RELIABLE FACILITY LOCATION DESIGN AND TRAFFIC SENSOR DEPLOYMENT UNDER PROBABILISTIC DISRUPTIONS

BY

XIAOPENG LI

DISSERTATION

Submitted in partial fulfillment of the requirements
for the degree of Doctor of Philosophy in Civil Engineering
in the Graduate College of the
University of Illinois at Urbana-Champaign, 2011

Urbana, Illinois

Doctoral Committee:

Professor Yanfeng Ouyang, Chair
Professor Christopher P.L. Barkan
Professor Rahim F. Benekohal
Professor Xin Chen

Abstract

Many private enterprises and public agencies have faced the problem of locating facilities over spatial dimensions to provide certain service functions. In the supply chain context, we often need to locate a variety of private or public facilities (e.g., manufacturing, assembly plants, schools and hospitals) to serve distributed customers. In the traffic engineering context, various types of surveillance sensors (e.g., induction loops, video cameras and radio frequency transponders) are deployed in transportation networks to estimate real-time traffic states, which are valuable information for both private sectors (e.g., tracking fleets for trucking companies, providing real-time traveler information) and public agencies (e.g., congestion mitigation, accident management). In every case, the operational efficiency and system benefit depend on the choices of facility locations. A good location design can maximize the system benefit while saving as much infrastructure investment as possible.

Due to natural disasters or human hazards (e.g., power outages, operational accidents, labor actions or terrorist attacks), facility disruptions are frequently observed in many contexts in the real world. These disruptions often adversely impair the benefit from these facilities. Proper redundancy in the location design is helpful to enhance system reliability and mitigate losses from such disruptions. However, reliable facility location problems are difficult mainly due to the large number of possible failure scenarios. In this Ph.D. research, we will overcome this challenge by developing a range of innovative modeling methods, and then generalize the methodologies to address supply chain design and traffic surveillance sensor location problems.

Traditional discrete location models (where customers and candidate facility locations are represented by discrete points) are NP-hard; i.e., they are suitable for small-scale problem instances, but suffer from excessive computational burden when problem size becomes large. To improve computational tractability, continuum approximation models (where customers and facilities are approximated by continuous spatial densities) are developed to approximate problems in a continuous metric space and provide good approximate solutions to large-scale instances.

We propose a continuum approximation (CA) model for the reliable uncapacitated fixed charge facility location problem to determine optimal facility locations that minimize the one-time investment for facility constructions and the long-run expected transportation costs for serving spatially distributed customers under correlated facility failures. Complex facility failure mechanisms such as spatial correlation or cascading failure effect are addressed. We identified a few interesting properties of the CA model and developed effective solution algorithms. We have tested this model over different types of numerical examples, and

useful managerial insights on how failure correlation impacts the location design are drawn.

There are many connections between supply chain facility location problems and sensor location design problems in the traffic surveillance context. For example, in traffic surveillance, we can view traffic surveillance sensors as facilities and traffic OD flow paths as customers being served (or inspected) by these facilities. For a traffic surveillance sensor system, benefits are generated by estimating the real-time traffic states with collected samples at installed sensors, and hence costs come from estimation errors, i.e., the differences between the estimated and the actual traffic states. Based on these connections, this research uses methodologies for supply chain facility location problems to determine surveillance sensor location design in a traffic network. We propose a discrete reliable sensor location model that takes into account the surveillance benefit from not only individual sensor data but also synthesized information from multiple sensors under probabilistic sensor failures. Like many other location design problems, the deterministic version of the sensor location model is already complex; considering an exponential number of possible failure scenarios will further increase the difficulty. Hence we propose efficient customized solution algorithms based on greedy heuristic and Lagrangian relaxation. We compare their performance with that of well-known commercial software (e.g., CPLEX). Numerical examples including a full-scale railroad wayside sensor location design are presented to show that the innovative model significantly improves the state of practice, and the proposed algorithms solve the problem efficiently even when commercial software fails to provide reasonable solutions. We further encapsulated the solution algorithm into a piece of stand-alone software for railroad wayside sensor location design, which has been adopted by the industry.

This sensor location model is further extended to generalize surveillance effectiveness measures and accommodate site-dependent failure probabilities. In the extended sensor location design framework, traffic surveillance effectiveness is defined as the reduction of "generalized estimation errors" on all highway segments between neighboring sensor pairs, such that most existing measures can be expressed as special cases. The problem is first formulated into a compact mixed-integer program, and we develop a variety of solution algorithms (including a custom-designed Lagrangian relaxation algorithm) and analyze their properties. We also propose alternative formulations including a continuum approximation model for single corridor problems and reliable fixed-charge sensor location models. Numerical case studies are conducted to test the performances of the proposed algorithms and draw managerial insights on how different parameter settings (e.g., failure probability and spatial heterogeneity) affect the optimal sensor deployment and the overall surveillance effectiveness.

To my family

Acknowledgments

I am deeply indebted to my adviser, mentor and friend, Professor Yanfeng Ouyang, for his constant guidance, inspiration, and encouragement throughout my graduate studies at the University of Illinois at Urbana-Champaign. It was Professor Ouyang who guided me into the transportation area, provided me with access to diverse challenge problems, and helped me conceive and carry out this dissertation work. His challenging mind, endless passion and persistent pursuit for understanding of our surroundings shaped my determination to pursue an academic career.

I am also very grateful to the three other members of my committee, Professor Christopher P.L. Barkan, Professor Rahim F. Benekohal and Professor Xin Chen. From their classes and seminars, I have learned a great deal of fundamental knowledge and methodologies on which my research is based. Their valuable comments and suggestions have substantially improved this thesis.

This research has much benefited from the courses I have taken from the transportation engineering program and several other departments at the University of Illinois. I want to express my sincere thanks to all the professors whom I have taken classes or interacted with.

My graduate study at the University of Illinois has been very stimulating and delightful also because of many fellow students and friends. I would like to particularly thank my colleague Fan Peng, who collaborated with me on a number of research topics and was my best consultant on modeling techniques and algorithm development. I would thank Eunseok Choi for his help on data preparation for some of my case studies. I also sincerely appreciate the friendship and generous help (both in research and in life) from my officemates, Yun Bai, Eunseok Choi, Leila Hajibabai, Taesung Hwang, Seyed Mohammad Nourbakhsh and Fan Peng, and all other students in our program.

I want to send my special thanks to my wife, Xiying Mi. It would not have been possible for me to finish this Ph.D. research without her love and unwavering support.

Finally, I acknowledge financial support from the National Science Foundation, CSX Transportation, Inc., and the USDOT NEXTRANS Center.

Table of Contents

List of Tables

List of Figures

Chapter 1

Introduction

1.1 Motivation

Most private enterprises and public agencies have faced the problem of locating facilities over spatial dimensions to provide certain service functions to their distributed clients or customers. Industrial firms need to locate a variety of facilities in the supply chain including manufacturing and assembly plants, warehouse and retail outlets. Government agencies must determine locations of public service facilities such as schools, hospitals, fire stations, ambulance bases and landfill. In every case, the operational efficiency and system benefit depend on the choices of facility locations. A good location design could maximize the service benefit while saving as much infrastructure investment as possible.

Uncertainties such as demand fluctuations and probabilistic facility disruptions are often observed in many real-world contexts and impose significant challenges to facility location planning. Although demand uncertainties have been extensively studied in the past few decades, only limited research has been conducted on the uncertainties of facilities. In reality, facility operations may be disrupted from time to time due to reasons such as natural disasters, power outages, operational accidents, labor actions or terrorist attacks. The failure of a facility will force its customers to either seek service at some other functioning facility (albeit less convenient) or completely give up service. Either way, system operation cost increases and service quality deteriorates. The adverse effect may be further exacerbated if multiple facilities fail simultaneously. Furthermore, many facility disruption cases exhibit not only site-dependent failure probabilities but also strong spatial correlations (e.g., due to shared exposure to common hazards). All these challenges and complexities raise the need for a reliable facility design framework that hedges against all possible scenarios of facility failures.

In the traffic engineering context, facility problems are also quite common. Sensing and

information technologies have been successfully applied in many ways and hold the promise for efficient estimation, monitoring, and management of many complex engineering systems. Traffic surveillance technologies, which are critical components of intelligent transportation systems, are also getting mature. A variety of sensor technologies, such as induction loops, video cameras and radio frequency identification (RFID), have been applied in transportation networks. These technologies can provide crucial real-time information and help improve estimation of transportation states. Such information is valuable for both private sectors (e.g., tracking fleets for trucking companies, providing real-time traveler information) and public agencies (e.g., congestion mitigation, accident management). Real-time information enables road users to choose routes that avoid congestion, traffic operators to promptly respond to congestion patterns and efficiently select control strategies, and the homeland security to locate most hazardous parts of a large transportation network in real time and carry out preventive actions.

Compared to the facilities in supply chains, traffic surveillance sensors have different types of service and benefit measures. For example, supply-chain facilities provide service to discretely or continuously distributed customer demand in a space while traffic surveillance sensors inspect traffic flows along O-D paths in a network. The utility of supply-chain facilities is quantified by the reduction of logistic cost such as inventory holding cost and customer traveling cost while the benefit of traffic surveillance sensors is measured by the improvement of network traffic state estimation by sensor data. In addition, different types of sensors provide different data and may have different benefit measures. Traditional traffic surveillance sensors (e.g., loop detectors) usually provide aggregated statistical data such as volume count and vehicle speed. Newer sensors (e.g., RFID) can identify individual vehicle and enable synthesis of disaggregated data from multiple sensors.

Properly locating surveillance sensors is critical to accurate real-time traffic estimation over transportation networks. Ideally, sensors can be densely deployed over a transportation network and each of them collects real-time traffic data around its location. Then the estimation for the whole network can be obtained by merging and interpolating local estimates by each individual sensors, which apparently has very high accuracy and can promptly capture anomalous traffic states (e.g, traffic accident detection). However, implementing such a sensor system requires enormous infrastructure investment, which is not realistic given limited resources. Furthermore, like many other IT technologies, most sensors are subject to performance disruptions due to technology flaws, system errors, adverse weather conditions, or intentional sabotages (Rajagopal and Varaiya, 2007; Carbunar et al., 2005). Such failures may substantially impair traffic network coverage and surveillance effectiveness. A practical solution would be to utilize available samples from a number of operational sensors

to reconstruct traffic states of the entire network based on traffic fundamental properties. In this case, the sensor locations are critical to obtain the most representative information over a network that maximizes expected estimation accuracy. A reliable sensor location framework shall be established that optimizes the trade-off between infrastructure investment and expected surveillance benefit across all possible sensor failure scenarios.

1.2 Objectives

This study will investigate location design for both supply chain facilities and traffic surveillance sensors. We will first review existing studies on facility location problems with both discrete and continuous modeling techniques in the supply chain context. Discrete models, though well developed, are generally not suitable for large-scale problem instances, especially those involving complex facility failure patterns. In this thesis, we first propose a continuum approximation (CA) approach to solve large-scale facility location problems with facility failure correlations.

We also aim to adapt these methodologies into those suitable for traffic sensor location problems. We propose a discrete reliable sensor location model for travel time estimation over general transportation networks. The model is extended to address a variety of sensor technologies, general surveillance benefit measures and complex sensor failure mechanisms. A continuum approximation approach for sensor location design along highway corridors is also proposed, and it is shown to be computationally very efficient.

1.3 Contribution Statement

This work proposes methodologies that address reliable location design in both supply chain and traffic surveillance contexts. In spite of decades of efforts on facility location problems, reliable location design that hedges against facility failures is still a challenging research topic due to the difficulty associated with modeling an exponential number of possible facility failure scenarios.

Building on the continuum approximation approach, this study proposes a continuous model to solve the reliable supply chain location problem under general facility failure probabilities. Compared to discrete models, this continuous model significantly reduces computational complexity and allows for more complex failure mechanisms such as spatially correlated failures. Numerical experiments are conducted to illustrate how the proposed model can be used to optimize facility location design, and how spatial correlations influence the total system cost.

This study also applies the reliable location methodologies to deploy surveillance sensors over transportation networks. We try to address the question on how to deploy surveillance sensors in a transportation network to maximize the utility (or minimizing the estimation error) from integrating disaggregated vehicle information from multiple locations. We have formulated novel mixed-integer mathematical programming models that optimize traffic surveillance benefits under different surveillance effectiveness measures (e.g., traffic volume coverage, vehicle-mile coverage and traffic state estimation error). These models also allow sensors to be subject to probabilistic failures (e.g., due to technical flaws or environmental hazards), even with complex failure patterns such as site-dependent failures. To our best knowledge, no existing literature has addressed these two issues in the context of traffic sensor deployment. Alternative models including single corridor continuum approximation and fixed charge location models are also formulated so as to investigate general properties of this class of problems and provide more flexible methodologies for various relevant applications.

We will develop a set of efficient customized solution algorithms (greedy, interchange, linear relaxation, Lagrangian relaxation) and discuss their performances on the proposed models versus that of well-known commercial software CPLEX. Numerical examples (including full-scale railroad wayside detector location design and Chicago intermodal network sensor location design) are presented to show that these innovative models significantly improve the state of practice, and the proposed algorithms solve these problems efficiently when commercial optimization software fails to provide reasonable solutions. This leads to the development of a piece of stand-alone software, Railroad Wayside Detector Location Solver (RWDLS) (Li and Ouyang, 2007), which has been adopted by the industry. With numerical examples, we also draw managerial insights on how optimal sensor deployment and surveillance benefits vary with the surveillance effectiveness measure and system parameters (e.g., sensor failure patterns and investment budget).

In summary, from an academic point of view, our study advances the knowledge on reliable location design in both supply chain and traffic surveillance contexts; from a practice point of view, it lays the foundation for the development of decision supporting tools (e.g., RWDLS) for the deployment of reliable facility (or sensor) systems.

1.4 Outline

This dissertation is organized as follows. Chapter 2 reviews discrete and continuous modeling techniques for supply chain facility location problems. Traditional discrete models formulate facility location problems into integer linear programs. They in general suffer from huge computational burdens for large problem instants. Continuous models significantly improve

computational tractability by approximating problems in a continuous metric space, and they are more suitable for large-scale instances. Recently, significant disbenefits from probabilistic facility disruptions have been recognized. Hence, researchers have become increasingly interested in reliable versions of these models.

Building on the continuum approximation approach, Chapter 3 proposes a continuous uncapacitated fixed charge location model for reliable facility location design under correlated probabilistic disruptions. This model seeks optimal facility locations to minimize the one-time investment for facility constructions and the long-run transportation costs for serving spatially distributed customers. This model greatly reduces computational complexity and provides flexibility to model general failure patterns (including correlated failures).

Chapter 4 adapts the methodologies for reliable location problems to address traffic sensor location design in a general transportation network. A reliable sensor deployment model is proposed to find optimal locations for advanced vehicle ID identification sensors (which can synthesize disaggregated vehicle information from multiple locations) under potential sensor failures. We consider the cases where the traffic surveillance benefit is from both individual sensor flow coverage (e.g., for traffic volume statistics) and synthesized sensor pairs (e.g., for travel time estimation) and sensors fail independently with an identical failure probability. Efficient solution algorithms are proposed and tested with numerical examples. A simplified version of this model has been encapsulated into a piece of stand-alone software, which have been adopted by the railroad industry.

Chapter 5 extends the sensor location model into a more general framework that incorporates general surveillance effectiveness measures and site-dependent sensor failure probabilities. We define a novel surveillance effectiveness measure that encompasses flow coverage, path coverage and estimation error reduction and formulate the design problem into a compact model. Alternative formulations including fixed-charge location and continuum approximation models are investigated. A range of customized solution algorithms are developed to solve this problem efficiently.

Chapter 6 summarizes this dissertation and recommends a few future research directions.

Chapter 2

Facility Location Problem Review

This chapter reviews several major discrete and continuous facility location models. Discrete models, which are most-commonly seen in facility location literature, formulate facility location problems into integer linear programs. Discrete models can be solved with commercial solvers or customized algorithms if the problem sizes are small. Continuous models significantly improve computational tractability by approximating problems in a continuous metric and are more suitable for large-scale instances. Experiments have shown that the solution quality of continuous models is comparable to that of discrete models if system parameters only vary slowly across the spatial domain.

2.1 Discrete Models

Facility location studies can be traced back to its original formulation in 1909 and the Weber Problem (Weber, 1957). Daskin (1995) and Drezner (1995) have systematically introduced classic discrete location models for deterministic problems including covering problems (Christofides, 1975; Church and ReVelle, 1974), center and median problems (Hakimi, 1964) and fixed-charge location problems (Cornuejols et al., 1977; Mirzain, 1985). These models are later extended to handle reliable problems that allow possible facility failures (Daskin, 1983; Snyder and Daskin, 2005; Cui et al., 2009). All these models are NP hard, and known algorithms (or commercial software) can only solve small-size instances to exact optimal solutions efficiently. Solving large-scale problems generally relies on heuristic algorithms that usually yield near-optimal solutions.

2.1.1 Classic Models

This section, mainly referring to Daskin (1995), reviews a set of classical facility location problems including covering, center, median, fixed charge facility location problems. In all of these problems, customer demand is distributed in a set of nodes $\mathcal{I}$ and each $i \in \mathcal{I}$ generates λ_i units of demand. Facilities can be built at locations in candidate set $\mathcal{J}$ to serve demand.

The set covering problem aims to find the facility location design with minimum number of facilities that can serve all demand. In this problem, a facility can only cover (or serve) a portion of demand. We use $\{a_{ij}\}_{i\in\mathcal{I},j\in\mathcal{J}}$ to represent the coverage relationship such that demand at i can (not) be served by a facility at j if $a_{ij} = 1$ ($a_{ij} = 0$). The binary integer decision variables $\mathbf{x} = \{x_j\}_{j\in\mathcal{J}}$ indicate where to build facilities; i.e., a facility is built at j if $x_j = 1$. The objective is to minimize the total number of facilities that can provide a complete coverage. The mathematical model can be written as follows

$$\min_{\mathbf{x}} \sum_{j\in\mathcal{J}} x_j, \tag{2.1a}$$

subject to

$$\sum_{j\in\mathcal{J}} a_{ij}x_j \geq 1, \forall i \in \mathcal{I}, \tag{2.1b}$$

$$x_j \in \{0, 1\}, \forall j \in \mathcal{J}, \tag{2.1c}$$

When the total coverage requirement is relaxed and budget limit is imposed, the above model (2.1) becomes the maximum covering problem. In this problem, no more than $N < |\mathcal{J}|$ facilities can be built in total due to the budget constraint. A set of binary auxiliary variables $\mathbf{y} = \{y_i\}_{i\in\mathcal{I}}$ are introduced such that $y_i = 1$ indicates that demand at i is covered or $y_i = 0$ otherwise. Note that once facility deployment $\mathbf{x}$ are given, all auxiliary variables are uniquely determined. This is also true for other auxiliary variables in all the following models in this chapter. Now the objective is to maximize the served demand.

$$\max_{\mathbf{x},\mathbf{y}} \sum_{i\in\mathcal{I}} \lambda_i y_i, \tag{2.2a}$$

subject to

$$\sum_{j\in\mathcal{J}} x_j \leq N, \tag{2.2b}$$

$$\sum_{j\in\mathcal{J}} a_{ij}x_j \geq y_i, \forall i \in \mathcal{I}, \tag{2.2c}$$

$$x_j \in \{0,1\}, \forall j \in \mathcal{J}, \tag{2.2d}$$

$$y_i = \{0,1\}, \forall i \in \mathcal{I}. \tag{2.2e}$$

Model (2.2) can be adapted to other problems if travel distance is taken into account. The center problem tries to minimize the maximum travel distance of a customer and is suitable for locating facilities of public services such as hospitals and schools where the service level and equity are priorities. Let d_{ij} denote the travel distance from customer i to facility j, and the auxiliary variables become $\mathbf{y} = \{y_{ij}\}_{i\in\mathcal{I},j\in\mathcal{J}}$, where $y_{ij} = 1$ if customer i is served by facility j. The center problem is formulated as follows.

$$\min_{\mathbf{x},\mathbf{y}} W, \tag{2.3a}$$

subject to

$$\sum_{j\in\mathcal{J}} x_j \leq N, \tag{2.3b}$$

$$\sum_{j\in\mathcal{J}} y_{ij} = 1, \forall i \in \mathcal{I} \tag{2.3c}$$

$$y_{ij} \leq x_j, \forall i \in \mathcal{I}, j \in \mathcal{J}, \tag{2.3d}$$

$$W \geq \sum_{j\in\mathcal{J}} d_{ij}y_{ij}, \forall i \in \mathcal{I}, \tag{2.3e}$$

$$x_j \in \{0,1\}, \forall j \in \mathcal{J}, \tag{2.3f}$$

$$y_{ij} = \{0,1\}, \forall i \in \mathcal{I}, j \in \mathcal{J}. \tag{2.3g}$$

As compared to social benefit or equity that concerns public sectors, private agencies who provide delivery services to customers are concerned more about their own profits. Thus

reducing the operating cost, which is closely related to the total travel distance between facilities and customers, is the primary consideration. This fits the median problem model that aims to minimize the total travel distance for all trips (or deliveries). After minor modification of model (2.3), the median problem is formulated as follows.

$$\min_{\mathbf{x},\mathbf{y}} \sum_{i\in\mathcal{I}} \sum_{j\in\mathcal{J}} \lambda_i d_{ij} y_{ij}, \tag{2.4a}$$

subject to

$$\sum_{j\in\mathcal{J}} x_j \leq N, \tag{2.4b}$$

$$\sum_{j\in\mathcal{J}} y_{ij} = 1, \forall i \in \mathcal{I} \tag{2.4c}$$

$$y_{ij} \leq x_j, \forall i \in \mathcal{I}, j \in \mathcal{J}, \tag{2.4d}$$

$$x_j \in \{0,1\}, \forall j \in \mathcal{J}, \tag{2.4e}$$

$$y_{ij} = \{0,1\}, \forall i \in \mathcal{I}, j \in \mathcal{J}. \tag{2.4f}$$

In addition to operating cost, one-time facility investment is sometimes a significant component of the total system cost. We can prorate one-time facility investment over years or aggregate long-term operating cost together to unify facility cost and operating cost. It is intuitive that these two costs form a trade-off; i.e., the more facilities, the better accessibility customers will have and thus the less operating cost. The fixed charge facility location problem is looking for a balance of this trade-off in order to minimize total system cost. Let f_j denote the unified one-time building cost of a facility at j. The uncapacitated fixed charge facility location (UFL) model can be obtained by adding facility cost to objective (2.4a),

$$\min_{\mathbf{x},\mathbf{y}} \sum_{j\in\mathcal{J}} f_j x_j + \sum_{i\in\mathcal{I}} \sum_{j\in\mathcal{J}} \lambda_i d_{ij} y_{ij}, \tag{2.5a}$$

subject to

$$(2.4b) - (2.4f). \tag{2.5b}$$

Constraint (2.4b) can be relaxed in the UFL model if the consideration of minimizing the total system cost dictates the allocation of the budget. In problems where the facility capacities restrict the location design, we add one more constraint to the model (2.5),

$$\sum_{j \in \mathcal{J}} \lambda_i y_{ij} \leq k_j x_j, \forall j \in \mathcal{J},$$

where k_j is the capacity of facility at j. The new model is called capacitated fixed charge location problem.

