LHS).

The proportional sampling based approach allocates initial points randomly based on density of points in a pre-defined area. For example, the area under study can be divided into smaller sub-areas. In each sub-area, we randomly allocate initial points proportional to the number of demand points contained in this subarea. For this test, we divided the square plane into 4 equal squares, and then, the number of facilities was proportionally assigned to the number of demand points in each square. The coordinates of the facilities generated at each small square were generated randomly following a Uniform distribution.

The latin hypercube sampling is a stratified-random procedure where samples are obtained from their distributions. The cumulative distribution for each variable is divided into N equi-probable intervals. Assuming that the variables are independent, a value is selected randomly from each interval. For this sampling procedure we assume that the distribution is known *apriori*, by developing an empirical distribution from the demand points.

We tested the performance of each strategy procedure using a square plane with vertices at (0,0), (0,1), (1,0) and (1,1) as area of study. In this area, we randomly generate a set of 10 problems fixing the number of facilities at 10 and using 100 (small problem) and 1000 (large problem) demand points. That is, at each combination of facilities and demand points we randomly generated 10 different sets of demand points. Each problem was solved using 20 different starting points for the facilities for each strategy. The coordinates (x,y) of the demand points were generated using two different distributions: (1) Uniform distribution U(0,1) and a truncated Normal distribution with mean 0.5 and standard deviation of 0.15, where values under 0 where set at 0.01 and values over 1 were set at 0.99 in order to avoid points in the boundary of the square region. For these evaluations, the social cost function used was $0.1e^{2.2\max\{\|x_j - x_i\|\}} = 0.1e^{2.2d_{ij}^{\max}}$ where p_j = 1 in Equation 6.

Comparison between sampling strategies

Tables 1a and 1b present the performance in the solution achieved using all strategies. We use three measures to evaluate the quality of the solution (see Table 1a).

The first measure is the number of problems where the strategy achieved the best solution (Problems –Best solution achieved). That is, after running the experiments we counted the number of problems out of the 10 problems where the best solution (minimum) was obtained from all 20 initial random points under all strategies. In both, large and small problems, the sampling strategy achieved the best solution in more problems. The Monte Carlo procedure only achieved the best solution in two of the large problems under the truncated normal distribution and in one of the small problems.

The other two measures are (1) the performance of the average case (Average Case) by comparing how far (in percentage) is the average of each of the solutions in each problem from the best solution among all strategies; and (2) the performance of the worst solution (Worst Case) achieved as compared with the best solution. Under all strategies, the heuristic provided an acceptable range of solutions. The maximum deviation from the best solution (10.54%) is obtained using the Monte Carlo procedure in the truncated small problem. The best results are obtained using the sampling procedures. For small problems (m =10 and n = 100), it can be observed that the proportional sampling based approach does not differ significantly from the non-sampling approach. However this difference becomes more significant as the problem size increases. Another important insight is that as the problem becomes larger the solution found is closer to the best solution achieved for all the strategies compared. In large problems the worst performance is obtained using the random sampling (around 2.64%) whenever the other sampling procedures have worst performances of 1.22% (latin hypercube) and 2.22% (proportional sampling).

In terms of CPU time, we computed the mean, the minimum time and the worst time. The results (see Table 1b) showed similar running times for small problems for all strategies. However, for large problems, the latin hypercube sampling obtained a solution faster than the other two methods.

For the comparison between sampling strategies different than Monte Carlo, we use the same performance metrics as described above in addition to the CPU time. In terms of CPU time, in small and large problems, the latin hypercube average running time is lower than the average running time using the proportional sampling based approach (see Table 1b). The situation is similar for the minimum running time and the maximum running time. The latin hypercube sampling shows a small standard deviation on the final solutions as compared to the other solution techniques. It was observed that the solution quality using the proportional sampling technique is slightly better than the latin hypercube sampling approach.