Models (2.1)-(2.5) lay the foundation of many location models that have been used in locating public and private facilities in various application contexts. These models can be extended in a variety of ways to deal with more realistic situations. Multiobjective optimization techniques are needed in case more than one conflicting or competing objectives are present. Distinguishing facility types is necessary when the system involves multiple types of facilities in a hierarchy. In complex supply chain systems, multiple stages of service and interactions among facilities may be considered. Sometimes, the distribution cost can not be simply measured by the direct shipment distance between facilities and customers, but detailed delivery routing has to be taken into account.

2.1.2 Reliable Models

The traditional facility location models assume that facilities, once built, will remain operational forever (or at least within the life cycle). However, in reality, one or more of the facilities may become unavailable from time to time—for example, due to adverse weather, natural disasters, labor action, or failure of a related infrastructure system. Well-known examples include the 2005 Hurricane Katrina that idled all industrial and transportation facilities in the entire U.S. Gulf Coast region (Godoy, 2007), the 2002 west-coast port lockout that strangled all U.S. freight shipment routes and supply lines (D'Amico, 2002), and the 2003 power outage that disabled all transportation systems in the New England area (Schewe, 2004). In addition, transportation infrastructure (such as surveillance sensors — for real-time information provision and traffic management) presents inferior performance under adverse environments; for example, more than 40% of the loop detectors on California highways are not functioning properly at any time (Rajagopal and Varaiya, 2007). Performance of the more advanced radio frequency identification systems is often impaired by factors such as radio frequency interference (Carbunar et al., 2005).

Early models considering system uncertainties focus on mitigating the facility congestions from stochastic demand by increasing the system availability through redundant coverage

(Daskin, 1982, 1983; Revelle and Hogan, 1989; Ball and Lin, 1993; Batta et al., 1989). Recently, facility disruptions due to unexpected events gains more attentions (Snyder and Daskin, 2005; Berman et al., 2007; Cui et al., 2009). Kleindorfer and Saad (2005) discussed the difference of these two types of risks and addressed conceptual strategies to counteract risks of facility disruptions. Mathematical models have been developed to determine reliable facility locations hedging against adverse impact of possible facility disruptions. Snyder and Daskin (2005) studied the reliable uncapacitated fixed charge location problem, RUFL, assuming that facility disruptions occur independently with equal probability. The problem is formulated into a mixed integer program and solved with Lagrangian relaxation. Cui et al. (2009) further developed mixed integer program models to allow site-dependent disruption probabilities. Compared with traditional UFL, these new models have significantly improved system reliability and reduced the expected overall cost across normal and failure scenarios. These models are recently applied to deploy sensors for network traffic surveillance (Ouyang et al., 2009; Li and Ouyang, 2010).

In reliable models, the failure of a supply chain facility will force its customers to either travel longer distances to obtain service from another facility, or give up service and incur a penalty. Either way, system operation cost increases and customer satisfaction deteriorates. Traffic sensor failures will decrease traffic flow network coverage and compromise real-time traffic surveillance benefit (e.g., estimation of traffic volume, speed, and travel time). This may lead to significant societal disbenefits due to ineffective traffic control practice. Such adverse effects may be further exacerbated if multiple facilities fail simultaneously. Hence, planning of facilities requires careful consideration about possible failure scenarios such that the facility location design not only is optimized for the normal non-failure scenario, but also hedges against potential cost increase (or benefit reduction) under rare and unexpected disruptions.

Reliable Models with Identical Failure Probability

Model (2.2) can be extended to a reliable problem in the following way. Each facility now may fail independently with an identical failure probability q. We allow multiple facilities at the same location (i.e., x_j can be greater than one) as back-ups to each other. Demand at a location is served if and only if at least one functioning facility covers it. The binary auxiliary variables become $\mathbf{y} = \{y_{ir}\}_{i\in\mathcal{I},r=0,1,\cdots,N-1}$ such that $y_{ir} = 1$ indicates that demand at i is covered by no less than $r+1$ facilities. The objective is to maximize the total expected served demand. The mathematical model can be written as follows

$$\max_{\mathbf{x},\mathbf{y}} \sum_{i\in\mathcal{I}} \sum_{r=0}^{N-1} (1-q)q^r \lambda_i y_{ir}, \tag{2.6a}$$

subject to

$$\sum_{j\in\mathcal{J}} x_j \leq N, \tag{2.6b}$$

$$\sum_{j\in\mathcal{J}} a_{ij} x_j \geq \sum_{r=0}^{N-1} y_{ir}, \forall i \in \mathcal{I}, \tag{2.6c}$$

$$x_j \in \{0, 1, 2, \cdots, N\}, \forall j \in \mathcal{J}, \tag{2.6d}$$

$$y_{ir} = \{0, 1\}, \forall i \in \mathcal{I}, r = 0, 1, \cdots, N-1. \tag{2.6e}$$

In the same way, (2.4) can be adapted to a reliable problem with independent and identically distributed facility failure probabilities. For each facility j, customers can be partitioned to levels starting with 0 such that customers at level r have other r closer facilities and can be served by j only if all these r facilities fail. Accordingly, we introduce binary auxiliary variables $\mathbf{y} = \{y_{ijr}\}_{i\in\mathcal{I},j\in\mathcal{J},r=0,1,\cdots,N-1}$ such that $y_{ijr} = 1$ indicates that demand at i is served by facility j at level r. To guarantee every customer gets service, N_N emergency facilities that will never fail are installed at locations among the emergency candidate location set $\mathcal{J}_N$ while no more than N_F regular fallible facilities with failure probability q can be installed at locations among $\mathcal{J}_F$, where $\mathcal{J}_N \bigcup \mathcal{J}_F = \mathcal{J}$. The cost for emergency facilities to serve customers can be alternatively interpreted as penalty cost when these customers do not receive regular service. The model formulation is

$$\max_{\mathbf{x},\mathbf{y}} \sum_{i\in\mathcal{I}} \lambda_i \left[\sum_{j\in\mathcal{J}_N} \sum_{r=0}^{N-1} q^r \lambda_i d_{ij} y_{ijr} + \sum_{j\in\mathcal{J}_F} \sum_{r=0}^{N-1} (1-q)q^r d_{ij} y_{ijr} \right], \tag{2.7a}$$

subject to

$$\sum_{j\in\mathcal{J}_F} x_j \leq N_F, \tag{2.7b}$$

$$\sum_{j\in\mathcal{J}_N} x_j = N_N, \tag{2.7c}$$

$$\sum_{j\in\mathcal{J}} y_{ijr} + \sum_{j\in\mathcal{J}_N} \sum_{s=0}^{r-1} y_{ijs} = 1, \forall i \in \mathcal{I}, r = 0, 1, \cdots, N-1, \tag{2.7d}$$

$$y_{ijr} \leq x_j, \forall i \in \mathcal{I}, j \in \mathcal{J}, r = 0, 1, \cdots, N-1, \tag{2.7e}$$

$$\sum_{r=0}^{N-1} y_{ijr} \leq 1, \forall i \in \mathcal{I}, j \in \mathcal{J}, \tag{2.7f}$$

$$x_i \in \{0, 1\}, \forall i \in \mathcal{I}, \tag{2.7g}$$

$$y_{ijr} = \{0, 1\}, \forall i \in \mathcal{I}, j \in \mathcal{J}, r = 0, 1, \cdots, N-1. \tag{2.7h}$$

In a similar manner, other models can be also modified to the reliable version with independently and identical failure probabilities.

Reliable Models with Site-Dependent Failure Probability

The assumption that all facility locations have identical failure probabilities might not represent practical situations. In reality, facilities closer to hazard sources are more vulnerable than those far away. For example, in hurricane related disasters, facilities located in the Gulf coast area (TX, LA, MS, AL and FL) will have a much higher chance of disruption than those in other locations. Cui et al. (2009) developed a reliable fixed-charge location model to handle site-dependent failure probabilities. The problem setting is the same as (2.7) except that (a) facility at j has a site-dependent failure probability q_j, (b) no explicit budget constraint is imposed but building a facility at j will incur a fixed cost f_j, (c) and $\mathcal{J}_N$ is a singleton $\{J\}$ with $q_J = 0$. Assume each customer can potentially go to at maximum R facilities for service, and if they all fail the customer goes to the emergency facility J (or equivalently subject to certain penalty cost). A second set of auxiliary variables $\mathbf{P} = \{P_{ijr}\}_{i \in \mathcal{I}, j \in \mathcal{J}, r=0,\cdots,R}$ are introduced such that P_{ijr} represents the probability that customer i is served by facility j at its r^{th} choice. The model can be written as follows.

$$\textbf{(NSPC)} \quad \min_{\mathbf{xy},\mathbf{P}} \sum_{j \in \mathcal{J}} f_j x_j + \sum_{i \in \mathcal{I}} \sum_{j \in \mathcal{J}} \sum_{r=0}^{R} \lambda_i d_{ij} P_{ijr} y_{ijr} \tag{2.8a}$$

subject to

$$\sum_{j \in \mathcal{J} \setminus \{J\}} y_{ijr} + \sum_{s=0}^{r-1} y_{iJs} = 1, \forall i \in \mathcal{I}, r = 0, 1, \cdots, R, \tag{2.8b}$$

$$y_{ijr} \leq x_j, \forall i \in \mathcal{I}, j \in \mathcal{J}, r = 0, 1, \cdots, R, \tag{2.8c}$$

$$\sum_{r=0}^{R} y_{ijr} \leq 1, \forall i \in \mathcal{I}, j \in \mathcal{J}, \tag{2.8d}$$

$$P_{ij0} = (1 - q_j), i \in \mathcal{I}, j \in \mathcal{J} \tag{2.8e}$$

$$P_{ijr} = (1 - q_j) \sum_{k \in \mathcal{J} \setminus \{J\}} \frac{q_k}{1 - q_k} P_{ik(r-1)} y_{ik(r-1)}, \quad \forall i \in \mathcal{I}, j \in \mathcal{J}, r = 0, 1, \cdots, R \tag{2.8f}$$

$$x_i \in \{0, 1\}, \forall i \in \mathcal{I}, \tag{2.8g}$$

$$y_{ijr} = \{0, 1\}, \forall i \in \mathcal{I}, j \in \mathcal{J}, r = 0, 1, \cdots, R. \tag{2.8h}$$

2.1.3 Algorithm Discussion

All the aforementioned models are (or can be converted to) linear mixed-integer programs. Small-size instances of these models can be solved to exact optimality by commercial software (e.g., CPLEX) or methodologies such as the branch and bound method. However, since all these models are known to be NP hard (which means that solution complexity increases exponentially with the problem size), heuristic algorithms are often applied to obtain near-optimal solutions for large-scale instances.

Greedy heuristic is a simple algorithm to find a good feasible solution. The greedy algorithm selects facility locations sequentially. At each step, it enumerates the marginal objective improvement by adding any extra facility location(or any few extra facilities) and selects the one (or few) bringing in the best improvement. This is repeated until the budget is exhausted or no additional facility can bring in any marginal improvement. In all these aforementioned models, once facility locations are given, it is very easy to evaluate the objective, and the enumeration space for next facility (or next few facilities) is not too large. Thus the greedy algorithm can efficiently identify the next best facility location (or few locations) and hence has very small computational burdon. Greedy heuristic is widely applied to many practical problems not only because of its simplicity but also due to its reasonable practical

performance (Feige, 1998; Ageev and Sviridenko, 1999).

The greedy algorithm can be improved by an interchange heuristic (or neighborhood search). Given a feasible solution, the interchange heuristic searches for a better solution within a certain neighborhood, e.g., only allowing changing one or two facility locations. This approach can be repeated until no better neighbor can be found. The neighborhood size needs to be carefully selected. If it is too small, the algorithm can be easily trapped at some local optimum; if it is too large, the computation will be too time-consuming. With proper selection of the neighborhood, empirical experience shows that algorithms combining greedy and interchange heuristics can often yield very good solutions for practical problems.

However, these heuristic algorithms can not give any performance bound to quantify the solution quality. Linear programming relaxation can be used to find an optimality gap, which nevertheless is often very loose. Instead, Lagrangian relaxation, a dual algorithm, is usually adopted in location problems to obtain a tighter optimality gap. The Lagrangian relaxation usually decomposes a location problem into subproblems which each is simple to solve. Its overall computational complexity is quite reasonable. Furthermore, a solution of a relaxed problem, though maybe not feasible for the original problem, can be easily modified into a feasible solution with certain heuristics, and from experience this feasible solution is probably very close to the true optimum. Due to all these advantages, the Lagrangian relaxation algorithm has been frequently adopted by researchers in this field.

2.2 Continuum Approximation (CA) Models

The CA models (Newell, 1971, 1973; Daganzo, 1984a,b; Daganzo and Newell, 1986; Ouyang and Daganzo, 2006) are often developed to provide good approximate solutions to large-scale logistics problems in various contexts (Hall, 1984, 1986, 1989; Campbell, 1993a,b; Daganzo, 1999; Dasci and Verter, 2001). See Langevin et al. (1996) and Daganzo (2005) for reviews.

Early CA studies stem from seeking simplified algorithms for the lost size problem with variable demand (Newell, 1971). Figure 2.1 illustrates this problem. The cumulative demand is denoted by the solid curve $D(t)$ over the finite time horizon $\mathbf{T} := [t_0, t_{\max})$. Curve $R(t)$ is the count of received items and jumps of $R(t)$ at time points $t_0, t_1, \cdots, t_{\max -1}$ represent discrete orders with ordering amounts equal to the step heights. Items are dispatched as soon as an order takes place and the lead time is ignored. Each dispatch at time t incurs a fixed cost $f(t)$ and covers demand up to the next ordering period. Each received item is stored in a warehouse and will be consistently incuring inventory holding cost c per unit time until consumed by the demand. Thus the total inventory holding cost is proportional to the shaded area with factor c. This problem aims to design optimal ordering time points and

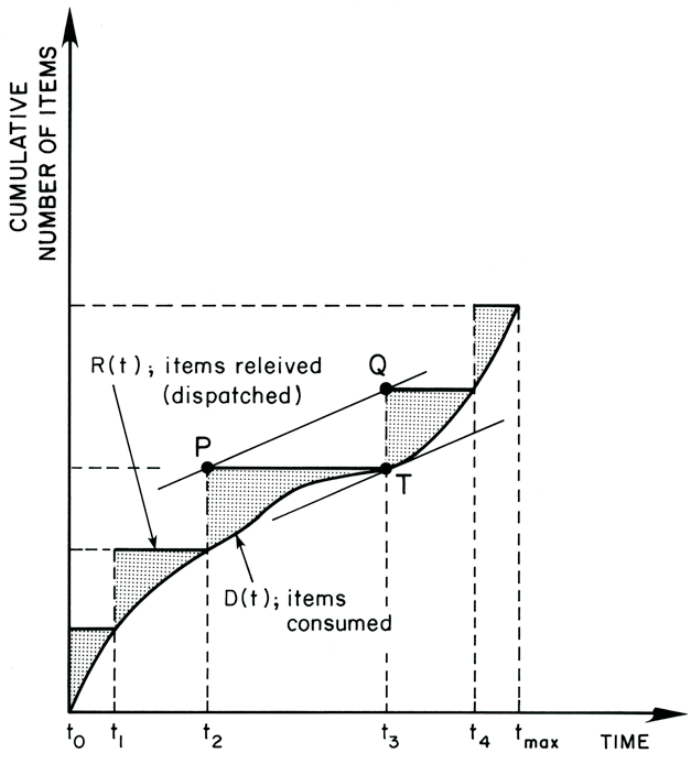


Figure 2.1: Lost size problem with variable demand.[a]

[a]Source: Daganzo (2005).

corresponding ordering amounts to minimize the total system cost (including fixed ordering cost and inventory cost).

Numerical approaches have been developed to solve the lot size problem. However, they may take excessively long time, especially in the early 1970s when computer technologies were limited. Newell (1971) proposed a CA approach that can solve such a problem with slowly-varying setting (i.e., $D'(t_i) \approx D'(t_{i+1})$ and $f(t_i) \approx f(t_{i+1})$) to a near-optimal solution in much shorter time. Assume that the ordering time points of an optimal solution are $t_0, t_1, \cdots, t_{\max -1}$. The total cost in interval $\mathbf{T}_i := [t_{i-1}, t_i)$ is

$$C_i := \int_{\mathbf{T}_i} \left[\frac{f(t_i)}{A_s(t)} + \frac{cA_s(t)}{2} D'(t'_i) \right] dt, \tag{2.9}$$

where $A_s(t)$ is a step function such that $A_s(t) = (t_i - t_{i-1})$, if $t \in \mathbf{T}_i$, and $t'_i \in \mathbf{T}_i$ satisfies that $\frac{1}{2}(t_i - t_{i-1})^2 D'(t'_i)$ equals the shaded area in this interval.

Based on the assumption of slow-varying setting, the key of this CA approach is to approximate $D'(t'_i)$ with $D'(t)$ and $A_s(t)$ with a continuous function $A(t)$. Then the total cost can be approximated as

$$C :\approx \int_{\mathbf{T}} \left[\frac{f(t)}{A(t)} + \frac{cA(t)}{2} D'(t) \right] dt. \tag{2.10}$$

Clearly, the $A(t)$ that minimizes (2.10) minimizes the integrand at every t; thus:

$$A(t) = \left[2f(t)/(cD'(t)) \right]^{1/2}. \tag{2.11}$$

Since the continuous function $A(t)$ does not directly specify discrete ordering time points, the discretization mehod illustrated in Figure 2.2 is taken to determine them. Draw a 45° line starting at the origin t_0 and find a horizontal segment from a point on the vertical axis, such as P_1 in the figure, to the intersection with the 45° line. The elevation of P_1 should be such that the area below the segment equals the area above it. The abscissa of the right ending point of the segment locates the next ordering time, t_1. The construction is repeated to find every ordering time.

The above one-dimensional CA framework has been extended to solve the facility location problem (Newell, 1973; Daganzo and Newell, 1986). For example, Figure 2.1 can be interpreted as a specific facility location problem in one-dimensional space $\mathbf{T}$. Customer demand is distributed over $\mathbf{T}$ and cumulates to $D(t)$ in $[t_0, t)$. Facilities each with opening cost $f(t)$ are built at locations $t = t_0, t_1, ..., t_{\max -1}$. A facility at t_i serves demand in $\mathbf{T}_i$ and the travel cost for a unit demand at t to reach the service equals $c[t - t_i]$. Thus the

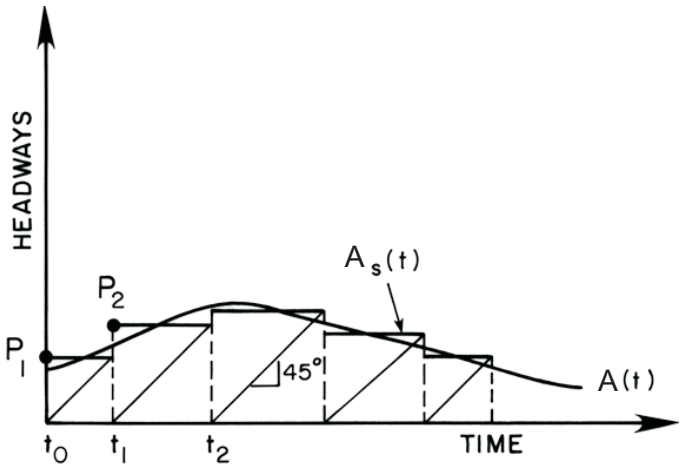


Figure 2.2: Discretization of $A(t)$[b].

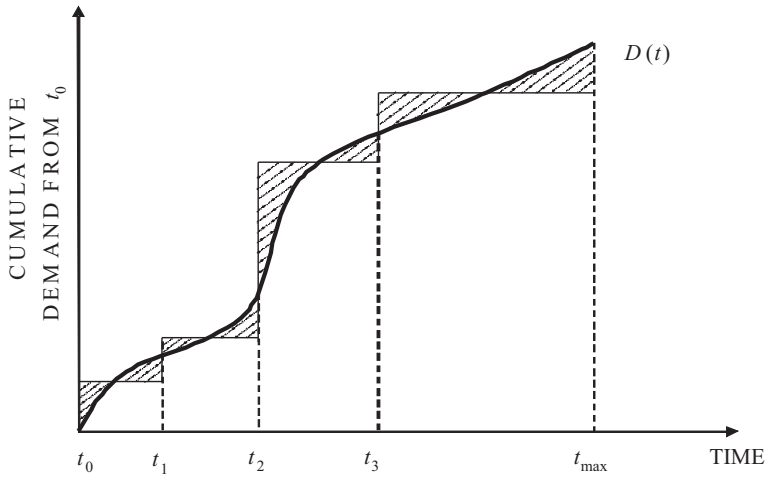


Figure 2.3: One-dimensional uncapacitated fixed-charge location problem.

[b]Source: Daganzo (2005).

total travel cost can be too represented as the shaded area and the summation of all costs in $\mathbf{T}_i$ can be written as C_i (2.9). Then the same CA approach can be taken to solve the problem. This model may seem unreasonable due to the assumption that each customer is always served by its immediate left facility. Nevertheless, it can be easily modified to the well-known UFL problem by letting each customer be served by its closest facility. Figure 2.3 shows this one-dimensional UFL. Despite that the travel cost or the shaded area is specified differently, the CA framework is still applicable to the new problem.

This one-dimensional CA approach has been generalized to solve UFL problems in a two-dimensional space (Daganzo and Newell, 1986; Ouyang and Daganzo, 2006). Figure 2.4 describes this problem. In this two-dimensional space $\mathbf{T}$, a set of facilities, each denoted by t_i, serve distributed customer that has density $D'(t), \forall t \in \mathbf{T}$. Since each customer is served by its closest facility, $\mathbf{T}$ is tessellated into regions such that facility t_i serves customers in region $\mathbf{T}_i$ (which is also called Voronoi Tessellation). Similar to (2.9), the total cost in $\mathbf{T}_i$ can be written as follows

$$C_i = \int_{\mathbf{T}_i} \left[\frac{f(t_i)}{A_s(t)} + \alpha_i \sqrt{A_s(t)} c D'(t_i') \right] dt, \tag{2.12}$$

where $A_s(t)$ is the area size of $\mathbf{T}_i$ if $t \in \mathbf{T}_i$, α_i is a scaler such that $\alpha_i \sqrt{A_s(t)}$ is the average distance from a unit demand in $\mathbf{T}_i$ to t_i and t_i' is a certain point in $\mathbf{T}_i$ such that $D'(t_i')$ is the average demand over $\mathbf{T}_i$. If everything varies slowly in this space and the space is large enough such that the boundary shape does not affect too much the tessellation for an optimal deployment, then each $\mathbf{T}_i$ is not too different from a circle and thus each $\alpha_i \approx \frac{2}{3\sqrt{\pi}}$. Then the total cost C over the whole space $\mathbf{T}$ can be approximated in the same way as (2.10),

$$C \approx \int_{\mathbf{T}} \left[\frac{f(t)}{A(t)} + \frac{2c\sqrt{A(t)}}{3\sqrt{\pi}} D'(t) \right] dt. \tag{2.13}$$

Again, the minimizer of (2.13) can be solved for each integrand as follows,

$$A(t) = \left[\frac{3\sqrt{\pi} f(t)}{c D'(t)} \right]^{2/3}. \tag{2.14}$$

The integrand of (2.13) also represents that each local neighborhood around $t \in \mathbf{T}$ is approximated with a plane with homogeneous settings $f(t)$ and $D'(t)$, and in equation (2.14) the facility density in this plane is simply determined by an economic order quantity (EOQ) model.

Though it is not as simple to discretize $A(t)$ in the two-dimensional space, a disc model

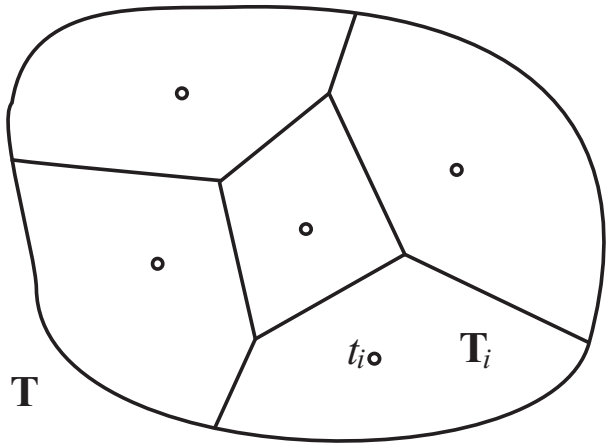


Figure 2.4: Two-dimensional uncapacitated fixed-charge location problem.

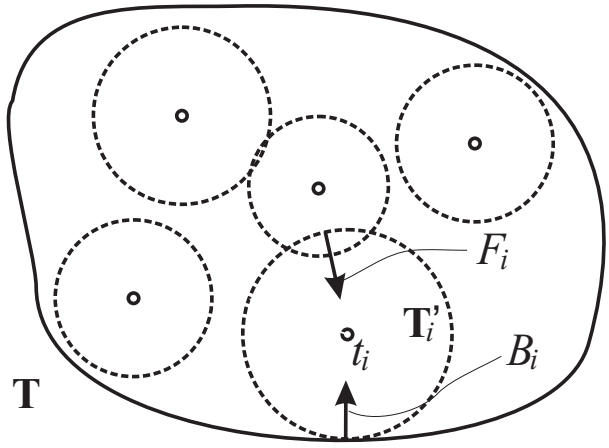


Figure 2.5: Disc model illustration.

has been developed to convert $A(t)$ to discrete facility locations (Ouyang and Daganzo, 2006). This is illustrated in Figure 2.5. We can determine near-optimal number of facilities N as the closest integer to $\int_{\mathbf{T}} 1/A(t)dt$. We distribute N seed facilities $\{t_i\}$ at random locations in $\mathbf{T}$. Assign each facility t_i a disc $\mathbf{T}'_i$ centered at t_i whose size approximately equals $A(t_i)$. Then introduce a terminal force F_i that repels $\mathbf{T}'_i$ away from other disc(s) overlapped with it, and a boundary force B_i that keeps $\mathbf{T}'_i$ within space $\mathbf{T}$. At each step, these forces nudge each $\mathbf{T}'_i$ to a new position and t_i, which is bonded to the center of $\mathbf{T}'_i$, also moves with it. After this movement of t_i, $A(t_i)$ changes accordingly and the size of $\mathbf{T}'_i$ is too updated. Such movements repelled by F_i and B_i are repeated until the position of each t_i converges, which yields the near-optimal locations for facility installations.

Based on this framework, Cui et al. (2009) developed a reliable CA model as an alternative for solving large-scale RUFL problems, and compared its performance with that of its discrete counterparts. Chapter 3 will generalize this reliable CA model by accommodating correlated facility disruptions.

2.3 List of Symbols

a_{ij}: Demand at i can (not) be served by a facility at j if $a_{ij} = 1$ ($a_{ij} = 0$)
$A(t)$: Continuous function to approximate $A_s(t)$
$A_s(t)$: Size (or length) of T_i that contains t
B_i: Boundary force that keeps $\mathbf{T}'_i$ within space $\mathbf{T}$
c: Holding cost factor
C_i: Total cost within $\mathbf{T}_i$
C: Total cost within $\mathbf{T}$
d_{ij}: Travel distance from i to j
$D(t)$: Cumulative demand at t
f_j: Unified one-time building cost of a facility at j
$f(t)$: Fixed cost at t
F_i: Terminal force that repels $\mathbf{T}'_i$ away
i: Index of a demand node or a time point
$\mathcal{I}$: Set of all demand nodes
j: Index of a candidate location
$\mathcal{J}$: Set of all candidate locations
$\mathcal{J}_N$: Set of emergency candidate locations
$\mathcal{J}_F$: Set of regular candidate locations
N: Maximum number of facilities that the budget allows to build
N_N: Number of emergency facilities that will never fail
N_F: Number of regular facilities that may probabilistically fail
P_{ijr}: Probability that customer i is served by facility j at its r^{th} choice
$\mathbf{P}$: $\{P_{ijr}\}_{i\in\mathcal{I},j\in\mathcal{J},r=0,\cdots,R}$
q: Site-independent failure probability of a regular facility
q_j: Site-dependent failure probability of a regular facility at j
r: Index of a facility number or a customer service level
R: Maximum number of facilities that each customer can potentially visit
$R(t)$: Count of received items
t: Point $\in \mathbf{T}$
t_0: Initial point in a one dimensional space (or time range)
$t_{\max}$: Ending point in a one-dimensional space (or time range)
$\mathbf{T}$: One-dimensional or two-dimensional space or time span
$\mathbf{T}_i$: Connected subset of $\mathbf{T}$

$\mathbf{T}'_i$: Disc centered at t_i
$\mathbf{x} = \{x_j\}$: $x_j = 1$ ($x_j = 0$) if a facility is (not) built at j
y_i: $y_i = 1$ ($y_i = 0$) if demand at i is (not) covered
y_{ij}: $y_{ij} = 1$ ($y_{ij} = 0$) if demand at i is (not) served by a facility at j
y_{ir}: $y_{ir} = 0$ ($y_{ir} = 1$) if demand at i is covered by (no) less than $r + 1$ facilities
y_{ijr}: $y_{ijr} = 1$ ($y_{ijr} = 0$) if demand at i is (not) served by facility j at level r
$\mathbf{y}$: $\{y_i\}_{i\in\mathcal{I}}$, $\{y_{ij}\}_{i\in\mathcal{I},j\in\mathcal{J}}$, $\{y_{ir}\}_{i\in\mathcal{I},r=0,1,\cdots,N-1}$ or $\{y_{ijr}\}_{i\in\mathcal{I},j\in\mathcal{J},r=0,1,\cdots,N-1}$
λ_i: Amount of demand at i

Chapter 3

A Continuum Approximation Approach to Reliable Facility Location Design Under Correlated Probabilistic Disruptions

Reliable facility location problems have been studied recently with both discrete and continuous modeling techniques. However, complex facility failure mechanisms such as spatial correlation have not yet been addressed. Due to the formidable complexity associated with such complex failure mechanisms, discrete facility models are not suitable to solve (or even model) such problems. Building on the continuum approximation approach, this chapter proposes a reliable model for the uncapacitated fixed charge location problem (UFL), which seeks optimal facility locations to minimize the one-time investment for facility constructions and the long-run transportation costs for serving spatially distributed customers. This model greatly reduces computational complexity and provides flexibility to model general failure patterns (including correlated failures). We have tested this model over different types of numerical examples and useful managerial insights are drawn on how failure correlation impacts the location design.

3.1 Motivation

In the real world, many facility disruption cases exhibit strong spatial correlations, probably because neighboring facilities are likely to be exposed to similar hazards. Such correlations significantly influence the facility failure pattern over space and hence the system operation.

For example, under positive correlations (e.g., due to natural disasters, power grid outages), neighboring facilities are more likely to fail simultaneously, and the customers will find it more costly to reach a functioning facility. In contrast, under negative correlations[a], neighboring facilities tend to back up each other to avoid long distance travels of the customers.

However, to the authors' best knowledge, spatial correlation among facility disruptions has not been addressed in the reliable UFL (RUFL) literature. This chapter aims to fill this gap by developing a reliable facility location design framework that allows correlated and site-dependent facility disruptions. Accounting for such correlations in the discrete location modeling framework generally requires scenario-based formulation, which is computationally prohibitive due to the exponential number of possible scenarios. Hence we build our model upon the continuum approximation approach to estimate and design the complex system. The structure of the spatial correlation is modeled in a variety of ways to provide flexibility in addressing real-world scenarios. Numerical experiments are conducted to illustrate applications of the model. Insights are also drawn through comparisons between the optimal solutions under various spatial correlation patterns and those under independent failures. The impact of disruption correlation on the total system cost (including yearly-prorated facility construction cost, expected annual customer traveling and penalty costs) is found to be significant when both failure probabilities and penalty costs for unserved customer demand are high.

The remainder of the chapter is organized as follows. Section 3.2 introduces the notation and problem definition. Section 3.3 presents the formulation and solution techniques for the CA model. Section 4 presents multiple ways to model spatial correlation under different application contexts. Section 3.5 applies the CA model to numerical examples and draw insights into the impact of correlations.

3.2 Model Formulation

In a two-dimensional space $\mathcal{S} \subseteq \mathbb{R}^2$, the customer demand per unit area is denoted by $\lambda(x)$, $\forall x \in \mathcal{S}$. A facility can be built at any location $x \in \mathcal{S}$ with a fixed opening cost $f(x)$. The decision variables are the number of facilities, N, and their locations $\mathbf{x} := \{x_1, x_2, ..., x_N\} \subseteq \mathcal{S}$. Suppose that the transportation cost for facility j to serve a unit demand at x is $\alpha_t \|x - x_j\|$, where α_t is a constant factor and $\|x - x_j\|$ is the Euclidean distance. We further assume that a customer at x, if served, shall only be served by a facility within a distance $D(x)$; if not

[a]In reality, negative failure correlation is rare but possible. For example, the facilities may compete against each other for limited critical resources (such as material supply or maintenance service), such that the failure of one facility helps other facilities to survive. In the context of terrorist attacks, the failure of one facility may raise alert and help prevent other facilities from being attacked.

served, the customer will incur a penalty cost $\alpha_p D(x)$, where $\alpha_p \geq \alpha_t$. Snyder and Daskin (2005) attributed the penalty cost to lost-sales or emergency-purchases.

We assume that the customers have complete information on facility disruptions[b] and choose facilities for service accordingly. This is different from Cui et al. (2009) where each customer is preassigned to a sequence of prioritized facilities regardless of the failure scenario. We also assume, for simplicity, that the failure scenario does not change during the time that customers are traveling. At any time, customers at x will either visit a functioning facility within distance $D(x)$ if one is available, or bear the penalty cost $\alpha_p D(x)$. The optimal strategy has the following simple property.

Proposition 1. *Given a facility failure scenario, each customer should always visit the closest functioning facility within distance $D(x)$.*

Proof. If a customer visits an operational facility other than the closest one, redirecting this customer to the closest operational facility will always strictly reduce the transportation cost. Thus the original solution cannot be optimal. This completes the proof. □

Given facility location design $\mathbf{x}$, let $\bar{P}(x|\mathbf{x})$ denote the probability for the customer at x not to be served (which occurs if all facilities within distance $D(x)$ from x have failed), and let $P(x, x_j|\mathbf{x})$ denote the probability for this customer to be served by facility j (which occurs if facility j is functioning, $\|x - x_j\| \leq D(x)$, and all facilities closer to x have failed). The values of these probabilities should always satisfy

$$\bar{P}(x|\mathbf{x}) + \sum_{j=1}^{N} P(x, x_j|\mathbf{x}) = 1, \, \forall x \in \mathcal{S}, \tag{3.1}$$

because any customer either receives service or incurs the penalty.

The objective is to minimize the expected overall cost with respect to $\mathbf{x}$, as follows,

$$\min_{\mathbf{x}} \sum_{j=1}^{N} f(x_j) + \alpha_p \int_{x \in \mathcal{S}} \lambda(x) D(x) \bar{P}(x|\mathbf{x}) dx + \alpha_t \int_{x \in \mathcal{S}} \sum_{j=1}^{N} \lambda(x) \|x - x_j\| P(x, x_j|\mathbf{x}) dx. \tag{3.2}$$

The three terms in (3.2) respectively represent the fixed facility opening costs, the expected penalty costs for unserved demands and the expected transportation costs for served demands.

[b]This assumption is reasonable given the rapid advancement of modern information technologies (such as Internet- and PDA-enabled service applications). It may not be totally realistic, however, if information availability is limited in certain situations (e.g., catastrophic disaster).

Following the ideas in Cui et al. (2009), (3.2) can be transformed by partitioning $\mathcal{S}$ into service areas. From the perspective of a generic facility j, every customer in $\mathcal{S}$ can be assigned a service rank $r \in \{0, 1, 2, ...\}$ if facility j is the $(r+1)^{\text{th}}$ nearest facilities to this customer. We define the *rank-r service area* of facility j, $\mathcal{A}_{j,r}$, as the subset of customers who are assigned a rank r by facility j. Obviously, the definition of $\{\mathcal{A}_{j,r}, \forall j, r\}$ are purely based on the facility locations $\mathbf{x}$. For any j, $\{\mathcal{A}_{j,r}, \forall r\}$ forms a non-overlapping partition of $\mathcal{S}$ when boundaries are ignored, i.e.,

$$\bigcup_{r=0}^{\infty} \mathcal{A}_{j,r} = S \text{ and } \mathcal{A}_{j,r} \bigcap \mathcal{A}_{j,r'} = \emptyset, \forall r \neq r'.$$

With this, (3.2) can be rewritten as follows,

$$\min_{\mathbf{x}} \sum_{j=1}^{N} f(x_j) + \alpha_p \int_{x \in \mathcal{S}} \lambda(x) D(x) \bar{P}(x|\mathbf{x}) dx + \alpha_t \sum_{j=1}^{N} \sum_{r} \int_{x \in \mathcal{A}_{j,r}} \lambda(x) \|x - x_j\| P(x, x_j|\mathbf{x}) dx. \tag{3.3}$$

For notation convenience, from now on we will use $\bar{P}(x)$ and $P(x, x_j)$ to represent $\bar{P}(x|\mathbf{x})$ and $P(x, x_j|\mathbf{x})$ respectively.

3.3 Continuum Approximation Framework

This section presents a continuum approximation approach to the RUFL problem. Section 3.3.1 first discusses the optimal solution to an idealized case where the problem is IHI; i.e., $\mathcal{S}$ is an *infinite* and *homogeneous* plane and the facilities fail *independently*. Building on the results for IHI, Section 3.3.3 discusses how to incorporate correlated disruptions into the framework, and Section 3.3.4 further develops the continuum approximation (CA) model for the general problem where $\mathcal{S}$ is finite and heterogeneous.

3.3.1 Building Block: The IHI Problem

In an IHI problem, $\mathcal{S}$ is an infinite and homogeneous plane (i.e., $\mathcal{S} = \mathbb{R}^2$), all relevant parameters are constant everywhere (i.e, $D(x) = D$, $f(x) = f$, $\lambda(x) = \lambda$, $\forall x \in \mathcal{S}$), and every facility fails independently with an equal probability $q(x) = q$. Some properties of the optimal solution to IHI have been discussed in the literature. Toth (1959) has proven that for $q = 0$, the total cost is minimized when the *initial service areas* $\{\mathcal{A}_{j,0}, \forall j\}$ form a regular hexagonal tessellation of $\mathcal{S}$ and each facility is located at the centroid of a hexagon. Cui et

al. (2009) showed that the optimal partition for $q > 0$ should also follow the same regular hexagonal tessellation pattern.