Table 1. Comparison between Strategies for Initial Points

a. Objective Function

		Random			Proportional Sampling (PS)			Latin Hypercube (LH)		
		Problems - Best solution achieved	**Average Perfomance**	**Worst Case**	**Problems – Best solution achieved**	**Average Perfomance**	**Worst Case**	**Problems - Best solution achieved**	**Average Perfomance**	**Worst Case**
Normal Distribution	m= 10, n = 100	0	7.17%	10.54%	0	6.16%	9.23%	10	0.62%	1.57%
	m = 10, n = 1000	2	1.44%	2.64%	6	0.76%	2.22%	2	0.91%	1.22%
Uniform Distribution	m= 10, n = 100	1	3.3%	6.79%	5	1.81%	2.42%	4	2.35%	3.61%
	m = 10, n = 1000	0	1.21%	1.96%	4	0.52%	1.22%	6	0.93%	1.52%

b. CPU Time (in seconds)

		Random				Proportional Sampling				Latin Hypercube			
		Average	**Min**	**Max**	**Stdev**	**Average**	**Min**	**Max**	**Stdev**	**Average**	**Min**	**Max**	**Stdev**
Normal Distribution	m= 10, n = 100	1.3265	0.7819	2.0567	0.3702	1.3091	0.7273	2.0263	0.3426	1.2782	0.7896	2.0997	0.3286
	m = 10, n = 1000	13.2530	7.7829	22.0181	3.9435	13.7738	7.4976	23.7390	4.5062	11.4983	19.9209	6.5616	3.4931
Uniform Distribution	m= 10, n = 100	1.1344	0.7110	1.8606	0.3072	1.1244	0.6855	1.7373	0.2945	1.1021	0.6879	1.7608	0.2977
	m = 10, n = 1000	16.0034	8.9597	26.9445	5.0202	13.0964	7.3041	21.6188	4.0050	12.8036	6.7021	20.4223	3.8583

3.5 Computational Evaluation

Our experimental results have confirmed the insight from Suzuki and Okabe (Suzuki and Okabe 1995) that, given the non-convexity of the problem, the initial set of solutions for constructing the Voronoi diagrams has an influence on the final solution. However, to assess the quality of the result as compared with the solution obtained using other solvers, we coded the algorithm in Matlab 7.0 using the built-in function for computing the Voronoi diagram using Delaunay triangulation while the Nelder-Mead method has been coded also in Matlab 7.0 (Mathworks, 2010) as an additional function. We generated a set of 16 problems varying the number of facilities from 1-50 and the demand points from 10-5000 (see Table 2 and 3). Similar to section 5 we used a square region with vertices at (0,0), (0,1), (1,0) and (1,1). Each demand point was generated randomly following the two distributions, Uniform(0,1) and Normal distribution as discussed in section 5. The facilities were initially located using the proportional hypercube sampling method and the cost was the same as the one used in Section 5.

To assess the quality of the solution we chose a subset of these problems and solved the p-median Uncapacited Facility Location version of the problem using a specialized global solver for mixed non-linear integer programming, Branch-And-Reduce Optimization Navigator (BARON). BARON is a computational system for solving non-convex optimization problems to global optimality for purely continuous, purely integer, and mixed-integer nonlinear problems. The subset of problem corresponds to the first eight problems generated, which were solved using the NEOS Server. On the other hand, the heuristic is run on a Pentium Core 2 Duo desktop of 2.3 Ghz and 2 Gb of RAM.

Tables 2 and 3 report the best solution obtained by the heuristic after running 10 different starting points. We also report the best solution obtained by BARON.

In the truncated normally generated demand points the heuristic reports a very close lower bound of the solution in the range of -12% and 0.02%. For the uniformly generated data, the gap between the solutions obtained by the heuristic and BARON were in the range of -6.45% and 7.56%. Moreover, the variation among solutions (see standard deviation of the solution in Tables 2 and 3) obtained through each different initial point generated, were very low, which shows the robustness of the heuristic.