The regular hexagonal tessellation pattern and the homogeneity of $\mathcal{S}$ imply that the only decision variable for the optimal design of IHI is the size of the hexagonal initial service area, which we denote by A. Figure 3.1[c] illustrates how the service areas for an arbitrary facility j would partition $\mathcal{S}$. Proposition 2 below shows that all these service areas have the same size.

Proposition 2. *For an IHI problem,* $|\mathcal{A}_{j,r}| = A, \ \forall j, r.$

Proof. Assume first that $\mathcal{S}$ is bounded but sufficiently large so that the boundary effect can be ignored. Each customer has one and only one facility as its r^{th} choice. This implies that the service areas of all different facilities with the same service rank form a mutually exclusive partition of $\mathcal{S}$, i.e.,

$$\bigcup_j \mathcal{A}_{j,r} = \mathcal{S} \quad \text{and} \quad \mathcal{A}_{i,r} \bigcap \mathcal{A}_{j,r} = \emptyset, \forall i \neq j, r. \tag{3.4}$$

Since almost every facility in $\mathcal{S}$ is translationally symmetric, $|\mathcal{A}_{j,r}| = |\mathcal{A}_{i,r}|$, for almost all i, j (except those near the boundary), Equation (3.4) implies that $|\mathcal{A}_{j,r}| \approx |\mathcal{S}|/N$ for all service rank r and facility j, where N is the total number of facilities. When $\mathcal{S} \to \mathbb{R}^2$, the boundary effect can be totally eliminated. Thus $|\mathcal{A}_{j,r}| = |\mathcal{A}_{j,0}| = A, \forall j, r$. This completes the proof. □

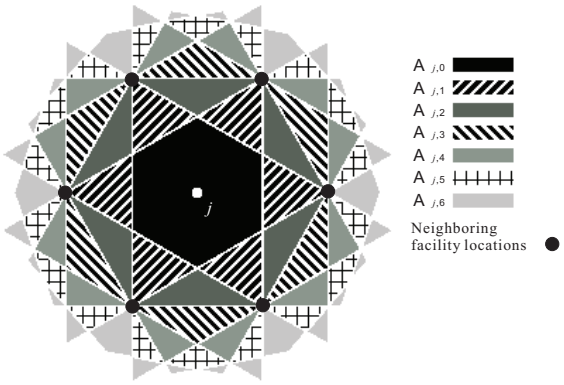


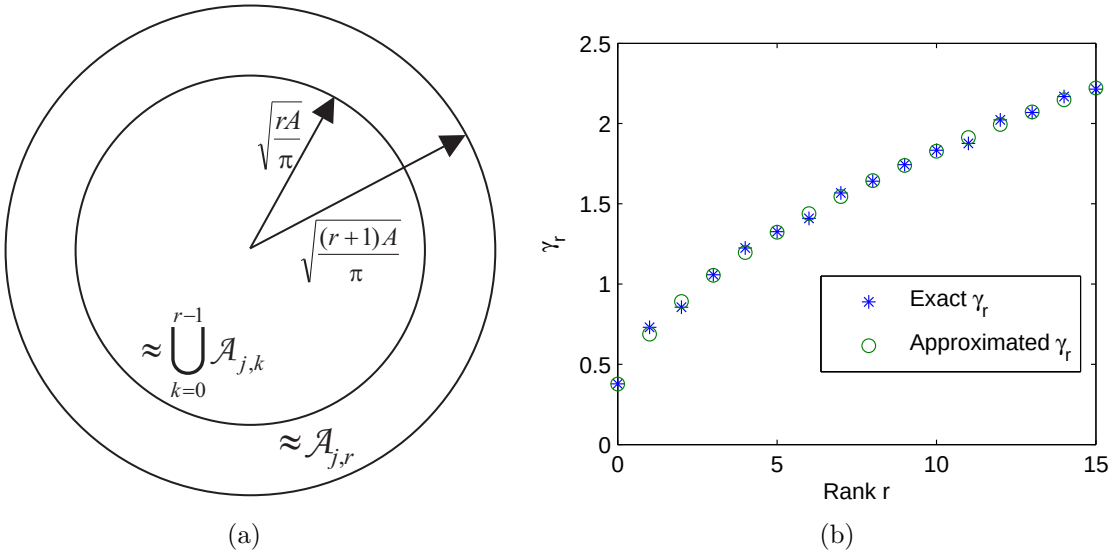
Figure 3.1: Service area partition $\{\mathcal{A}_{j,r}, \forall r\}$ for the IHI problem.

[c]This is adapted from Cui et al. (2009).

The scalability of hexagons on the infinite plane implies that the average travel distance from the customers in $\mathcal{A}_{j,r}$ to facility j is proportional to $A^{1/2}$ and does not depend on j. We denote this average distance by $\gamma_r A^{1/2}$, where constant scalar γ_r can be calculated exactly for all r.[d].

$$\gamma_r \approx \frac{2}{3\sqrt{\pi}} \left[(r+1)^{3/2} - r^{3/2} \right]. \tag{3.5}$$

Figure 3.2(b) plots both the approximation (3.5) and the exact γ_r values. The approximation error is no more than 2% except for $r = 1, 2$, and it vanishes as r increases.



(a) (b)

Figure 3.2: Service cost calculation: (a) Approximation of $\mathcal{A}_{j,r}$ by a ring; (b) Exact and approximated γ_r.

Optimizing objective function (3.3) for infinite and homogeneous $\mathcal{S}$ is equivalent to minimizing the expected total cost per unit area, which includes the unit-area facility opening cost C_f, the unit-area expected penalty cost C_p, and the unit-area expected transportation cost C_t. Obviously,

$$C_f = f/A. \tag{3.6}$$

The rest of this section provides closed-form approximations for C_p and C_t.

Note that $N_D(x)$, the number of facilities that a customer at x can visit within distance D, varies slightly with x. Hence, $\bar{P}(x) = q^{N_D(x)}$ varies with x as well, Let $\theta \in \mathbb{R}_+$ be the average value of $N_D(x)$ across $x \in \mathcal{S}$, and $\bar{P}$ the average value of $\bar{P}(x)$. The customer demand in $\mathcal{S}$ that each facility can potentially reach is $\lambda \pi D^2$, while asymptotically, each facility corresponds to λA customer demand. Hence $\lambda \pi D^2 = \lambda A \cdot \theta$, which yields $\theta = \pi D^2 / A$.

[d]See how the exact values are computed numerically in Section 3.3.2

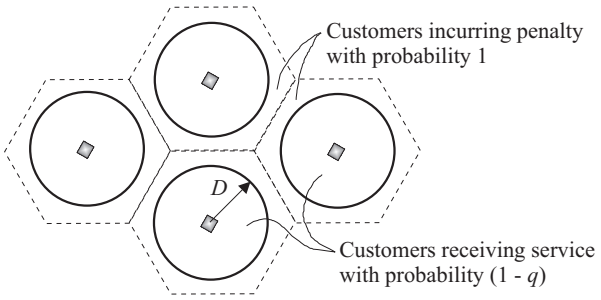


Figure 3.3: Customer partition when $\theta \leq 1$.

If $\theta \leq 1$ ($\pi D^2 \leq A$), the situation is shown in Figure 3.3. Only those customers within distance D from a facility will receive service with probability $(1-q)$; they incur penalty with probability q. All other customers incur penalty with probability 1.[e] Simple geometry yields $\bar{P}$ as follows,

$$\bar{P} = [\pi D^2 \cdot q + (A - \pi D^2) \cdot 1]/A = 1 - (1-q)\theta.$$

More generally, for $\theta > 1$ ($\pi D^2 > A$), customers lie in service areas of different ranks, as shown in Figure 3.1. Exact calculation of $\bar{P}$ is tedious. However, since $N_D(x)$ obviously does not vary significantly across $\mathcal{S}$, $\bar{P}$ can be approximated by $\bar{P} \approx q^\theta$. Hence, we have

$$\bar{P} \approx \begin{cases} q^\theta, & \theta > 1; \\ 1 - (1-q)\theta, & \text{otherwise}; \end{cases} \tag{3.7}$$

and

$$C_p = \alpha_p \lambda D \bar{P}, \tag{3.8}$$

Figure 3.4(a) shows that Equation (3.7) accurately predicts the exact value of $\bar{P}$.[f] The prediction error is almost 0 for $\theta > 3$ (for most realistic cases) and $\theta \leq 1$, and no more than 0.04 for θ around 2.

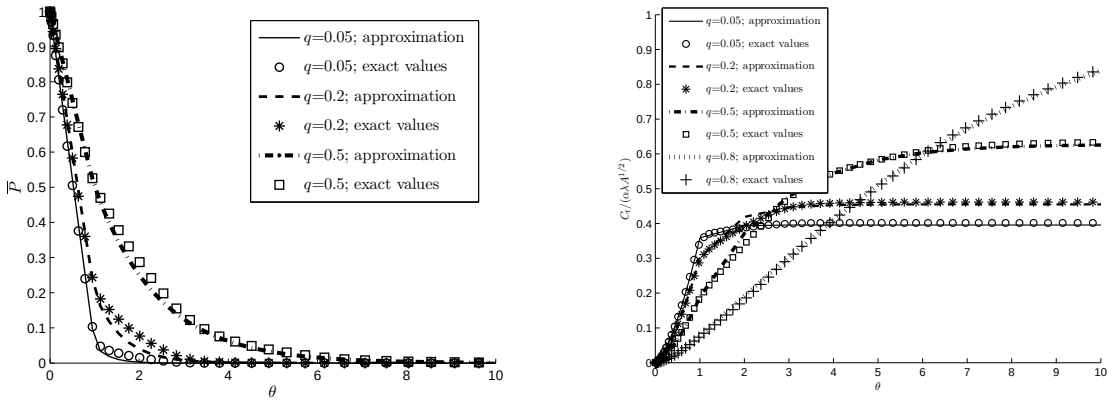
All customers in $\mathcal{A}_{j,r}$ receive service from facility j with equal probability, which we denote by P_r. Due to the independence of facility failures in IHI,

$$P_r = P(x, x_j) = (1-q)q^r, \forall x \in \mathcal{A}_{j,r}. \tag{3.9}$$

In case $\theta > 1$ and $D \to \infty$, Proposition 2 implies that C_t can be directly calculated as

[e] If θ is very close to 1, there may be a very small fraction of customers near the hexagon boundaries with $N_D(x) > 1$. This exception is numerically negligible.

[f] See Section 3.3.2 for details on how these exact values are computed.



(a) (b)

Figure 3.4: Exact and approximated values of (a) $\bar{P}$ and (b) C_t.

follows,

$$C_t = \alpha_t \lambda A^{1/2} \sum_{r=0}^{\infty} P_r \gamma_r. \tag{3.10}$$

For finite D but $\theta > 1$, since θ may not be an integer, interpolation of (3.10) yields

$$C_t \approx \alpha_t \lambda A^{1/2} \left[\sum_{r=0}^{\lfloor \theta \rfloor - 1} P_r \gamma_r + \bar{\theta} P_{\lfloor \theta \rfloor} \gamma_{\lfloor \theta \rfloor} \right], \tag{3.11}$$

where $\lfloor \cdot \rfloor$ is the floor operation and $\bar{\theta} = \theta - \lfloor \theta \rfloor$. For $\theta < 1$ (see Figure 3.3), almost all served customers of facility j lie in the circle within $\mathcal{A}_{j_0}$ and their average distance from the facility is $\frac{2}{3}D$. When facility j does not fail (with probability $(1-q)$), these customers collectively incur service cost $\frac{2}{3}\alpha_t \lambda \pi D^3$. Hence, the expected service cost per unit area can be averaged across $\mathcal{A}_{j_0}$ as follows:

$$C_t = \frac{2}{3}\alpha_t \lambda \pi D^3 (1-q)/A = \frac{2}{3}\alpha_t \lambda A^{1/2} P_0 \sqrt{\theta^3/\pi}. \tag{3.12}$$

Equations (3.11) and (3.12) can be expressed as

$$C_t = \alpha_t \lambda A^{1/2} U(\theta, \mathbf{P}), \tag{3.13}$$

where $\mathbf{P} := \{P_r : \forall r\}$ and

$$U(\theta, \mathbf{P}) \approx \begin{cases} \sum_{r=0}^{\lfloor \theta \rfloor - 1} P_r \gamma_r + \bar{\theta} P_{\lfloor \theta \rfloor} \gamma_{\lfloor \theta \rfloor} & \text{if } \theta \geq 1, \\ \frac{2}{3} P_0 \sqrt{\theta^3/\pi} & \text{otherwise.} \end{cases} \tag{3.14}$$

Obviously, the term $A^{1/2}U(\theta, \mathbf{P})$ represents the expected travel distance for a customer to reach a functioning facility. Figure 3.4(b) compares the approximation formula (3.14) with the exact values.[g] Again, the error is almost 0% for $\theta > 3$ or $\theta \leq 1$, while the maximum percentage error is about 6%.

From (3.6), (3.8) and (3.13), the total cost per unit area for the IHI problem is

$$C := C_f + C_p + C_t = f/A + \alpha_p \lambda D \bar{P} + \alpha_t \lambda A^{1/2} U(\theta, \mathbf{P}). \tag{3.15}$$

In general, the optimal solution A^* does not have a simple analytical form because $\bar{P}$ and $U(\theta, \mathbf{P})$ are both functions of A. Section 3.3.4 introduces a simple bisecting method to find A^* efficiently.

3.3.2 Computing Exact Values of γ_r, P_r, C_t and $\bar{P}$ for the IHI Problem

We deploy facilities such that the initial service areas form a regular hexagonal partition (each with hexagon size A) on a sufficiently large area $\mathcal{S}$ (e.g., with $|\mathcal{S}| > 100A$) centered at $(0, 0)$. Then $\mathcal{S}$ is diced into infinitesimal squares (e.g., with size $< 0.001A$), each representing a customer neighborhood. To eliminate the influence from the boundary of $\mathcal{S}$, only those squares sufficiently far away from the boundary are considered. For any given values of q and θ, we conduct the following computations to obtain exact values of γ_r, P_r, C_t and $\bar{P}$.

Without losing generality, we focus on facility j which is located at $x_j = (0, 0)$ and determine the service area partition $\{\mathcal{A}_{j,r}, \forall r\}$ as shown in Figure 1. For any customer neighborhood at $x \in \mathcal{A}_{j,r}$, the travel distance to facility j is $\|x\|$ and the corresponding service probability is $(1-q)q^r$. For each r, the exact values of γ_r, P_r are calculated by averaging $\|x\|/A^{1/2}$ and $(1-q)q^r$ respectively across all the corresponding infinitesimal squares in $\mathcal{A}_{j,r}$.

The expected total transportation cost to facility j, $C_{t,j}$, is the summation of $\|x\|(1-q)q^r$ across all customer neighborhoods that satisfy $\|x\| \leq D$. Due to symmetry, the value of $C_{t,j}$ is identical for all j, and hence the transportation cost per unit area is $C_t = \frac{C_{t,j}}{A}$. For every customer neighborhood at $x \in \mathcal{A}_{j,0}$, also count $N_D(x)$, the number of facilities that are within distance D. Penalty probability $\bar{P}(x) = q^{N_D(x)}$, and $\bar{P}$ is computed as the average value of $\bar{P}(x)$ across x.

[g] See Section 3.3.2 for details.

3.3.3 Penalty & Service Probabilities under Correlated Disruptions

In the previous section, Equations (3.7) and (3.9) hold only when the facilities fail independently. Using these formulas will cause significant errors if facility disruptions are actually correlated, as indicated in the following proposition.

Proposition 3. *For any facility location design, the existence of positive (or negative) facility failure correlation increases (or decreases) the expected transportation and penalty cost per unit demand.*

Proof. Any customer at $x \in \mathcal{S}$ may travel a distance $\delta \in [0, D(x))$ to receive service. Define $c(\delta), \forall \delta \leq D(x)$ to be the cost for one unit of demand at x; i.e.,

$$c(\delta) = \begin{cases} \alpha_t \delta, & \delta < D(x); \\ \alpha_p \delta, & \delta = D(x). \end{cases}$$

Obviously, $c(\delta)$ is an increasing function of δ since $\alpha_p \geq \alpha_t$. Under correlated facility failure, let $F(\delta)$ denote the probability for the customer to travel farther than distance δ. The expected cost for one unit of demand at x is

$$\begin{aligned} E[c(\delta)] = \int_{\delta=0}^{D(x)} c(\delta) d[1 - F(\delta)] + c(D(x))F(D(x)) \\ = \int_{\delta=0}^{D(x)} F(\delta) dc(\delta) + c(D(x))F(D(x)). \end{aligned} \quad (3.16)$$

If facility failure is independent, the probability for the customer to travel farther than distance δ is denoted by $F_I(\delta)$. The expected cost becomes

$$\begin{aligned} E_I[c(\delta)] = \int_{\delta=0}^{D(x)} c(\delta) d[1 - F_I(\delta)] + c(D(x))F_I(D(x)) \\ = \int_{\delta=0}^{D(x)} F_I(\delta) dc(\delta) + c(D(x))F_I(D(x)). \end{aligned} \quad (3.17)$$

Note that $F(\delta)$ and $F_I(\delta)$ are the probabilities for all facility within distance δ from x to fail. By definition, for any δ, $F(\delta) \geq F_I(\delta)$ under positive correlations, or $F(\delta) \leq F_I(\delta)$ otherwise. Hence, comparison between (3.16) and (3.17) clearly shows that $E[c(\delta)] \geq E_I[c(\delta)]$ when the correlation is positive; the contrary is also true. This completes the proof.

□

Hence, accurate estimation of the total cost mandates that the penalty probability $\bar{P}$

and the service probability P_r accommodate failure correlations. In the rest of this subsection, we provide a general formulation framework for $\bar{P}(x)$ and $P_r(x)$ based on conditional probabilities.

In the literature, conditional probabilities have been used to model general correlations of symmetric binary events.[h] Based on the Pascal's triangle, probabilities of symmetric disruptions can be represented as the product of facility failure probabilities conditional on the number of neighboring disruptions.

On the infinite homogeneous plane, we let $\{q_l, l = 0, 1, 2...\}$ denote the conditional failure probability of a facility given that (i) this facility is the $(l+1)^{th}$ closest facility to a certain customer, and (ii) all l closer facilities to this customer have failed (regardless of all other facilities on the plane). As such, q_0 represents the unconditional individual failure probability.[i] Generally, if q_l increases with l, the failure correlation is positive. For example, $q_l = 1, \forall l \geq 1$ yields the case of perfect correlation (i.e., facilities either all survive or all fail).[j] On the other hand, facility failure is negatively correlated if q_l decreases with l.

When $\{q_l, \forall l\}$ is known (e.g., from historical data), from the perspective of a customer, the probability for all m nearest facilities to fail simultaneously equals $\prod_{l=0}^{m-1} q_l$. For general $\theta \in \mathbb{R}_+$, $\bar{P}$ can be approximated by interpolating the probabilities for $\lfloor\theta\rfloor$ and $\lfloor\theta\rfloor + 1$ simultaneous failures, as follows,

$$\bar{P} \approx \begin{cases} (1-\bar{\theta}) \prod_{l=0}^{\lfloor\theta\rfloor-1} q_l + \bar{\theta} \prod_{l=0}^{\lfloor\theta\rfloor} q_l, & \theta > 1; \\ 1 - (1-q_0)\theta, & \text{otherwise}, \end{cases} \tag{3.18}$$

It is easy to observe that (3.7) bounds (3.18) from below/above under positive/negative correlations, indicating over-/under-estimation of penalty probability when correlation is ignored. Probability P_r equals the probability that all r nearest facilities to a customer fail while the $(r+1)^{\text{th}}$ facility survives; i.e., (3.9) shall be replaced by

$$P_r \approx (1 - q_r) \prod_{l=0}^{r-1} q_l. \tag{3.19}$$

In certain cases, the conditional probabilities may be dependent of A especially when the correlation magnitude is sensitive to the distance among facilities. The modeling framework described above remains applicable by simply specifying the appropriate $\{q_l(A), \forall l\}$.

[h] Interested readers are referred to Bakkaloglu et al. (2002) and Tang and Iyer (1992) for reviews on this topic.

[i] If failure correlation is ignored, $\bar{P}$ and $P_r, \forall r$, shall be computed from (3.7) and (3.9) with probability $q = q_0$.

[j] Such an extreme case is usually induced by a shared failure source that causes simultaneous disruptions of all facilities. Examples may include outage in a power grid due to failure of the power plant.

3.3.4 CA Model for Heterogeneous Space

We assume that in a finite heterogeneous space $\mathcal{S} \subset \mathbb{R}^2$, parameters $f(x)$, $\lambda(x)$, $D(x)$ vary slowly over $x \in \mathcal{S}$. Instead of looking for $\mathbf{x}$ directly, we propose to look for a continuous function, $A(x) \in \mathbb{R}_+, x \in \mathcal{S}$, that approximates the initial service area size of a facility near x; i.e., $A(x) \approx |\mathcal{A}_{j,0}|$ if $x \in \mathcal{A}_{j,0}$. We assume that $\mathcal{S}$ is far larger than $A(x)$; i.e., $|\mathcal{S}| \gg A(x), \forall x \in \mathcal{S}$. When all parameters are *approximately constant* over a region comparable to the size of several initial service areas, $\bar{P}(x), P_r(x), \theta(x)$ and $A(x)$ should also be approximately constant on that scale.[k]

We apply the cost formulation (3.15) to the neighborhood of x (i.e., imagining that this neighborhood is part of an infinite and homogeneous plane), while using the values of $f(x)$, $\lambda(x)$, $D(x)$ as the parameter input. Incorporating (3.15) into (3.3) yields the following

$$\min_{A(x)} \int_{x \in \mathcal{S}} C(x, A(x)) dx, \tag{3.20}$$

where the total cost per unit area near x is

$$C(x, A(x)) := f(x)/A(x) + \alpha_p \lambda(x) D(x) \bar{P}(x) + \alpha_t \lambda(x) A^{1/2}(x) U(\theta(x), \mathbf{P}), \forall x \in \mathcal{S}. \tag{3.21}$$

Since the inverse of $A(x)$ represents the facility density at x, the number of facilities is

$$N \approx \int_{x \in \mathcal{S}} [A(x)]^{-1} dx. \tag{3.22}$$

For any x, (3.21) has only one scalar decision variable $A(x)$. We shall note that the values of $\bar{P}(x)$ and $U(\theta(x), \mathbf{P})$ depend on $\theta(x)$ and hence on $A(x)$, although $\mathbf{P}$ is independent of $A(x)$ as suggested by (3.9) and (3.19). Intuitively, the second and third terms in (3.21) should be increasing with $A(x)$, while the first term is decreasing with $A(x)$. Hence, the function $C(x, A(x))$ is likely to have a "V" shape with regard to $A(x)$. The optimal solution $A^*(x)$ can be obtained from a simple bisecting search.

The estimated optimal cost per unit area $C(x, A^*(x))$ and the optimal facility density function $[A^*(x)]^{-1}$ can be integrated across $\mathcal{S}$ to yield the total system cost C^* and the optimal number of facility N^*, respectively. Function $A^*(x), x \in \mathcal{S}$ and N^* can be used in the disk model (Ouyang and Daganzo, 2006) to design the optimal discrete facility locations. The disk model exerts repulsive forces to N^* disks that each represents a facility and its initial service area, and iteratively adjusts positions and sizes of these disks to achieve optimal

[k] Interested readers are referred to Sections 4.2, 4.4 and Section B in Cui et al. (2009) for discussions on the applicability and accuracy of the continuum approximation method.

layout. Interested readers are referred to Ouyang and Daganzo (2006) and Ouyang (2007) for more implementation details. These references have also shown that the total cost of the discrete design obtained from the disk model is very close to that estimated by (3.20).

3.4 Alternative Correlation Structures

Facility failure correlations can be modeled in a variety of ways. This section discusses two special cases. Section 3.4.1 simplifies the formulation with beta-binomial distributions when the correlation is always positive. Section 3.4.2 shows how to decompose $\bar{P}(x)$ and $P_r(x)$ into scenario-based probabilities in case that the facility failure mechanisms are known.

3.4.1 Positively Correlated Beta-Binomial Facility Failure

The modeling approach in Section 3.2 requires a whole set of conditional failure probabilities $\{q_l, \forall l\}$ to be specified (most likely from historical data). This may be tedious in certain practical situations. As an alternative, the beta-binomial distribution has been used in various fields (such as computer science (Bakkaloglu et al., 2002; Goyal and Nicola, 1990) and biometrics (Griffiths, 1973)) to model positive failure correlations. The beta-binomial distribution, which we denote by $B_{n,a,b}$ with $a, b > 0$, only has three parameters. It is defined as the distribution for the number of failures in n symmetric success/failure experiments, while each experiment has a random failure probability p whose probability density function is

$$\frac{p^{a-1}(1-p)^{b-1}}{\int_0^1 p^{a-1}(1-p)^{b-1}dp}, p \in [0,1].$$

Accordingly, the probability that m out of n experiments fail is

$$B_{n,a,b}(m) := \binom{n}{m} \frac{[(a+m-1)(a+m-2)\cdots a][(b+n-m-1)(b+n-m-2)\cdots b]}{(a+b+n-1)(a+b+n-2)\cdots(a+b)}. \tag{3.23}$$

Equation (3.23) can be equivalently structured in terms of the general conditional probabilities

$$q_l = \frac{B_{l+1,a,b}}{B_{l,a,b}} = \frac{a+l}{a+b+l}, \forall l. \tag{3.24}$$

$B_{n,a,b}$ has mean $n\frac{a}{a+b}$ and variance $n\frac{ab}{(a+b)^2}\frac{1+n/(a+b)}{1+1/(a+b)}$. Compared with the regular binomial distribution with n experiments and independent failure probability $\frac{a}{a+b}$, $B_{n,a,b}$ has the same mean but a larger variance; the amplification factor $\frac{1+n/(a+b)}{1+1/(a+b)}$ captures positive corre-

lation among facility failure.[1] The positive correlation can be also seen from (3.24) where q_l obviously increases over l.

For RUFL, we assume that the probability for all n nearest facilities to a customer at $x \in \mathcal{S}$ to fail is given by $B_{n,a(x),b(x)}(n)$ with varying parameters $a(x)$, $b(x)$. Through interpolation (similar to (3.18)), $\bar{P}(x)$ can be represented as follows,

$$\bar{P}(x) \approx \begin{cases} (1-\bar{\theta})B_{\lfloor\theta(x)\rfloor,a(x),b(x)}(\lfloor\theta(x)\rfloor) \\ \qquad +\bar{\theta}B_{\lfloor\theta(x)\rfloor+1,a(x),b(x)}(\lfloor\theta(x)\rfloor+1), & \theta(x)>1; \\ 1-[1-B_{1,a(x),b(x)}(1)]\theta(x), & \text{otherwise,} \end{cases} \tag{3.25}$$

We also assume that the probability for all n nearest facility to fail but the $(n+1)^{th}$ facility to survive is given by $B_{n+1,a(x),b(x)}(n)$, and the service probability $P_r(x)$ can be approximated by

$$P_r(x) \approx B_{r+1,a(x),b(x)}(r). \tag{3.26}$$

Again, if the facilities fail independently, probabilities $\bar{P}(x)$ and $P_r(x)$ could be obtained from (3.7) and (3.9) respectively, with probability $q(x) = B_{1,a(x),b(x)}(1) = \frac{a(x)}{a(x)+b(x)}$.

3.4.2 Correlation Induced from Shared Hazard Exposure

Sometimes the sources and causal mechanisms of facility disruptions are well understood. In such cases, the disruption probabilities can be conditioned on a set of mutually exclusive hazard occurrence states, $\mathcal{H}$. Each state $h \in \mathcal{H}$ corresponds to a possible scenario of hazard occurrence (e.g., earthquake, hurricane). Suppose state h happens with a probability Q_h ($\sum_{h\in\mathcal{H}} Q_h = 1$ if "no-disaster" is considered one of the states). Conditional on each state h, each facility near x fails independently with probability $\chi_h(x)$. It should be noted that although the facilities fail independently within each hazard occurrence state, the overall facility disruptions (due to all hazards) can be correlated.

In each state h, the penalty probability for a customer at x can be approximated by (3.7). Based on conditional expectation, the overall penalty probability $\bar{P}(x)$ across all possible hazard occurrence state is

$$\bar{P}(x) \approx \begin{cases} \sum_h Q_h[\chi_h(x)]^{\theta(x)}, & \theta(x)>1 \\ \sum_h Q_h\{1-[1-\chi_h(x)]\theta(x)\}, & \text{otherwise.} \end{cases} \tag{3.27}$$

[1] A larger value of $\frac{1}{a+b}$ corresponds to a greater variance and hence more significant correlation (Bakkaloglu et al., 2002). For example, when $\frac{1}{a+b} \to 0$, facilities fail almost independently; when $\frac{1}{a+b} \to \infty$, facility failure is almost perfectly correlated.

Similarly, given h, the service probability at rank r for a customer at x can be approximated from (3.9). The expected value across all states yields $P_r(x)$ as follows

$$P_r(x) \approx \sum_h Q_h[1-\chi_h(x)]\chi_h^r(x). \tag{3.28}$$

Note that $\bar{P}(x)$ and $P_r(x)$ can be expressed equivalently in the form of (3.18) and (3.19), respectively, by setting

$$q_l = \frac{\sum_h Q_h \chi_h(x)^{l+1}}{\sum_h Q_h \chi_h(x)^l}, \forall l.$$

Now we briefly discuss how the penalty and service probabilities will be erroneous if correlations are ignored. Note that the single facility failure probability $q(x) = \sum_h Q_h\chi_h(x)$. $\bar{P}(x)$ and $P_r(x)$ could be calculated from (3.7) and (3.9) as follows,

$$\bar{P}(x) \approx \begin{cases} \left[\sum_h Q_h\chi_h(x)\right]^{\theta(x)}, & \theta(x) > 1 \\ 1-[1-\sum_h Q_h\chi_h(x)]\theta(x), & \text{otherwise} \end{cases},$$

$$P_r(x) \approx \left[1-\sum_h Q_h\chi_h(x)\right]\left[\sum_h Q_h\chi_h(x)\right]^r.$$

Note that much of the difference in the corresponding probability formulas (with or without correlations) comes from the fact that

$$\mu_r(x) := \sum_h Q_h\chi_h^r(x) - \left[\sum_h Q_h\chi_h(x)\right]^r \geq 0, \forall r > 2, \tag{3.29}$$

due to the Jensen's Inequality. Note that $\mu_r(x)$ becomes even larger as r increases.

3.5 Numerical Examples

This section presents four numerical examples to illustrate how the CA model can be applied to problems with correlated facility disruptions. Each example uses an aforementioned failure correlation structure. The space $\mathcal{S}$ is a $[0,1] \times [0,1]$ unit square. Customer demand is distributed with density function $\lambda(x) = \bar{\lambda}[1+\tau_\lambda \cos(\omega\|x\|)]$, and the facility opening cost at x is $f(x) = \bar{f}[1+\tau_f \cos(\omega\|x\|)]$, where $\tau_\lambda \in [-1,1]$ and $\tau_f \in [-1,1]$ control the heterogeneity of $\lambda(x)$ and $f(x)$ over $\mathcal{S}$, respectively. The scalar ω is selected to normalize the average customer density and the average facility cost (e.g., $\int_{\mathcal{S}} \lambda(x)dx = \bar{\lambda}$ and $\int_{\mathcal{S}} f(x)dx = \bar{f}$). The travel cost factor $\alpha_t = 1$.

The estimated optimal total cost C^* and the estimated optimal facility number N^* are computed from (3.20) and (3.22) respectively. For comparison, we let $A_I(x)$, C_I and N_I respectively denote the optimal $A(x)$, total cost, and facility number when correlation is erroneously ignored. These three values may be relevant to strategic resource allocation and budget planning. Let C_{IC} denote the actual cost under correlation while solution $A_I(x)$ is implemented. The percentage difference $\varepsilon_I = \frac{C_I - C^*}{C^*}$ indicates the error in estimated system cost caused by ignoring correlations, while $\varepsilon_{IC} = \frac{C_{IC} - C^*}{C^*}$ indicates the actual cost difference after implementing the "wrong" design.

3.5.1 Correlation Specified by Conditional Probabilities

Following the framework presented in Section 3.3.3, we set the conditional probabilities to be

$$\begin{aligned} &q_1(x) = q_0(x) + \Delta q(x), \\ &q_l(x) = \min\left\{ q_{l-1}(x) + \frac{q_{l-1}(x) - q_{l-2}(x)}{2}, \frac{q_{l-1}(x)+1}{2} \right\}, \forall l = 2, 3, ..., \end{aligned} \tag{3.30}$$

Here, positive/negative $\Delta q(x)$ yields positive/negative correlations; e.g., perfect correlation can be specified by setting $\Delta q(x) = 1 - q_0(x)$. For demonstration purposes, we simply assume $q_0(x) = q_0$, $\Delta q(x) = \Delta q, \forall x$.

Substituting Equation (3.30) into (3.18) and (3.19) yields the correct penalty and service probabilities, while the erroneous counterparts can be computed from (3.7) and (3.9). Define $\theta^* = \pi D^2 / A^*$. Table 3.1 illustrates the results for a range of problem instances with $\bar{f} = 1$, $\bar{\lambda} = 500$, $\omega = 11.73$, $\tau_\lambda, \tau_f \in \{0, 1\}$, $q_0 \in \{0.05, 0.2\}$, $\Delta q \in \{-q_0/2, (1-q_0)/2, 1-q_0\}$, $\alpha_p \in \{1, 10\}$, and $D \in \{0.1, 0.2\}$.

It can be observed that N^*, N_I, C^*, C_I and C_{IC} all increase with q_0 in almost all cases, indicating that facilities should be deployed closer to each other (as back-ups) under higher failure probabilities, and as a result the total system cost increases. The same trend is observed as α_p increases; this is intuitive because higher α_p implies higher penalty cost, which would motivate a denser facility deployment. As D increases (i.e., reducing the likelihood for customers to incur penalty), the optimal numbers of facilities, N^* and N_I, both decrease; however, the values of C^*, C_I and C_{IC} may still increase because a larger D also implies a proportionally larger penalty value.

The optimal total cost C^* is obviously influenced by the correlation, sometimes dramatically (when q_0 and α_p are large); positive correlation generally leads to higher total cost. On the contrary, the optimal number of facilities N^* decreases under positive correlation

Table 3.1: CA cost estimation when correlation is specified by conditional probabilities.

#	τ_λ	τ_f	q_0	Δq	α_p	D	θ^*	N^*	N_I	C^*	C_I	C_{IC}	ε_I	ε_{IC}
1	0	0	0.05	-0.025	1	0.2	2.7	21	21	64	64	64	0 %	0 %
2	0	0	0.05	-0.025	10	0.2	3	24	22	64	64	64	0 %	0 %
3	0	0	0.05	-0.025	10	0.1	1.4	44	44	88	80	88	-9 %	0 %
4	0	0	0.05	0.475	1	0.2	2.6	21	21	65	64	65	-1 %	0 %
5	0	0	0.05	0.475	10	0.2	3.2	26	22	83	64	84	-22 %	1 %
6	0	0	0.05	0.475	10	0.1	1	32	44	89	80	92	-10 %	4 %
7	0	0	0.05	0.95	1	0.2	2.5	20	21	65	64	65	-2 %	0 %
8	0	0	0.05	0.95	10	0.2	2.5	20	22	110	64	110	-42 %	0 %
9	0	0	0.05	0.95	10	0.1	1	32	44	89	80	96	-10 %	8 %
10	0	0	0.2	-0.1	1	0.2	3	24	21	70	70	70	0 %	0 %
11	0	0	0.2	-0.1	10	0.2	3.2	25	31	71	74	72	4 %	1 %
12	0	0	0.2	-0.1	10	0.1	2	64	67	101	110	103	9 %	2 %
13	0	0	0.2	0.4	1	0.2	2.6	21	21	72	70	72	-3 %	0 %
14	0	0	0.2	0.4	10	0.2	4.2	33	31	153	74	154	-52 %	1 %
15	0	0	0.2	0.4	10	0.1	2	64	67	146	110	148	-25 %	1 %
16	0	0	0.2	0.8	1	0.2	2.2	18	21	74	70	74	-5 %	1 %
17	0	0	0.2	0.8	10	0.2	2.2	18	31	254	74	258	-71 %	2 %
18	0	0	0.2	0.8	10	0.1	1	32	67	159	110	185	-31 %	17 %
19	1	0	0.2	-0.1	1	0.2	2.5	20	20	64	64	64	0 %	0 %
20	1	0	0.2	-0.1	10	0.2	3.2	25	28	67	69	68	3 %	1 %
21	1	0	0.2	-0.1	10	0.1	1.6	51	56	92	98	95	6 %	3 %
22	1	0	0.2	0.4	1	0.2	2.3	18	20	66	64	67	-4 %	0 %
23	1	0	0.2	0.4	10	0.2	3.9	31	28	147	69	147	-53 %	1 %
24	1	0	0.2	0.4	10	0.1	1.5	47	56	135	98	137	-28 %	1 %
25	1	0	0.2	0.8	1	0.2	2	16	20	69	64	70	-8 %	1 %
26	1	0	0.2	0.8	10	0.2	2.2	17	28	250	69	254	-72 %	2 %
27	1	0	0.2	0.8	10	0.1	0.9	27	56	156	98	175	-37 %	13 %
28	1	1	0.2	-0.1	1	0.2	3	24	21	70	70	70	0 %	0 %
29	1	1	0.2	-0.1	10	0.2	3.2	25	31	71	73	72	4 %	1 %
30	1	1	0.2	-0.1	10	0.1	2	64	67	101	110	102	9 %	2 %
31	1	1	0.2	0.4	1	0.2	2.6	21	21	72	70	72	-3 %	0 %
32	1	1	0.2	0.4	10	0.2	4.2	33	31	153	73	154	-52 %	1 %
33	1	1	0.2	0.4	10	0.1	2	64	67	146	110	148	-25 %	1 %
34	1	1	0.2	0.8	1	0.2	2.2	18	21	74	70	74	-5 %	1 %
35	1	1	0.2	0.8	10	0.2	2.2	18	31	254	73	258	-71 %	2 %
36	1	1	0.2	0.8	10	0.1	1	32	67	158	110	185	-31 %	17 %

in most of the cases, probably because positively correlated failures weaken the benefit of having more facilities as backups.

The error $|\varepsilon_I|$ always increases with $|\Delta q|$, which means that assuming independent disruptions yields a poor cost estimation when correlations are actually present. The error is large in cases of high failure probability q_0 and large penalty factor α_p. As expected, all ε_I values are negative for positive Δq (leading to underestimation of disruption risks) and non-negative for negative Δq. This is consistent with the discussions in Section 3. On the other hand, the actual cost error ε_{IC} is always non-negative. This is not surprising because $A_I(x)$ is suboptimal to the cost-minimization problem. For most of cases, $|\varepsilon_{IC}|$ is not large. This is probably because the objective function is quite flat near the optimal solution (similar to many other facility location problems). Nevertheless, $|\varepsilon_{IC}|$ is large for large α_p and $\theta^* \approx 1$, as the solutions under these scenarios impose a large penalty risk to the customers.

When $\lambda(x)$ is heterogeneous (i.e., $\tau_\lambda = 1$), N^* and C^* are lower than those in the

corresponding homogeneous cases (i.e., $\tau_\lambda = 0$). This suggests that uneven distribution of customers generally reduces the optimal total cost. In addition, heterogeneous $\lambda(x)$ seems to slightly inflate $|\epsilon_I|$ under positive correlations and reduce it under negative correlations. However, when the facility opening cost $f(x)$ varies in proportion to $\lambda(x)$ (i.e., $\tau_f = 1$), which may happen due to higher land prices in areas with high population density, the results are almost the same as those with homogeneous $\lambda(x)$ and $f(x)$ (i.e., $\tau_\lambda = \tau_f = 0$).

3.5.2 Correlations Specified by the Beta-Binomial Distribution

We assume that all parameters remain the same as those in the previous example, excepted that the correlations are expressed via beta-binomial distribution with parameters $a(x) = a$, $b(x) = b, \forall x$. Equations (3.25) and (3.26) are used to estimate the penalty and service probabilities $\bar{P}(x)$ and $P_r(x)$.

Table 3.2 shows the results for a range of problem instances, where $\bar{f} = 1$, $\bar{\lambda} = 500$, $\omega = 11.73$, $\tau_\lambda, \tau_f \in \{0, 1\}$, $a \in \{0.1, 0.01\}$, $b \in \{19a, 4a\}$, $\alpha_p \in \{1, 10\}$, and $D \in \{0.1, 0.2\}$. Since the beta-binomial formulation is simply a special case of the general conditional probability formulation, the results are consistent with those in Section 3.5.1. Facility number N^* and optimal cost C^* generally increase over the failure probability $\frac{a}{a+b}$, the correlation $\frac{1}{a+b}$, and the penalty factor α_p. The estimation error $|\epsilon_I|$ is large, especially when the penalty cost and the correlation are high, though $|\epsilon_{IC}|$ is only large for a few cases. Heterogeneities in the system again help reduce the optimal number of facilities and the total cost.