We do not report the CPU time for BARON because the conditions for comparison were not the same. Our heuristic instead solved the largest problem (m= 20 and n =5000) problem in 3094 and 2001 seconds for the truncated and uniformly distributed demands respectively. This includes the total time for 10 different initial points. Observe also that the solutions achieved had small standard deviations.

Table 2. Comparison between Voronoi Heuristc and BARON solver for points generated with a truncated normal distribution

		Proportional Sampling (10 different starting points)			**BARON**	% Diff
Facilities (m)	**Demand (n)**	**Best Solution**	**Tot. CPU Time (in sec)**	**Std. Dev. of Solutions**	**Optimal Solution**	
2	10	1.3823	3.3090	0.0000	1.5730	-12.12%
3	10	1.2586	2.0921	0.0411	1.3934	-9.67%
4	10	1.1807	2.0439	0.0384	1.3418	-12.01%
5	10	1.137	2.3077	0.0450	1.1813	-3.75%
3	20	2.8997	2.8554	0.0361	2.8990	0.02%
4	20	2.7069	3.4080	0.0621	2.7719	-2.34%
5	20	2.5235	5.2752	0.0516	2.5315	-0.32%
10	50	5.7735	7.6316	0.0431		
5	100	12.7765	9.2918	0.0720		
10	100	11.8284	10.5054	0.0591		
20	100	11.0720	17.4371	0.0441		
10	1000	119.0752	140.1177	0.2718		
20	1000	113.2694	177.0186	0.0592		
50	1000	107.9404	257.5415	0.0611		
20	5000	565.5630	1425.4045	0.1374		
50	5000	540.7252	3094.8271	0.2006		

Table 3 Comparison between Voronoi Heuristc and BARON solver for points generated with a uniform distribution

Facilities (m)	Demand (n)	Proportional Sampling (10 different starting points) Best Solution	Tot. CPU Time (in sec)	Std. Dev. of Solutions	BARON - Optimal Solution	% Diff
2	10	1.8852	1.0930	0.1090	1.9472	-3.18%
3	10	1.6679	0.4949	0.0518	1.5506	7.56%
4	10	1.342	0.9046	0.0992	1.3038	2.93%
5	10	1.2068	1.9060	0.1529	1.2367	-2.42%
3	20	3.4875	2.5949	0.0425	3.4875	0.00%
4	20	3.0621	3.2182	0.1293	3.0621	0.00%
5	20	2.8042	0.8209	0.1077	2.9976	-6.45%
10	50	6.0959	4.9932	0.0522		
5	100	14.5258	6.8529	0.1089		
10	100	12.7087	9.5084	0.1197		
20	100	11.5600	13.0731	0.0689		
10	1000	131.0937	112.3177	0.1776		
20	1000	119.7891	106.0767	0.2790		
50	1000	111.4911	424.5710	0.1483		
20	5000	604.6945	2001.9042	0.2937		
50	5000	561.8537	2090.6907	0.2435		

3.6 Example

In this section, we provide an application example using data from Hurricane Katrina. According to the Congressional Research Service (CRS) report for Congress about Hurricane Katrina (Gabe et al. 2005), more than 700,000 people were acutely impacted by hurricane Katrina; these people lived in neighborhoods either affected by the flooding or with significant structural damage. According to the White House report of the Federal Response to Hurricane Katrina (2006), assets were pre-deployed at different sites throughout the region to encircle the forecasted impact area.

For this application example, we chose 33 locations, among parishes and counties from 6 metropolitan areas in Louisiana and 5 metropolitan areas in Mississippi, as demand points (see Table 4). We transformed the data so that we can use the algorithm enclosing the resulting points and the study area into a square with vertices at (0,0), (0,1), (1,0) and (1,1).

We first evaluated the base case or actual locations for pre-positioning (Figure 3a) for which the objective function based on an arbitrarily chosen cost function $p_j\left(0.1e^{2.2d_{ij}}\right)$ was computed. Note also that, by constructing the Voronoi diagram using the actual facilities, one of the warehouses has no demand point assigned (see Figure 3a). Then we run the algorithm, using the proportional sampling based approach, for 10 different initial points, keeping the best value obtained. We repeated this process for 3, 4, and 5 facilities. The results obtained are presented in Table 5. To provide more meaningful results, we transformed the results in total distance traveled from each demand point to its corresponding facility using the real distance between latitudes and longitudes in miles. The calculation assumes the earth is a perfect sphere of radius 3,963.1 miles using the coordinates of the actual facilities and the ones obtained through the heuristic.

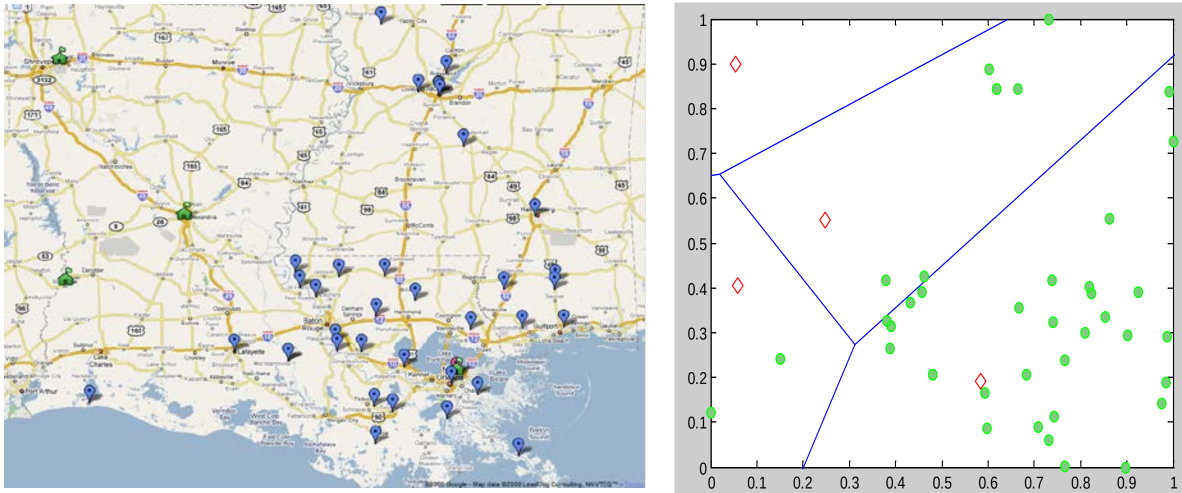
From the results it is interesting to note that even for the case with 3 facilities, the solution can be improved up to 52.5% in total miles traveled and 50% in average miles. Figure 3b shows the best solutions for 4 points, which reduces the total miles traveled from 4,644.45 miles to 2,304 miles and the average distance traveled from the facilities to their assigned demand points reduces from 164.5 miles to 69.83. Moreover, Table 5 shows that in all cases there are savings in terms of distance of around 50%. Furthermore, if we account for the coverage measured by population per mile traveled, the improvements are more significant (see Table 6).

These results have two important implications: (1) that using a Voronoi diagram to assess the coverage of the warehouse can give us an accurate estimate of the total area covered, and (2) that by using the proposed heuristic we were able to provide locations with less distance traveled from the warehouse to the demand point.