3.5.3 Flooding Hazard

Now we suppose that facility failure may be caused by a potential flooding hazard, and the flood, whenever happening, always immerses the whole $\mathcal{S}$.[m] Following the framework introduced in Section 3.4.2, there are $|\mathcal{H}| = 2$ exclusive hazard occurrence states; assume that state $h = 1$ represents no-disaster, which occurs with a high probability $Q_1 = 0.9$, while state $h = 2$ represents flooding disaster, which occurs with a low probability $Q_2 = 0.1$.[n] For $h = 1, 2$, the associated facility failure probability $\chi_h(x) = \chi_h$ for all $x \in \mathcal{S}$, where $\chi_1 \ll 1$ and $\chi_2 > 0$. Penalty and service probabilities $P_r(x)$ and $\bar{P}(x)$ are computed from (3.27) and (3.28) respectively.

Now that we have two hazard occurrence states, we use ε_{IC1} and ε_{IC2} to replace ε_{IC}, representing the actual total cost error under states 1 and 2, respectively. Table 3.3 shows

[m] For problems where only some subareas are subject to such hazards, we can partition $\mathcal{S}$ accordingly and solve a subproblem for each subarea.

[n] This is for illustration only; in the real world Q_2 should be much smaller.

Table 3.2: CA cost estimations when correlation is specified by the beta-binomial distribution.

#	τ_λ	τ_f	$\frac{a}{a+b}$	$\frac{1}{a+b}$	α_p	D	θ^*	N^*	N_I	C^*	C_I	C_{IC}	ε_I	ε_{IC}
1	0	0	0.05	0.5	1	0.2	2.4	19	21	68	64	68	-6 %	0 %
2	0	0	0.05	0.5	10	0.2	3.6	28	22	77	64	78	-16 %	2 %
3	0	0	0.05	0.5	10	0.1	1	32	43	89	80	91	-10 %	3 %
4	0	0	0.05	5	1	0.2	2.5	20	21	66	64	66	-10 %	2 %
5	0	0	0.05	5	10	0.2	3	24	22	101	64	101	-60 %	2 %
6	0	0	0.05	5	10	0.1	1	32	43	89	80	95	-31 %	3 %
7	0	0	0.2	2	1	0.2	1.9	15	21	78	70	80	-11 %	0 %
8	0	0	0.2	2	10	0.2	5	40	30	183	74	186	-60 %	0 %
9	0	0	0.2	2	10	0.1	1.1	36	66	159	110	163	-31 %	7 %
10	0	0	0.2	20	1	0.2	2.2	17	21	74	70	75	-6 %	1 %
11	0	0	0.2	20	10	0.2	3	24	30	246	74	247	-70 %	1 %
12	0	0	0.2	20	10	0.1	1	32	66	159	110	182	-31 %	15 %
13	1	0	0.2	2	1	0.2	1.6	12	19	74	64	77	-14 %	4 %
14	1	0	0.2	2	10	0.2	4.6	36	28	175	69	179	-61 %	2 %
15	1	0	0.2	2	10	0.1	1.3	42	55	147	98	151	-34 %	3 %
16	1	0	0.2	20	1	0.2	1.9	14	19	70	64	71	-9 %	2 %
17	1	0	0.2	20	10	0.2	2.7	21	28	241	69	242	-71 %	1 %
18	1	0	0.2	20	10	0.1	0.9	27	55	156	98	172	-37 %	10 %
19	1	1	0.2	2	1	0.2	1.9	15	21	78	70	80	-10 %	2 %
20	1	1	0.2	2	10	0.2	5	39	30	183	73	186	-60 %	2 %
21	1	1	0.2	2	10	0.1	1.1	35	66	158	110	163	-31 %	3 %
22	1	1	0.2	20	1	0.2	2.2	17	21	74	70	75	-6 %	1 %
23	1	1	0.2	20	10	0.2	3	23	30	245	73	247	-70 %	1 %
24	1	1	0.2	20	10	0.1	1	31	66	158	110	182	-31 %	15 %

the results for a range of instances where $\bar{f} = 1$, $\bar{\lambda} = 500$, $\omega = 11.73$, $\tau_\lambda, \tau_f \in \{0, 1\}$, $[\chi_1, \chi_2] \in \{[0, 0.5], [0, 1]\}$, $\alpha_p \in \{1, 10\}$, and $D \in \{0.1, 0.2\}$. Recall that $\mu_2 = \sum_h Q_h \chi_h^2(x) - [\sum_h Q_h \chi_h(x)]^2$ indicates the magnitude of positive correlations. We can observe that the impacts of failure probability, correlation, penalty and parameter heterogeneities on the optimal number of facilities and the total cost are similar to those seen in the previous numerical experiments.

3.5.4 Earthquake Hazard

This section considers a heterogeneous case where earthquake hazards impose site-dependent failure probability over $\mathcal{S}$. The setting is the same as that in Section 3.5.3 except that hazard occurrence state $h = 2$ is induced by an earthquake source centered at $(0, 0)$, and ω is set to be 2.038 to ensure that $\lambda(x)$ is monotone (either always decreases or always increases) as we move away from the earthquake center. When an earthquake occurs, a facility at $x \in \mathcal{S}$ fails with a probability $q(x) = \exp(-\beta \|x\|)$, where β is a scalar.

Table 3.4 shows the results for a range of instances. Define $\bar{q} = \int\limits_{x \in \mathcal{S}} \sum_h Q_h \chi_h(x) dx$ and $\bar{\mu}_2 = \int\limits_{x \in \mathcal{S}} \mu_2(x) dx$. The average values of $\sum_h Q_h \chi_h(x)$ and $\mu_2(x)$ are set to be comparable to those in the flooding examples. As expected, we observe consistent effects of failure probability, correlation, penalty and heterogeneities. Particularly, the spatial distribution patterns

Table 3.3: CA cost estimations for flooding hazard.

#	τ_λ	τ_f	$\sum_h Q_h \chi_h$	μ_2	α_p	D	θ^*	N^*	N_I	C^*	C_I	C_{IC}	ε_I	ε_{IC}	ε_{IC1}	ε_{IC2}
1	0	0	0.05	0.02	1	0.2	2.6	21	21	64	64	64	-1 %	0 %	0 %	0 %
2	0	0	0.05	0.02	10	0.2	4.1	33	22	73	64	78	-12 %	6 %	-5 %	54 %
3	0	0	0.05	0.02	10	0.1	1	32	44	89	80	91	-10 %	2 %	11 %	-14 %
4	0	0	0.1	0.09	1	0.2	2.4	19	21	68	66	68	-3 %	0 %	0 %	2 %
5	0	0	0.1	0.09	10	0.2	2.4	19	25	158	67	159	-58 %	1 %	1 %	1 %
6	0	0	0.1	0.09	10	0.1	1	32	51	112	90	125	-19 %	12 %	19 %	4 %
7	1	0	0.05	0.02	1	0.2	2.3	19	19	59	58	59	-1 %	0 %	0 %	0 %
8	1	0	0.05	0.02	10	0.2	3.6	29	22	67	60	69	-9 %	4 %	-3 %	36 %
9	1	0	0.05	0.02	10	0.1	1.1	36	38	83	74	84	-12 %	1 %	2 %	-3 %
10	1	0	0.1	0.09	1	0.2	2.1	17	19	63	60	63	-4 %	0 %	0 %	1 %
11	1	0	0.1	0.09	10	0.2	2.3	19	24	154	63	155	-59 %	1 %	1 %	0 %
12	1	0	0.1	0.09	10	0.1	0.9	28	44	109	82	119	-25 %	9 %	14 %	3 %
13	1	1	0.05	0.02	1	0.2	2.6	21	21	65	64	65	-1 %	0 %	0 %	0 %
14	1	1	0.05	0.02	10	0.2	4.1	33	22	73	64	78	-12 %	6 %	-5 %	54 %
15	1	1	0.05	0.02	10	0.1	1	32	44	89	80	91	-10 %	2 %	11 %	-14 %
16	1	1	0.1	0.09	1	0.2	2.4	19	21	68	66	68	-3 %	0 %	0 %	2 %
17	1	1	0.1	0.09	10	0.2	2.4	19	25	158	67	159	-58 %	1 %	1 %	1 %
18	1	1	0.1	0.09	10	0.1	1	32	51	112	90	125	-19 %	12 %	19 %	4 %

of customer demand $\lambda(x)$ and facility failure probability $q(x)$ seem to jointly influence the optimal system design. For example, when customer density increases with the distance from the earthquake center (i.e., $\tau_\lambda = -1$), the optimal cost C^* drops. This desirable situation is probably due to not only the demand heterogeneity but also the concentration of demand in places with lower facility failure risks. On the contrary, when customer density decreases with the distance from the earthquake center (i.e., $\tau_\lambda = 1$), the change of C^* is not always monotone. Although the heterogeneity of $\lambda(x)$ tends to reduce the total cost, the fact that more customers live in places with higher facility failure risk tends to increase the total system cost.

Table 3.4: CA cost estimations for earthquake hazard.

#	τ_λ	β	$\bar{q}$	$\bar{\mu}_2$	α_p	D	θ^*	N^*	N_I	C^*	C_I	C_{IC}	ε_I	ε_{IC}	ε_{IC1}	ε_{IC2}
1	0	1	0.05	0.02	1	0.2	2.6	21	21	64	64	64	-1 %	0 %	0 %	0 %
2	0	1	0.05	0.02	10	0.2	4.1	32	22	75	64	79	-14 %	6 %	-5 %	47 %
3	0	1	0.05	0.02	10	0.1	1	32	43	88	80	90	-9 %	3 %	10 %	-12 %
4	0	0.05	0.1	0.08	1	0.2	2.5	20	21	68	66	68	-3 %	0 %	0 %	1 %
5	0	0.05	0.1	0.08	10	0.2	3.4	27	25	148	67	149	-55 %	0 %	-1 %	1 %
6	0	0.05	0.1	0.08	10	0.1	1	32	51	110	89	122	-19 %	11 %	19 %	2 %
7	1	1	0.05	0.02	1	0.2	2.5	20	20	63	62	63	-1 %	0 %	0 %	0 %
8	1	1	0.05	0.02	10	0.2	4.2	33	22	74	63	80	-16 %	8 %	-7 %	60 %
9	1	1	0.05	0.02	10	0.1	1.1	36	43	91	78	92	-14 %	1 %	6 %	-8 %
10	1	0.05	0.1	0.08	1	0.2	2.3	19	20	66	64	66	-3 %	0 %	0 %	1 %
11	1	0.05	0.1	0.08	10	0.2	3.2	25	25	148	65	149	-56 %	0 %	0 %	0 %
12	1	0.05	0.1	0.08	10	0.1	1	32	49	111	87	120	-21 %	9 %	15 %	2 %
13	-1	1	0.05	0.02	1	0.2	2.4	19	19	61	61	61	0 %	0 %	0 %	0 %
14	-1	1	0.05	0.02	10	0.2	3.5	28	22	69	62	71	-9 %	3 %	-3 %	27 %
15	-1	1	0.05	0.02	10	0.1	1.1	36	41	83	76	84	-8 %	1 %	4 %	-7 %
16	-1	0.05	0.1	0.08	1	0.2	2.3	18	19	65	63	65	-3 %	0 %	0 %	1 %
17	-1	0.05	0.1	0.08	10	0.2	4	32	24	140	65	143	-54 %	2 %	-6 %	7 %
18	-1	0.05	0.1	0.08	10	0.1	1	31	48	109	86	118	-21 %	8 %	15 %	1 %

3.6 List of Symbols

A: Size of the hexagonal initial service area in an IHI
$A_I(x)$: Optimal $A(x)$ when correlation is ignored
$\mathcal{A}_{j,r}$: Subset of customers who are assigned a rank r by facility j
$B_{n,a,b}$: Beta-binomial distribution
$c(\delta)$: Cost for one unit of demand at x to travel δ
C^*: Estimated optimal total cost
C_f: Unit-area facility opening cost
C_I: Optimal total cost when correlation is ignored
C_{IC}: Actual cost under correlation when $A_I(x)$ is implemented
C_p: Unit-area expected penalty cost
C_t: unit-area expected transportation cost
$D(x)$: Penalty distance at x
$\bar{f}$: Parameter to specify average $f(x)$ over space $\mathcal{S}$
$f(x)$: Fixed facility opening cost at x
$F(\delta)$: Probability for a customer to travel farther than δ under dependent failures
$F_I(\delta)$: Probability for a customer to travel farther than δ under independent failures
$\mathcal{H}$: Set of mutually exclusive hazard occurrence states
IHI: *infinite* and *homogeneous* plane and the facilities fail *independently*
N: Total number of facilities
N^*: Estimated optimal facility number
N_I: Optimal facility number when correlation is ignored
$N_D(x)$: Number of facilities that a customer at x can visit within D
$\bar{P}$: Average value of $\bar{P}(x)$
$P(x, x_j|\mathbf{x})$: Probability for this customer at x to be served by facility j
P_r: Probability that $\mathcal{A}_{j,r}$ receives service from facility j
$\bar{P}(x|\mathbf{x})$: Probability for the customer at x not to be served
$q(x)$: Failure probability of a facility at x
q_l: Conditional failure probability of a facility
Q_h: Probability of state $h \in \mathcal{H}$
r: Index of a facility number or a customer service level
$\mathcal{S}$: Two-dimensional space
$\bar{\lambda}$: Parameter to specify average $\lambda(x)$ over space $\mathcal{S}$
$\lambda(x)$: Demand density at $x \in \mathcal{S}$

x_j: Location of the j^th facility,$j = 1, 2, \cdots N$

α_t: Coefficient of transportation cost

α_p: Coefficient of penalty cost

$\chi_h(x)$: Probability that each facility near x fails in state $h \in \mathcal{H}$

δ: Variable for distance

$\Delta q(x)$: Parameter for correlation in conditional probabilities

ε_I: $\frac{C_I - C^*}{C^*}$

ε_{IC}: $\frac{C_{IC} - C^*}{C^*}$

γ_r: Scalar for the average distance from $\mathcal{A}_{j,r}$ to facility j

θ: $S\pi D^2/A$

τ_f: Parameter to specify variation of $f(x)$ across space $\mathcal{S}$

τ_λ: Parameter to specify variation of $\lambda(x)$ across space $\mathcal{S}$

ω: Parameter to normalize the average customer density and facility cost

Chapter 4

Reliable Traffic Surveillance Sensor Design: Homogeneous Failure

The optimization methodologies for supply chain location problems can be extended to traffic sensor location problems. This chapter proposes a reliable sensor deployment model for advanced vehicle ID identification sensors that can synthesize disaggregated vehicle information from multiple locations. This model optimizes traffic surveillance benefit from synthesized sensor pairs (e.g., for travel time estimation) in addition to individual sensor flow coverage (e.g., for traffic volume statistics), while considering probabilistic sensor failures. Customized greedy and Lagrangian relaxation algorithms are proposed to solve this problem, and their performance is discussed. We test our algorithm with a moderate-scale network, which shows that the proposed algorithms solve the problem efficiently. Then we apply it to the Chicago intermodal network and discuss managerial insights on how optimal sensor deployment and surveillance benefits vary with surveillance objective and system parameters (such as sensor failure probabilities).

The proposed model can also be applied to the railroad context. We conduct a case study on railroad wayside detector deployment. According to railroad specifications, we only need to consider individual sensor flow coverage of railcars without sensor failures. Our model (after adaption) is able to solve very large-scale problems with over ten thousand nodes and around half a million railcars. We have implemented the wayside detector location model into a standalone solver, which has been used by a class-I railroad to make sensor deployment decisions.

4.1 Introduction

Sensor technologies (e.g., loop detectors, surveillance cameras, radio frequency identifications/RFID) have been widely used on highway networks. Real-time traffic information is sampled by these sensors to monitor traffic status and to develop control strategies. The effectiveness of a traffic surveillance system depends on not only the accuracy of the sampled information but also the coverage over the transportation network. However, implementing these new technologies usually requires large investment. Accuracy and coverage are often two conflicting objectives due to limited resources: collecting high-quality information usually relies on sophisticated and expensive technologies and thus limited budget would restrict the number of installations; on the other hand, due to the limited effective range of most sensors, complete coverage over a network usually requires dense installations. To balance this trade-off, intensive studies have been conducted to determine efficient and reliable deployment of surveillance systems. Early studies mostly focused on deploying traditional sensors (e.g., inductive loop detectors) that provide aggregated statistics (e.g., vehicle counts) for purposes related to origin-destination (O-D) flow volume estimation. Lam and Lo (1990) proposed a heuristic approach to select locations for traffic flow volume counting sensors in a roadway network. Yang et al. (1991) conducted a robust analysis on the utility of traffic counting point, and Yang and Zhou (1998) proposed a sensor deployment framework to maximize such utilities. This framework has been extended to accommodate turning traffic information (Bianco et al., 2001, 2006), existing installations and O-D information content (Ehlert et al., 2006), the screen line problem (Yang et al., 2006), time-varying network flows (Fei et al., 2007; Fei and Mahmassani, 2008), railcar inspection under potential sensor failures (Ouyang et al., 2009), and unobserved link flow estimation (Hu et al., 2009). The emergence of advanced traffic sensor technologies (e.g., automatic vehicle identification tag readers, radio frequency identification sensors) that are able to track vehicle identifications has further enabled flow volume estimation for individual O-D paths. For example, Gentili and Mirchandani (2005) investigated guidelines for locating advanced traffic sensors that are able to read both a vehicle's identification and its route information. Castillo et al. (2008) proposed a location model to determine the optimal locations of vehicle plate scanning sensors for path flow reconstruction. Recent studies have investigated the potential use of sensor data for network O-D travel time estimation. The location of traditional sensors in a single freeway corridor (Bartin et al., 2007; Ban et al., 2009), and deployment of vehicle identification technologies on a highway network (Sherali et al., 2006; Mirchandani et al., 2009) have been considered in support of network travel time estimation.

Despite numerous studies on O-D flow coverage, research on the usage of sensors for

network O-D travel time estimation has been relatively scarce. To the best of our knowledge, only Ban et al. (2009) developed sensor deployment algorithm for travel time estimation in a single freeway corridor—little research has addressed the problem in general networks. Accurate travel time estimation provides important information for decision support in both private sectors (e.g., tacking fleets for trucking companies, traveler information provision) and public agencies (e.g., congestion mitigation, accident management). For a transportation network, we may want to know as much as possible the real-time travel time between all possible O-D pairs. However, traditional surveillance technologies (e.g., loop detectors) would encounter significant challenges due to their inability to accurately capture O-D flows (Kerner and Rehborn, 1996; Li et al., 2010). New sensor technologies, on the other hand, are able to identify vehicle IDs and therefore hold the promise to overcome these challenges by synthesizing vehicle ID information from different sensors. For example, the consecutive time stamps of a vehicle at two sensor locations would provide an accurate estimate of travel time.

Like many other IT technologies, most existing sensors are subject to performance disruptions due to system errors, adverse weather conditions, or intentional sabotage (Rajagopal and Varaiya, 2007; Carbunar et al., 2005). Intuitively, such failures may substantially impair the surveillance effectiveness. Potential disruptions need to be addressed in a reliable design so that the sensor system not only has a good performance in the normal scenario but also is resilient against possible loss in failure scenarios. In recent years, reliable facility location problems have been studied in the supply chain design (Daskin, 1983; Snyder and Daskin, 2005; Cui et al., 2009) and railroad defect detection sensor design contexts (Ouyang et al., 2009). However, despite these recent efforts, few studies in the network traffic surveillance context have addressed the possibility of sensor failures.

This chapter aims to fill these gaps. It builds on the reliable facility location literature and develops a linear integer model to determine optimal locations for vehicle ID inspection sensors for travel time estimation as well as traffic O-D flow count. The model allows probabilistic sensor failures in general transportation networks. The formulated problem is complex by nature, and the real-world instances are generally of large scale. This imposes prohibitive computational burden if we solve this model with standard solvers. We therefore propose customized algorithms to solve the problem efficiently. Case studies are conducted to test the algorithms and to draw insights.

The chapter has the following layout. Section 2 introduces the notation and develops the mathematical model. Section 3 proposes customized algorithms to solve this problem. Section 4 conducts numerical experiments to draw managerial insights. Section 5 concludes this chapter and briefly discusses future study directions.

4.2 Model Formulation

We select sensor locations in a transportation network to maximize the expected benefit from both O-D volume estimation and travel time measurement. For any O-D flow, the total traffic volume can be inspected by a single sensor if and only if the flow passes the sensor (Yang and Zhou, 1998). In this case, we say that the flow is covered by the sensor in the sense of *flow coverage*. Such individual sensor information can also be used to infer travel time based on speed measurements (Ban et al., 2009). However, sensors (particularly those with vehicle-ID capabilities) can work in pairs to provide an accurate measurement of travel time between their installation locations. Assume that the traffic state along the traffic paths remains relatively stable during the nominal travel time.[a] Intuitively, accurate travel time estimation for an O-D path benefits all traffic on this path, while the accuracy depends on the span of sensors—the wider a pair of sensors span over an O-D path, the larger portion of the path is measured and the better it helps to estimate travel time of that O-D path. Thus the travel time surveillance benefit, which we denote by *path coverage*, depends on not only the inspected traffic volume but also the lengths of covered O-D paths by sensor pairs. We assume for simplicity that path coverage for an O-D path is proportional to both its traffic volume and covered length.