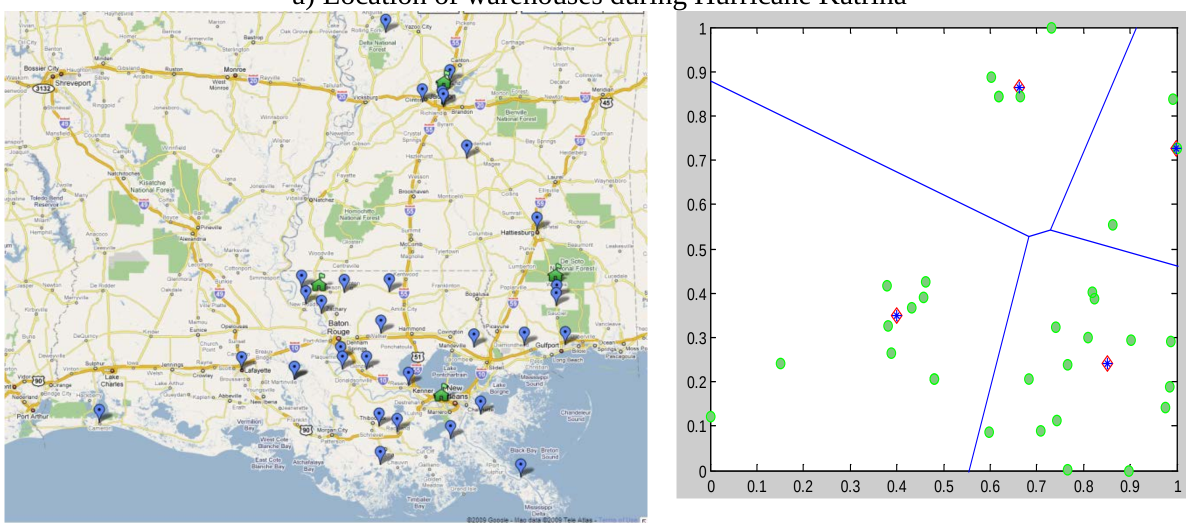
Table 4. Location of Demand Points and the warehouses used during Katrina Hurricane

No.	**City**	**Population**
1	Ascension	94,520
2	East Baton Rouge (Central City: Baton Rouge)	424,597
3	East Feliciana	20,802
4	Iberville	32,526
5	Livingston	112,445
6	Pointe Coupee	22,212
7	West Baton Rouge	22,126
8	West Feliciana	15,111
9	Lafourche	89,974
10	Terrebonne (Central City: Houma)	104,503
11	St. Martin	48,583
12	Lafayette (Central City: Lafayette)	190,503
13	Calcasieu	183,577
14	Cameron (Central City: Lake Charles)	9,991
15	Jefferson	455,466
16	Orleans (Central City: New Orleans)	484,674
17	Plaquemines	26,757

No.	**City**	**Population**
18	St. Bernard	67,229
19	St. Charles	48,072
20	St. John the Baptist	43,044
21	St. Tammany	191,268
22	Hancock	42,967
23	Harrison (Central City: Gulfport-Biloxi)	189,601
24	Stone	13,622
25	Forrest (Central City: Hattiesburg)	72,604
26	Perry	12,138
27	Copiah	28,757
28	Hinds (Central City: Jackson)	250,800
29	Madison	74,674
30	Rankin	115,327
31	Simpson	27,639
32	George	19,144
33	Jackson (Central City: Pascagoula)	131,420



a) Location of warehouses during Hurricane Katrina



b) Location of warehouses (Heuristic solution)

Figure 3. Comparison between location of warehouses under Hurricane Katrina and the solution achieved using the Voronoi heuristic

Table 5 Results for Real World Case: Katrina actual warehouse locations vs. results using the heuristic (using 10 different initial points)

	Total Distance		**Average Distance**		**Min Distance**	**Max Distance**
	Miles	**% Improv.**	**Miles**	**% Improv.**	**Miles**	**Miles**
n=3	2,203.88	52.55%	72.00	50.39%	15.79	271.22
n=4	2,304.30	50.39%	69.83	51.89%	15.8	158.52
n=5	2,223.77	52.12%	67.39	53.57%	2.38	140.05
Actual Locations	4,664.45		145.14		70.26	223.73

Table 6 Results using miles-population

	Total Population/Mile	**Average Pop/Mile**	**Min Pop/Mil**	**Max Pop/Mile**
n=3	78,341.9	2,448.2	63.0	26,892.0
n=4	82,203.0	2,491.0	63.0	26,865.0
n=5	147,925.8	4,482.6	117.3	80,171.3
Actual Locations	25,319.6	791.2	59.7	4,248.1

CHAPTER 4. CONCLUSIONS

This chapter summarizes the research, highlights its contributions, and proposes directions for future research.

4.1 Summary

This study addresses the two primary objectives:

1. Determines a methodological framework for planning of service facility locations under high demand requires consideration of customers' en-route travel and their in-facility delay, as well as the reliability of service facilities against natural or man-made hazards. This project first presents a scenario-based mixed-integer non-linear program (MINLP) model that integrates service disruption risks, en-route traffic congestion and in-facility delay into an integrated facility location problem. We derive lower and upper bounds to this highly complex problem by approximating the expected system costs associated with customer arrival, en-route travel, in-facility delay, and service. This allows us to develop a more tractable approximate MINLP formulation to minimize the expected overall system cost under all probabilistic service disruption scenarios. We also develop a Lagrangian relaxation (LR) based solution approach to decompose the integer and continuous variables of this approximation model, and reformulate the relaxed sub-problem for location and service allocation decisions into a conic program. Numerical experiments show that the approximation model and LR solution approach are capable of overcoming the computational difficulties associated with reliable service facility location design. Managerial insights are drawn from a series of case studies and sensitivity analyses.

2. In this project we have discussed the issue of finding appropriate location for sitting distribution centers for disasters by addressing one of the challenges of humanitarian logistics: response time. Given that social cost functions have gained relevance in models for disasters, we have proposed a model that incorporates these social costs. The social cost function in this case has been defined as urgency delivery cost, which is distance based, and results in a nonlinear non-convex model. Due to the non-convex nature of the social cost function chosen the problem might have a multiple solutions. Moreover, in order to approximate continuous demand functions a large number of demand points are required, increasing the complexity of finding optimal solutions. For these reasons we propose a heuristic constructed under the idea of Voronoi diagrams. Computational results showed that the heuristic achieves high-quality solutions in reasonable time. We have also provided an evaluation of strategies for selecting the initial starting points for the heuristic, which helps to speed it up and reduces the standard deviation of the solutions that can be achieved, improving its robustness. Contrary to this approach, which is limited in the sense that we are assuming direct deliveries, similar to the ones proposed by Balcik and Beamon (Balcik and Beamon, 2008) and uncapacitated (Akkihal, 2006), future research is intended to not only find optimal locations but to also include routes in the optimization problem (last mile distribution).

4.2 Future research directions

The present research addressed the problem of facility location and rerouting of traffic in disasters. Future research directions include considering network equilibrium conditions, facility congestion costs and external validation with multiple disasters. This project is an important step in that direction.

REFERENCES

R. K. Ahuja, T. L. Magnanti, J. B. Orlin, (1993). Network Flows: Theory Algorithms and Applications, Prentice Hall.

Garcia, L.A., Sozen, M.A. and Ramirez, J.A. (2004). Emergency Earthquake Routes for the State of Indiana; Part II: Route Seismic Vulnerability Aspects, JTRP Draft Final.

Report FHWA/INDOT-2003/22.

Peeta, S. and Viswanath, K. (2004). Emergency Earthquake Routes for the State of Indiana; Part I: Criteria for Selection of Primary Routes: Transportation Aspects, JTRP

Draft Final Report FHWA/INDOT-2003/22.

Viswanath K. and Peeta S., 2003. The multicommodity maximal covering network design problem for planning critical routes for earthquake response. Transportation Research Record, 1857, p. 1-10.

Wong, R.T., (1980). Integer programming formulations of the travelling salesman problem, Proc. IEEE Conf. on Circuits and Computers, p. 149-152.