Let $\mathcal{I}$ be the set of O-D paths on the network. Each path $i \in \mathcal{I}$ is specified by its traffic volume f_i, which is assumed to be deterministic and known. Each path i passes a set of candidate locations, $\mathcal{J}_i$, where sensors can be potentially installed. Each candidate location j on path i has a corresponding mileage, m_{ij}, increasing along the traffic direction of f_i. The collection of all candidate locations over the network is $\mathcal{J} := \bigcup_{\forall i} \mathcal{J}_i$. For convenience of notation, let $\mathcal{I}_j$ denote the set of paths that pass the same location j. Note that $\bigcup_{\forall j} \mathcal{I}_j = \mathcal{I}$.

Due to limited budget, no more than N sensors can be built on the network. For $\forall i \in \mathcal{I}_j$, f_i is inspected if an operational sensor is located at j. Similar to the traditional maximal covering models (Yang and Zhou, 1998), if f_i is inspected by at least one sensor, the benefit of flow coverage is $b_c f_i$, where b_c is a nonnegative coefficient. If f_i passes at least two sensors, we can record its travel time between the first functioning (*head*) sensor it passes, at location j^h, and the last functioning (*rear*) sensor it passes, at location j^e. The benefit of path coverage can be expressed as $b_t f_i (m_{ij^e} - m_{ij^h})$, where b_t is also a nonnegative coefficient.

In the long run, sensors may be disrupted or malfunctional from time to time. When sensors fail, the flow coverage and path coverage patterns in the network also change. Hence we consider the expected surveillance benefit across all sensor failure scenarios in addition

[a] Without losing generality a path can be divided into multiple short segments to make this assumption reasonable.

to the ideal non-failure scenario. The head (or rear) sensor for each i may vary over different failure scenarios. In other words, different head (or rear) sensors are assigned to i according to failure scenarios. Sensors on i can be ranked into different priority levels according to such head (or rear) assignment such that in any scenario the sensor with the lowest level among all functioning ones, if available, is the head (or rear) sensor. In the normal scenario (without any failure), the most upstream sensor on i serves as the head sensor, and thus it is the level-zero head sensor for i. If this sensor fails, its immediately downstream sensor takes over to serve i, and thus this second sensor is the level-one head sensor for i. This process can be repeated to label every installed sensor on i with a unique head sensor assignment level. Similarly, each sensor on i can be labeled with a unique rear sensor assignment level that starts from zero for the most downstream sensor and increases upstream. Supposing that there are S_i sensors installed on path i, we see that once the locations with installations on i are given (i.e., $\{j_0^i, j_1^i, .., j_{S_i-1}^i\}$ ordered from upstream to downstream), their head and rear assignment levels are determined by the following simple rule

Definition 1. *(Valid assignment rule)* A sensor at location j_s^i is the level-s head sensor and the level-$(S_i - 1 - s)$ rear sensor for traffic path i.

Since each sensor installed on i receives a unique head (or rear) assignment level to i, there are at most $R_i := \min(|\mathcal{J}_i|, N)$ levels of possible head (or rear) assignment. Let $r = 0, 1, \cdots, R_i - 1$ denote a possible head (or rear) assignment level for a sensor on i.

The primal decision variables $\mathbf{x} := \{x_j\}$ determine where to install sensors, where

$$x_j = \begin{cases} 1, & \text{if a sensor is installed at location } j; \\ 0, & \text{otherwise.} \end{cases}$$

Given $\mathbf{x}$, the auxiliary variables $\mathbf{h} = \{h_{ijr}\}$ and $\mathbf{e} = \{e_{ijr}\}$ decide how sensors are assigned to paths according to the valid assignment rule; i.e.,

$$h_{ijr} = \begin{cases} 1, & \text{if a sensor is installed at } j \text{ and it is assigned to } i \text{ as a level-}r \text{ head sensor;} \\ 0, & \text{otherwise,} \end{cases}$$

and

$$e_{ijr} = \begin{cases} 1, & \text{if a sensor is installed at } j \text{ and it is assigned to } i \text{ as a level-}r \text{ rear sensor;} \\ 0, & \text{otherwise.} \end{cases}$$

Assume that each sensor fails independently with an identical probability $0 \leq q < 1$. This probability can be obtained from historic sensor performance statistics (Rajagopal and

Varaiya, 2007). The objective of this two-sensor-covering problem (TSC) is to maximize the expected total benefit of flow coverage and path coverage for all O-D paths.

$$\text{(TSC)} \quad \max_{\mathbf{x}} z(\mathbf{x}) := \max_{\mathbf{h},\mathbf{e}} \sum_{i\in\mathcal{I}} \sum_{j\in\mathcal{J}_i} \sum_{r=0}^{R_i-1} q^r(1-q) f_i[-b_t m_{ij} h_{ijr} + (b_t m_{ij} + b_c) e_{ijr}], \tag{4.1}$$

subject to

$$\sum_{j\in\mathcal{J}} x_j \leq N, \tag{4.2}$$

$$\sum_{r=0}^{R_i-1} h_{ijr} = x_j, \forall i \in \mathcal{I}, \forall j \in \mathcal{J}_i, \tag{4.3}$$

$$\sum_{r=0}^{R_i-1} e_{ijr} = x_j, \forall i \in \mathcal{I}, \forall j \in \mathcal{J}_i, \tag{4.4}$$

$$\sum_{j\in\mathcal{J}_i} h_{ijr} \leq 1, \forall i \in \mathcal{I}, r = 0; \tag{4.5a}$$

$$\sum_{j\in\mathcal{J}_i} h_{ijr} \leq \sum_{j\in\mathcal{J}_i} h_{ij(r-1)}, \forall i \in \mathcal{I}, \forall r = 1, \cdots, R_i - 1, \tag{4.5b}$$

$$\sum_{j\in\mathcal{J}_i} e_{ijr} \leq \sum_{j\in\mathcal{J}_i} h_{ijr}, \forall i \in \mathcal{I}, \forall r = 0, 1, \cdots, R_i - 1, \tag{4.6}$$

$$x_j, h_{ijr}, e_{ijr} \in \{0, 1\}, \forall i \in \mathcal{I}, \forall j \in \mathcal{J}_i, \forall r = 0, 1, \cdots, R_i - 1. \tag{4.7}$$

Constraint (4.2) enforces the budget limit, while constraints (4.3)-(4.7) postulate the valid assignment rule. Constraints (4.3) (or (4.4)) ensure that each installed sensor is assigned to each of its corresponding paths at one and only one head (or rear) assignment level. Constraints (4.5) and (4.6) indicate that no more than one head or rear sensor is assigned to each path at each level, and each rear assignment must be accompanied by a head assignment. Constraints (4.5) also imply that for each path i, all the implemented head assignment levels, $\{r | \sum_{j\in\mathcal{J}_i} h_{ijr} = 1\}$, start from 0 and form a consecutive sequence. Constraints (4.7) postulate all decision variables to be binary.

The following proposition reveals the relationship between the above formulation and the valid assignment rule.

Proposition 4. *The optimal solution to the TSC problem* (4.1)-(4.7) *satisfies the valid assignment rule.*

Proof. Let $\mathbf{x}^*$, $\mathbf{h}^*$, $\mathbf{e}^*$ denote the optimal solution to TSC. Again locations with installed sensors on each path i are indexed with $\{j_0^i, j_1^i, .., j_{S_i-1}^i\}$ from upstream to downstream. Let $\mathcal{R}_i^h$ denote the set of all implemented head assignment levels to i; i.e., $\mathcal{R}_i^h := \{r | \sum_{j\in\mathcal{J}_i} h_{ijr} = 1\}$. Similarly, let $\mathcal{R}_i^e := \{r | \sum_{j\in\mathcal{J}_i} e_{ijr} = 1\}$. For the case of $q = 0$, there is no failure and only the level-0 assignment affects the objective value. It is obvious that the optimal solution enforces all non-trivial assignments (at level-0) to be consistent with the valid assignment rule.

Now we consider the case with $q > 0$. Since each installed sensor on i corresponds to only one implemented head (or rear) assignment level (from (4.3) and (4.4)) and different sensors cannot have the same head (or rear) assignment level (from (4.5) and (4.6)), it is obvious that $|\mathcal{R}_i^h| = |\mathcal{R}_i^e| = S_i$. For the head assignment, due to constraints (4.5), $\mathcal{R}_i^h$ contains a sequence of levels from 0 to $S_i - 1$. Due to constraints (4.6), $\mathcal{R}_i^e \subseteq \mathcal{R}_i^h$. Thus $\mathcal{R}_i^h = \mathcal{R}_i^e = \{0, 1, \cdots, S_i - 1\}$, and we denote them by $\mathcal{R}_i$. Therefore on path i, each sensor j_s^i is labeled with a unique head (or rear) assignment level in $\mathcal{R}_i$. At optimality, a more upstream sensor shall have a lower head assignment level and a higher rear assignment level. Thus j_s^i corresponds to the level-s head assignment and the level-$(S_i - 1 - s)$ rear assignment to i, which is the valid assignment rule. □

It shall be noted that the TSC modal can be easily adapted for cases where existing installations are already present (Ehlert et al., 2006). We simply enforce $x_j = 1$ if a sensor is already installed at location j; the model still has the same structure and complexity.

4.3 Solution Algorithms

TSC is NP-hard because the maximal covering problem is a special case of TSC (with $b_t = 0$ and $q = 0$). As we will show in Section 4, commercial optimization software (e.g., CPLEX) would work well only for small-scale instances but it usually runs into difficulty when problem size increases. We hence propose customized algorithms to obtain near-optimal solutions for large-scale problems. The first algorithm is based on a simple greedy heuristic, which can yield good solutions for many realistic applications. But it does not provide information on how close these solutions are from the true optima. Hence we propose a second algorithm based on Lagrangian relaxation (LR), which provides not only good feasible solutions but also optimality gaps.

4.3.1 Greedy Algorithm

The greedy algorithm for TSC simply selects sensor locations sequentially based on the best marginal increase of objective (4.1), until all N installation locations have been selected. The exact steps are as follows.

Step 0: Initialization. Let the set of selected location indices $\mathcal{Q} := \emptyset$ and the iteration index $n := 1$. Set $x_j = 0, \forall j \in \mathcal{J}$;

Step 1: Search for the n^{th} location that will bring the maximum marginal improvement of objective (4.1); i.e., select

$$j^* = \arg \max_{k \in \mathcal{J} \setminus \mathcal{Q}} \{z(\mathbf{x}') : x'_j = 1, \text{ iff } j \in \mathcal{Q} \bigcup \{k\}\}.$$

The corresponding marginal objective improvement is denoted by $\rho_n := z(\mathbf{x}') - z(\mathbf{x})$, where $x'_j = 1, \text{ iff } j \in \mathcal{Q} \bigcup \{j^*\}$. Let $x_{j^*} = 1$ and $\mathcal{Q} = \mathcal{Q} \bigcup \{j^*\}$.

Setp 2: If $n = N$, stop and return $\mathbf{x}$ and the corresponding objective value $\sum_{n=1}^{N} \rho_n$; otherwise, $n = n + 1$, and go to step 1.

The greedy heuristic is widely applied to many practical problems not only because of its simplicity but also due to its reasonable practical performance. For example, in the case of the classic maximal covering problem (a special case of TSC where $q = 0$ and $b_t = 0$), Feige (1998) proved that the objective value of any greedy solution is no smaller than $(1 - 1/e)$ of the true optimum; i.e., the approximation ratio is $e/(e - 1)$. More importantly, no known polynomial-time algorithm can beat the greedy algorithm in terms of this approximation ratio bound (Feige, 1998).

We can obtain a similar approximation ratio for the maximal covering problem with probabilistic facility failures (a special case of TSC where $b_t = 0$ and $q > 0$), which is stated in the following proposition.

Proposition 5. *For TSC problems with $b_t = 0$ and $q > 0$, the objective value of the greedy algorithm solution is no smaller than $(1 - 1/e)$ of the true optimum.*

Proof. For any $\mathcal{J}' \in \mathcal{J}$, let $C(\mathcal{J}')$ denote the expected coverage benefit (the objective value of (4.1)) given that each location in $\mathcal{J}'$ has a sensor installed. Let j_n denote the n^{th} selected location by the greedy algorithm ($n = 1, 2, \cdots, N$). Define $\mathcal{J}_n^G := \{j_1, j_2, \cdots, j_n\}$ and $\mathcal{J}_0^G := \emptyset$. For convenience of notation, let $B^r := q^r(1 - q)f_i b_c$. Since $b_t = 0$, the valid

assignment rule yields $C(\mathcal{J}_n^G) = \sum_{i\in\mathcal{I}}(\sum_{r=0}^{|\mathcal{J}_n^G \bigcap \mathcal{J}_i|-1} B^r)$. Then

$$\rho^n = C(\mathcal{J}_n^G) - C(\mathcal{J}_{n-1}^G) = \sum_{i\in\mathcal{I}_{j_n}} (B^{|\mathcal{J}_{n-1}^G \bigcap \mathcal{J}_i|}) = \max_{j\in\mathcal{J}\setminus\mathcal{J}_{n-1}^G:} \sum_{i\in\mathcal{I}_j} (B^{|\mathcal{J}_{n-1}^G \bigcap \mathcal{J}_i|}). \tag{4.8}$$

Since $C(\mathcal{J}_0^G) = 0$, $z^G := C(\mathcal{J}_N^G) = \sum_{n=1}^N \rho^n$.

Let the optimal installations be $\mathcal{J}^* := \{j_1^*, j_2^*, \cdots, j_N^*\}$, which yield the true optimal objective value $z^* := C(\mathcal{J}^*)$. Based on the valid assignment rule, $z^* = \sum_{i\in\mathcal{I}} \sum_{r=0}^{|\mathcal{J}^* \bigcap \mathcal{J}_i|-1} B^r$. Also,

$$C(\{j_{n'}^*\} \bigcup \mathcal{J}_{n-1}^G) - C(\mathcal{J}_{n-1}^G) = \begin{cases} \sum_{i\in\mathcal{I}_{j_{n'}^*}} (B^{|\mathcal{J}_{n-1}^G \bigcap \mathcal{J}_i|}), & \text{if } j_{n'}^* \notin \mathcal{J}_{n-1}^G; \\ 0, & \text{otherwise,} \end{cases}$$

which is no greater than ρ_n based on (4.8).

Then

$$\begin{aligned} z^* - C(\mathcal{J}_{n-1}^G) &= \sum_{i\in\mathcal{I}} (\sum_{r=0}^{|\mathcal{J}^* \bigcap \mathcal{J}_i|-1} B^r - \sum_{r=0}^{|\mathcal{J}_{n-1}^G \bigcap \mathcal{J}_i|-1} B^r) \\ &\le \sum_{i\in\mathcal{I}} |(\mathcal{J}^* \bigcap \mathcal{J}_i) \setminus (\mathcal{J}_{n-1}^G \bigcap \mathcal{J}_i)| B^{|\mathcal{J}_{n-1}^G \bigcap \mathcal{J}_i|} \\ &= \sum_{j\in\mathcal{J}^*\setminus\mathcal{J}_{n-1}^G} \sum_{i\in\mathcal{I}_j} (B^{|\mathcal{J}_{n-1}^G \bigcap \mathcal{J}_i|}) \\ &= \sum_{j\in\mathcal{J}^*} [C(\{j\} \bigcup \mathcal{J}_{n-1}^G) - C(\mathcal{J}_{n-1}^G)] \\ &\le \sum_{j\in\mathcal{J}^*} \rho_n = N\rho_n. \end{aligned}$$

Hence,

$$\rho^n \ge \frac{z^* - C(\mathcal{J}_{n-1}^G)}{N}, \forall n = 1, 2, \cdots, N,$$

which yields

$$z^* - C(\mathcal{J}_n^G) \le (z^* - C(\mathcal{J}_{n-1}^G))(1 - 1/N) \le \cdots \le z^*(1 - 1/N)^n,$$

and

$$z^G \ge z^*[1 - (1 - 1/N)^N] \ge z^*(1 - 1/e).$$

□

For general TSC, however, the approximation ratio of the proposed greedy algorithm is not bounded. This can be seen from the following simple example. Suppose a network has three nodes $\mathcal{J} = \{1, 2, 3\}$, two links $\{(1, 2), (2, 3)\}$, and two consecutive O-D flow paths, i.e., $\mathcal{I} = \{a, b\}$ with $f_a = 0, f_b = 1$, and $\mathcal{J}_a = \{1, 2\}$ and $\mathcal{J}_b = \{2, 3\}$. If $b_c = 0$, $b_t > 0$ and $N = 2$, a possible solution from the greedy algorithm is $\mathcal{Q} = \{1, 2\}$, which yields $z(\mathbf{x}) = 0$. Yet the optimal solution is obviously $\mathcal{Q} = \{2, 3\}$, which gives a positive objective value. Hence, the proposed greedy algorithm for TSC does not have a performance bound and we propose an LR-based algorithm in the next section. In the following we also discuss generalized greedy algorithms (e.g., selecting $k = 2, 3, ...$ sensor locations simultaneously each time) and demonstrate that their approximation ratios remain unbounded.

The greedy algorithm described above can be generalized by selecting $k = 2, 3, \cdots$ sensor locations simultaneously, as follows.

Step 0: Initialization. Let $\mathcal{Q} := \emptyset$ and $x_j := 0, \forall j \in \mathcal{J}$;

Step 1: Search for the next k (or the maximum number allowed by the budget) locations that will bring the largest increase of (4.1); i.e., select

$$\mathcal{J}^* = \arg \max_{\substack{\mathcal{J}' \subset \mathcal{J} \setminus \mathcal{Q} \\ |\mathcal{J}'| = \min\{k, N - |\mathcal{Q}|\}}} \{z(\mathbf{x}) : x_j = 1, \text{ iff } j \in \mathcal{Q} \bigcup \mathcal{J}'\}.$$

Let $x_{j^*} = 1, \forall j^* \in \mathcal{J}^*$ and $\mathcal{Q} := \mathcal{Q} \bigcup \mathcal{J}^*$.

Setp 2: If $|\mathcal{Q}| = N$, stop and return $\mathbf{x}$; otherwise, go to step 1.

Again, the approximation ratio of the generalized greedy algorithm is unbounded below by any positive number. This can be seen from the example in Figure 4.1, where $G(d)$ denotes a complete subgraph containing d nodes. The network contains n subgraphs of type $G(k)$ $(n > 1)$, one subgraph of type $G(nk)$, and n connectors. The length of every edge in the network is 1. Each edge within a complete subgraph is an O-D flow path. The traffic volume is 1.1 if the edge is within a type $G(k)$ subgraph, or 1 if it is within the subgraph of type $G(nk)$. There is no traffic flow on connector edges. Suppose $b_c = 0$, $b_t > 0$ and $N = nk$.

At each step, the generalized greedy algorithm will select all k nodes from one of the type $G(k)$ subgraph in order to obtain the maximum marginal improvement of the objective. As a result, the greedy solution will select all nk nodes from all type $G(k)$ subgraphs, yielding an objective value of $z^G = 1.1nk(k-1)b_t/2$. However, the true optimal solution is obviously the set of nodes in the type $G(nk)$ subgraph, with an optimal objective value of $z^* = nk(nk-1)b_t/2$. Since $\lim_{n\to\infty} z^G/z^* = 0$, the generalized greedy algorithm does not have a positive approximation ratio bound for large scale cases (i.e., $n \to \infty$).

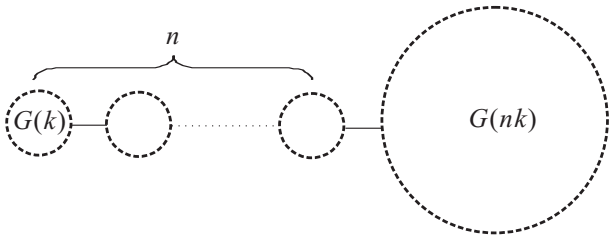


Figure 4.1: Example for the performance bound of the generalized greedy algorithm.