Aboolian, R., Berman, O., Drezner, Z., 2008. Location and allocation of service units on a congested network. IIE Transactions 40, 422-433.

Aboolian, R., Berman, O., Krass, D., 2012. Profit maximizing distributed service system design with congestion and elastic demand. Transportation Science 46(2), 247-261.

An, S., Cui, N., Li, X. P., Ouyang, Y., 2013. Location Planning for Transit-based Evacuation under the Risk of Service Disruptions. Transportation Research Part B 54, 1-16.

Atamtürk, A., Berenguer G., Shen, Z.-J. M., 2012. A conic integer programming approach to stochastic joint location-inventory problems. Operations Research 60(2), 366-381.

Bai, Y., Hwang, T., Kang S., Ouyang, Y., 2011. Biofuel refinery location and supply chain planning under traffic congestion. Transportation Research Part B 45, 162-175.

Ball, M. O., Lin., F. L., 1993. A reliability model applied to emergency service vehicle location. Operation Research 41(1), 18–36.

Bar-Gera, H., 2009. Transportation Network Test Problems. http://www.bgu.ac.il/~bargera/tntp/ (accessed 06.02.13)

Benjaafar, S., 1995. Performance bounds for the effectiveness of pooling in multi-processing systems. European Journal of Operational Research 87 (2), 375-388.

Berman, O., Larson, R. C., Chiu, S. S., 1985. Optimal server location on a network operating as an M/G/1 queue. Operations Research 33, 746–771.

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

Boyce, D.E., Chon, K.S., Ferris, M.E., Lee, Y. J., Lin, K-T., Eash, R.W., 1985. Implementation and evaluation of combined models of urban travel and location on a sketch planning network, Chicago area transportation study, xii + 169 pp.

Bundschuh, M., Klabjan, D., Thurston, D. L., 2003. Modeling robust and reliable supply chains. Optimization Online e-print. www.optimization-online.org (accessed 01.15.13)

Castillo, I., Ingolfsson A., Sim, T., 2009. Social Optimal Location of Facilities with Fixed Servers, Stochastic Demand, and Congestion. Production and Operations Management 18(6), 721-736.

Chen, Q., Li, X., Ouyang, Y., 2011. Joint inventory-location problem under the risk of probabilistic facility disruptions. Transportation Research Part B 45(7), 991-1003.

Corrêa, F. A., Lorena, L. A. N., Ribeiro, G. M., 2009. A decomposition approach for the probabilistic maximal covering location-allocation problem. Computers & Operations Research 36, 2729-2739.

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

Daskin, M. S., 1982. Application of an Expected Covering Model to EMS System Design. Decision Sciences 13 (3), 416-439.

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

Eash, R.W., Chon, K.S., Lee, Y. J., Boyce, D. E., 1983. Equilibrium traffic assignment on an aggregated highway network for sketch planning, Transportation Research Record, 994, 30-37.

Eiselt, H. A., Gendreau, M., Laporte, G., 1996. Optimal location of facilities on a network with an unreliable node or link. Information Processing Letters 58(2), 71–74.

Fisher, M., 1981. The lagrangian relaxation method for solving integer programming problems. Management Science 27 (1), 1{18.

Friesz, T. L., 2011. Supply chain disruption and risk management. Transportation Research Part B 45(8), 1125-1127.

Friesz, T. L., Lee, I., Lin C. C., 2011. Competition and disruption in a dynamic urban supply chain. Transportation Research Part B 45(8), 1212-1231.

Jaller, M., Holguín-Veras, J., 2012. Locating points of distribution in large urban disasters. Submitted to Transportation Science.

Hajibabai, L., Bai, Y., Ouyang, Y., 2014. Joint optimization of freight facility location design and pavement rehabilitation plan under traffic equilibrium. Submitted to Transportation Research Part B, 63: 38-52.

Hajibabai, L., Ouyang, Y., 2013. Integrated Planning of Supply Chain Networks and Multimodal Transportation Infrastructure expansion: model development and application to the biofuel industry. Computer-Aided Civil and Infrastructure Engineering 28, 247–259.

Konur, D., Geunes, J., 2011. Analysis of traffic congestion costs in a competitive supply chain. Transportation Research Part E 47, 1-17.

Konur, D., Geunes, J., 2012. Competitive multi-facility location games with non-identical firms and convex traffic congestion costs. Transportation Research Part E 48, 373-385.

Kwasnica, A. M., Stavrulaki, E., 2008. Competitive Location and Capacity Decisions for Firms Serving Time-Sensitive Customers. Naval Research Logistics 55(7), 704-721.

Larson, R. C., 1974. A hypercube queuing model for facility location and redistricting in urban emergency services. Computers & Operations Research 1, 67–95.

Larson, R. C., 1975. Approximating the performance of urban emergency service systems. Operations Research 23(5), 845–868.

Li, A. C. Y., Nozick, L., Xu, N. X., Davidson, R., 2012. Shelter location and transportation planning under hurricane conditions. Transportation Research Part E 48, 715-729.

Li, X., Ouyang, Y., 2010a. A continuum approximation approach to reliable facility location design under correlated probabilistic disruptions. Transportation Research Part B 44(4), 535-548.

Li, X., Ouyang, Y., 2010b. Reliable sensor deployment for network traffic surveillance. Transportation Research Part B 45(1), 218-231.

Li, X., Ouyang, Y., 2012. Reliable traffic sensor deployment under probabilistic disruptions and generalized surveillance effectiveness measures. Operations Research 60(5), 1183-1198.

Li, X., Ouyang, Y., and Peng, F. 2013. A supporting station model for reliable infrastructure location design under interdependent disruptions. Transportation Research Part E, 60: 80-93.

Marianov, V., ReVelle, C., 1996. The queuing maximal availability location problem: a model for the siting of emergency vehicles. European Journal of Operational Research, 93, 110–120.

Marianov, V., Rı´os, M., Icaza, M. J., 2008. Facility location for market capture when users rank facilities by shorter travel and waiting times. European Journal of Operational Research 191, 32-44.

Marianov, V., Serra, D., 2002. Location-Allocation of Multiple-Server Service Centers with Constrained Queues or Waiting Times. Annals of Operations Research 111, 35-50.

Nel, L.D., Colbourn, C. J., 1990. Locating a broadcast facility in an unreliable network, INFORMS 28 , 363-379.

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

Qi, L., Shen Z. M., Snyder L.V., 2010. The effect of supply disruptions on supply chain design decisions. Transportation Science 44(2), 274-289.

ReVelle, C., Hogan, K., 1989. The maximum availability location problem. Transportation Science 23, 192–200.

Shen, Z.J.M., Zhan, R., Zhang, J., 2011. The reliable facility location problem: formulations, heuristics, and approximation algorithms. INFORMS Journal on Computing 23(3), 470-482.

Sherali, H. D., Carter, T. B., 1991. A location-allocation model and algorithm for evacuation planning under hurricane/flood conditions. Transportation Research Part B 25(6), 439-452.

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

Snyder, L. V., Daskin, M. S., Teo, C. P., 2007. The stochastic location model with risk pooling. European Journal of Operational Research 179(3), 1221-1238.

Syam, S. S., 2008. A multiple server location–allocation model for service system design. Computers & Operation Research 35, 2248-2265.

Zhang L., Rushton, G., 2008. Optimizing the size and locations of facilities in competitive multi-site service systems. Computers & Operations Research 35, 327-338.

Zhang, Y., Berman, O., Marcotte, P., Verter, V., 2010. A bilevel model for preventive healthcare facility network design with congestion. IIE Transactions 42, 865-880.

Zhang, Y., Berman, O., Verter, V., 2009. Incorporating congestion in preventive healthcare facility network design. European Journal of Operational Research 198, 922-935.