4.3.2 LR-based Algorithm

Relaxed Subproblems and Bounds

We relax constraints (4.5) and (4.6), and add them to the objective (4.1) with nonnegative Lagrangian multipliers $\lambda = \{\lambda_{ir}\}$ and $\gamma = \{\gamma_{ir}\}$, respectively. The relaxed TSC (RTSC) becomes:

$$\text{(RTSC)} \quad \min_{\lambda,\gamma\geq 0} z_R(\lambda,\gamma) := \max_{\mathbf{x},\mathbf{h},\mathbf{e}} \left[\sum_{i\in\mathcal{I}} \sum_{j\in\mathcal{J}_i} \sum_{r=0}^{R_i-1} (H_{ijr}h_{ijr} + E_{ijr}e_{ijr}) + \sum_{i\in\mathcal{I}} \lambda_{i0} \right] \tag{4.9}$$

s.t. (4.2)-(4.4) and (4.7), where the benefit of an installation at location j as a level-r head sensor for any $i \in \mathcal{I}_j$ is

$$H_{ijr} = \begin{cases} -q^r(1-q)f_i b_t m_{ij} - \lambda_{ir} + \lambda_{i(r+1)} + \gamma_{ir}, & r = 0, 1, \cdots, R_i - 2; \\ -q^r(1-q)f_i b_t m_{ij} - \lambda_{ir} + \gamma_{ir}, & r = R_i - 1, \end{cases} \tag{4.10}$$

and the benefit of this installation as a level-r rear sensor is

$$E_{ijr} = q^r(1-q)f_i(b_t m_{ij} + b_c) - \gamma_{ir}. \tag{4.11}$$

For any given λ and γ, the exact value of $z_R(\lambda,\gamma)$ provides an upper bound of (4.1), and it can be obtained from the following decomposition scheme. When (4.5) and (4.6) are relaxed, assignments are no longer dependent across j. Constraints (4.3) require that the rear assignment of each j with sensor installed is conducted at exactly one level for each $i \in \mathcal{I}_j$. Thus to achieve the optimal benefit, j is assigned to i as a head sensor at the level corresponding to the maximum H_{ijr} across all r. Similarly, the corresponding rear assignment level is chosen to maximize E_{ijr} across all r. Therefore, in RTSC, the contribution of installing a sensor at j, in terms of objective (4.9), is

$$B_j = \sum_{i \in I_j} [\max_r(H_{ijr}) + \max_r(E_{ijr})]. \tag{4.12}$$

Obviously, the optimal solution to (4.9) is to set $x_j = 1$ for the N locations with the largest B_j values, and accordingly, set $h_{ijr} = 1$ (or $e_{ijr} = 1$) if $x_j = 1$ and r maximizes H_{ijr} (or E_{ijr}) across all r.[b] Then the optimal objective value of RTSC is

$$z_R(\lambda, \gamma) = \sum_{j \in \mathcal{J}} B_j x_j + \sum_{i \in \mathcal{I}} \lambda_{i0}. \tag{4.13}$$

Since the solution obtained from the above procedure is probably not feasible to the original TSC problem, heuristic methods are used to construct a feasible solution. Although such constructive heuristics do not guarantee the exact optimal solution, previous experiments (Cornuejols et al., 1977; Caprara et al., 1999) yield very good feasible, often exactly optimal, solutions (and tight lower bounds of the original objectives) if the Lagrangian multipliers are near convergence. One simple heuristic is that we install all facilities that are obtained from RTSC, and then apply the valid assignment rule to determine the feasible **h** and **e** accordingly. If the lower bound equals the upper bound at any iteration, then the optimal solution is found. Otherwise, the difference between these bounds provides an optimality gap - the difference between the true optimum and the feasible solution is sure to be no larger than this gap.

For the classic maximal covering problem ($q = 0$ and $b_t = 0$), Cornuejols et al. (1977) proved that the relative gap between the optimal LR solution and the optimal TSC solution is bounded by $1/e$. The following proposition provides conditions under which this bound holds for problems with positive failure probability $q > 0$.

Proposition 6. *For TSC problems with $b_t = 0$ and $q > 0$, the optimal objective value* (4.1) *for the original TSC is no smaller than* $(1 - 1/e)$ *of the optimal LR objective* (4.9) *if* $q \leq \min\{\rho_n/\rho_{n-1}, \forall n = 2, 3, \cdots, N\}$.

Proof. The notation follows Proof 4.3.1. Let z_R^* denote the optimal LR objective (4.9). Let

$$u_{ir}^n := \begin{cases} B^r, & r \leq |\mathcal{J}_n^G \bigcap \mathcal{J}_i| - 1; \\ 0, & \text{otherwise.} \end{cases} \quad \forall n = 0, 1, \cdots, N.$$

Then $\sum_{r=0}^{R_i - 1} u_{ir}^n = \sum_{r=0}^{|\mathcal{J}_G^n \bigcap \mathcal{J}_i| - 1} B^r$ represents the expected coverage benefit of $\mathcal{J}_G^n$ associated with path i, and thus $C(\mathcal{J}_n^G) = \sum_{i \in \mathcal{I}} \sum_{r=0}^{R_i - 1} u_{ir}^n$.

[b] Ties can be broken arbitrarily.

For any $n \in \{1, 2, \cdots, N\}$, let

$$\lambda_{ir} = \begin{cases} u_{ir}^{n-1}, & \forall i \in \mathcal{I}, r = R_i - 1; \\ u_{ir}^{n-1} + \lambda_{i(r+1)}, & \forall i \in \mathcal{I}, \forall r = 0, \cdots, R_i - 2, \end{cases} \tag{4.14}$$

which yields $\lambda_{i0} = \sum_{r=0}^{R_i-1} u_{ir}^{n-1}$, and let

$$\gamma_{ir} = u_{ir}^{n-1}, \forall i \in \mathcal{I}, \forall r = 0, 1, \cdots, R_i - 1. \tag{4.15}$$

It is obvious that λ and γ are nonnegative and feasible for equation (4.9). Since $b_t = 0$, plug (4.14) and (4.15) into and (4.10) and (4.11), respectively, and we obtain $H_{ijr} = 0$ and $E_{ijr} = B^r - u_{ir}^{n-1}$. Hence, equation (4.9) yields,

$$z_R^* \leq z_R(\lambda, \gamma) = \max_{\sum_{j\in\mathcal{J}} x_j = N} \sum_{j\in\mathcal{J}} x_j \sum_{i\in\mathcal{I}_j} \max_{r\in\{0,1,\cdots,R_i-1\}} (B^r - u_{ir}^{n-1}) + \sum_{i\in\mathcal{I}} \sum_{r=0}^{R_i-1} u_{ir}^{n-1}$$

Note that

$$\max_{r\in\{0,1,\cdots,R_i-1\}} (B^r - u_{ir}^{n-1}) = \begin{cases} B^{|\mathcal{J}_{n-1}^G \bigcap \mathcal{J}_i|}, & \text{if } |\mathcal{J}_{n-1}^G \bigcap \mathcal{J}_i| \leq R_i - 1; \\ 0, & \text{otherwise.} \end{cases} \tag{4.16}$$

Equation (4.8) and (4.16) yield the following: when $n = 1$, $\sum_{i\in\mathcal{I}_j} \max_{r\in\{0,1,\cdots,R_i-1\}}(B^r - u_{ir}^{n-1}) \leq \rho_n, \forall j \in \mathcal{J}$; when $n \in \{2, 3, \cdots, N\}$, $\sum_{i\in\mathcal{I}_j}, \max_{r\in\{0,1,\cdots,R_i-1\}}(B^r - u_{ir}^{n-1}) \leq \rho_n, \forall j \in \mathcal{J}\backslash\mathcal{J}_{n-1}^G$ and $\sum_{i\in\mathcal{I}_j} \max_{r\in\{0,1,\cdots,R_i-1\}}(B^r - u_{ir}^{n-1}) \leq q\rho_{n-1} \leq \rho_n, \forall j \in \mathcal{J}_{n-1}^G$. Hence, $z_R^* \leq N\rho_n + C(\mathcal{J}_{n-1}^G)$ and $\rho_n \geq \frac{z_R^* - C(\mathcal{J}_{n-1}^G)}{N}$. Similarly, we obtain $z^G \geq z_R^*(1 - 1/e)$, and hence $z^* \geq z^G \geq z_R^*(1 - 1/e)$.

Remark 1. In case that we allow multiple installations at the same location (i.e., $x_j = 0, 1, 2, \cdots$), the approximation bound stated in Proposition 6 will holds for all $q \in [0, 1)$. The greedy algorithm shall allow repeated selection of the same candidate location, and hence $\sum_{i\in\mathcal{I}_j} \max_{r\in\{0,1,\cdots,R_i-1\}}(B^r - u_{ir}^{n-1}) \leq \rho_n$, for all $j \in \mathcal{J}$, regardless of whether $q \leq \rho_n/\rho_{n-1}$.

□

It should be noted that the computational time for solving the RTSC problem (4.9) and for obtaining a feasible solution are bounded by $O(N \cdot |\mathcal{I}| + |\mathcal{J}| \sum_{i\in\mathcal{I}} R_i)$ and $O(N \cdot |\mathcal{I}|)$, respectively.

Multiplier Updating

Function $z_R(\lambda, \gamma)$ is known to be convex over λ and γ. RTSC can be solved with an iterative subgradient search. We update λ and γ iteratively to find the tightest upper bound $\min_{\lambda,\gamma\geq 0} z_R(\lambda, \gamma)$. We add subscript k to distinguish variables in iteration k. The initial values of the multipliers are obtained with heuristics (e.g., the dual solution to the linear relaxation of the original problem). At the end of each iteration, multipliers are updated as follows.

$$\lambda_{ir}^{k+1} = \max\left(0, \lambda_{ir}^{k} + t^k \Delta\lambda_{ir}^{k}\right), \forall i \in \mathcal{I}, \forall r = 0, 1, \cdots, R_i - 1, \tag{4.17}$$

$$\gamma_{ir}^{k+1} = \max\left(0, \gamma_{ir}^{k} + t^k \Delta\gamma_{ir}^{k}\right), \forall i \in \mathcal{I}, \forall r = 0, 1, \cdots, R_i - 1, \tag{4.18}$$

where the subgradients are $\Delta\lambda_{ir}^{k} := \sum_{j\in\mathcal{J}_i} h_{ijr} - \begin{cases} 1, & r = 0 \\ \sum_{j\in\mathcal{J}_i} h_{ij(r-1)}, & \text{otherwise} \end{cases}$, and $\Delta\gamma_{ir}^{k} := \sum_{j\in\mathcal{J}_i} (e_{ijr} - h_{ijr})$. Step size t_k is usually set to

$$t^k = \frac{\mu^k (z_R(\lambda^k, \gamma^k) - z^{LB})}{\sum_{i\in\mathcal{I}} \sum_{r=0}^{R_i-1} \left[(\Delta\lambda_{ir}^{k})^2 + (\Delta\gamma_{ir}^{k})^2\right]},$$

where μ^k is a control scaler, and z^{LB} is the objective value of the best-known feasible solution. Traditionally, control scaler μ^k is determined by setting $\mu^0 = 2$ and halving μ^k if $z_R(\lambda^k, \gamma^k)$ is not improved after a fixed number of iterations (Fisher, 1981). This approach is modified by Caprara et al. (1999) for faster convergence. The idea is to set $\mu^0 = 0.1$, and compare the best and worst values of $z_R(\lambda^k, \gamma^k)$ in every certain number (e.g., 20) of iterations: decrease μ^k if the difference is greater than a larger threshold (e.g., 1%) and increase μ^k if the difference is less than a smaller threshold (e.g., 0.1%). In our case study, we use the traditional approach when multipliers are far from their optimal values and then switch to the second approach near convergence.

In principle, the LR algorithm is terminated if one of the following conditions is satisfied: (i) the lower bound equals the upper bound, (ii) the optimality gap stops reducing, and (iii) the solution time exceeds a reasonable limit. Our experience shows that condition (ii) terminates the algorithm most of the time. In case that happens, we may use the following branch and bound procedure to further reduce the optimality gap.

Branch and Bound

If the aforementioned LR algorithm ends up having a non-zero optimality gap, we implement the LR algorithm into a branch and bound framework. We branch on variables $\mathbf{x}$ in a

depth-first manner, and use a greedy heuristic to choose the next variable x_j for branching: installation at j shall bring in the greatest increase of the objective value (4.1) given the variables that have already been branched. We branch each variable first to 1 (enforcing installation) and then to 0 (forbidding installation). At each node, we run the LR algorithm to determine the lower and upper bounds, while extra constraints for already-branched variables are exerted. If the upper bound is lower than the best feasible solution so far, the node no longer has potential and is trimmed. If the current node has already had N enforced or $|\mathcal{J}| - N$ forbidden installations, only one non-trivial feasible solution exists and is returned as both the lower and the upper bounds. At each branching, the multipliers of a parent node are passed down to its child nodes as their initial multipliers.

4.4 Case Studies

This section presents two numerical examples of the TSC model. All solution algorithms are implemented on a PC with 2.0 GHz CPU and 2 GB memory. For the LR-based algorithm, we denote the optimal objective value by z^*, the solution time by T, and the optimality gap by ϵ. Let z^G be the objective value found by the greedy algorithm. For comparison, we solve the same instances with commercial software CPLEX, and let z^C, T^C and ϵ^C be the objective value, the solution time and the residual optimality gap, respectively. Let $\alpha := b_t/(b_t + b_c)$ be an indicator of the relative importance of path coverage benefit.[c]

4.4.1 Sioux-Falls Network

The Sioux-Falls network has 24 vertices and 76 links, as shown in Figure 4.2[d]. Assume that all the vertices are candidate locations, i.e., $|\mathcal{J}| = 24$. There are 528 traffic O-D pairs. For simplicity, we assume that each O-D pair only has one flow path that is determined by the shortest path algorithm[e], and hence $|\mathcal{I}| = 528$. Assume too that the sensor at a node can detect all traffic passing that node from different directions.

We set a solution time limit of 1800 seconds, and run a series of instances for $b_t = 1$, $b_c \in \{0, 1, 10\}$, $N \in \{3, 5, 7\}$ and $q \in \{0, 0.05, 0.2, 0.5\}$. The results are summarized in Table 4.1. As we can see, the LR-based algorithm found optimal solutions for almost all the instances ($\epsilon = 0\%$). CPLEX has a comparable performance only when α is small (i.e., flow coverage dominates). Otherwise, the performance of CPLEX is significantly worse than the

[c]Note that once α is fixed, scaling the value of b_t (or b_c) does not affect the optimal sensor deployment.

[d]Source: http://www.bgu.ac.il/ bargera/tntp/.

[e]An alternative is to obtain traffic flows within a traffic assignment framework. This will not change the structure of the model though.

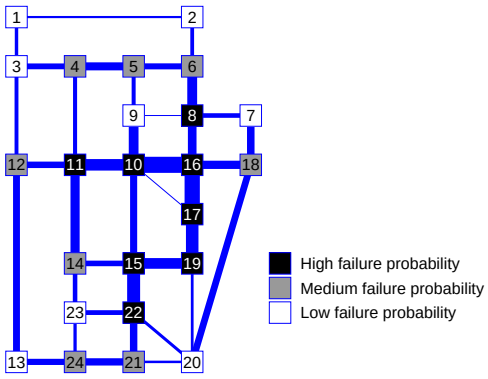


Figure 4.2: The Sioux-Falls test network.

LR-based algorithm: CPLEX cannot find the optimal solution within 1800 seconds for many instances, and sometimes it cannot even find a meaningful feasible solution (where $z^C = 0$ or $\epsilon^C = \text{INF}\%$). The greedy algorithm finds a good feasible solution (i.e., $z^G \approx z^*$) when α is small. For most instances with $\alpha = 1$, however, the results from the greedy algorithm are quite far from the optima. This implies that the greedy algorithm does not work as well when path coverage is the dominating objective. This is probably because a sensor's contribution to path coverage highly depends on other sensors' locations.

In Table 4.1, z^* increases with N and decreases with q, as expected. Figure 4.3 further reveals their relationships by plotting z^* over N and q for different parameter values. In Figure 4.3(a), curves 1 and 2 are for path coverage only ($\alpha = 1$) and curves 3 and 4 are for flow coverage only ($\alpha = 0$). We see that curves 3 and 4 quickly flatten out while curves 1 and 2 almost linearly increase until N is close to $|\mathcal{J}|$. This suggests that path coverage benefit is more sensitive to value of N. This is probably because the marginal path coverage benefit depends not only on the additional installation itself, but also on other installations that form pairs with the additional one. The differences between curves 1 and 2, and that between 3 and 4 represent the expected coverage loss due to probabilistic sensor failures. Although such loss is small for flow coverage, it is significant for path coverage. This is further confirmed by Figure 4.3(b) which shows how z^* varies with q. Curves 5 and 6 are for path coverage while curves 7 and 8 are for flow coverage. We see that when q is not too large (e.g., $q < 0.5$, which is true for most real-world cases), curves 5 and 6 drop much faster than curves 7 and 8. This confirms the observation that the benefit loss due to failures is more significant for path coverage. In this case, sensor failures should be addressed carefully. It is also interesting to notice that curves 5 and 6 are rather convex while 7 and 8 are rather

Table 4.1: Results for Sioux-Falls test network.

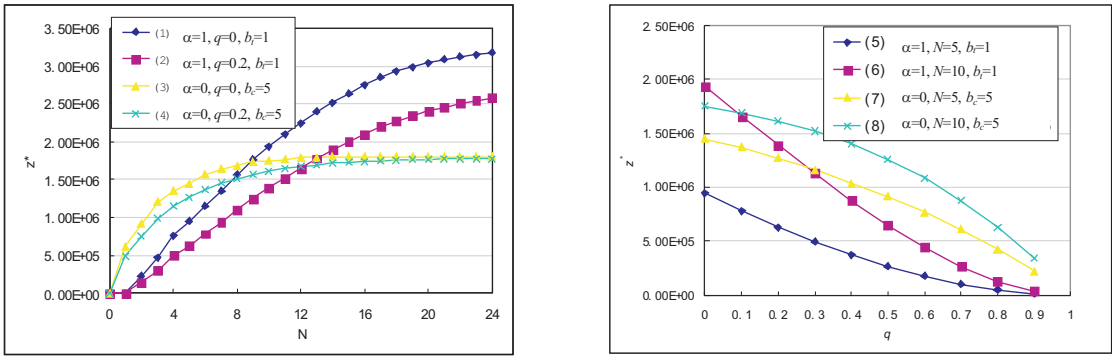
#	N	q	b_c	α	z^G	z^*	z^C	T/s	T^C/s	ϵ	ϵ^C
1	3	0	0	1.00	230600	469200	469200	59	66	0 %	0 %
2	3	0	1	0.50	692800	692800	692800	8	53	0 %	0 %
3	3	0	5	0.17	1.59E+06	1.59E+06	1.59E+06	1	1	0 %	0 %
4	3	0.05	0	1.00	208117	423453	423453	60	414	0 %	0.01 %
5	3	0.05	1	0.50	640371	640371	640371	8	288	0 %	0.01 %
6	3	0.05	5	0.17	1.51E+06	1.51E+06	1.51E+06	2	1	0 %	0 %
7	3	0.2	0	1.00	150426	300288	287168	73	>1800	0 %	18.29 %
8	3	0.2	1	0.50	494320	494320	494320	11	341	0 %	0.01 %
9	3	0.2	5	0.17	1.27E+06	1.27E+06	1.27E+06	2	1	0 %	0 %
10	3	0.5	0	1.00	62000	119838	92925	270	>1800	0 %	86.68 %
11	3	0.5	1	0.50	252775	252775	252775	29	>1800	0 %	3.24 %
12	3	0.5	5	0.17	794675	794675	794675	1	1	0 %	0 %
13	5	0	0	1.00	662800	947800	947800	44	113	0 %	0 %
14	5	0	1	0.50	1.22E+06	1.22E+06	1.22E+06	23	107	0 %	0 %
15	5	0	5	0.17	2.31E+06	2.31E+06	2.31E+06	6	8	0 %	0 %
16	5	0.05	0	1.00	607112	861901	861901	52	>1800	0 %	0.97 %
17	5	0.05	1	0.50	1.13E+06	1.13E+06	1.13E+06	16	270	0 %	0.01 %
18	5	0.05	5	0.17	2.19E+06	2.19E+06	2.19E+06	5	13	0 %	0.01 %
19	5	0.2	0	1.00	507213	625062	588058	123	>1800	0 %	17.26 %
20	5	0.2	1	0.50	872339	872339	865773	35	>1800	0 %	2.05 %
21	5	0.2	5	0.17	1.87E+06	1.87E+06	1.87E+06	4	5	0 %	0 %
22	5	0.5	0	1.00	213788	266725	0	642	>1800	0 %	INF %
23	5	0.5	1	0.50	449163	449163	443650	81	>1800	0 %	5.82 %
24	5	0.5	5	0.17	1.18E+06	1.18E+06	1.18E+06	2	2	0 %	0 %
25	7	0	0	1.00	1.28E+06	1.35E+06	1.35E+06	110	125	0 %	0 %
26	7	0	1	0.50	1.65E+06	1.65E+06	1.65E+06	59	176	0 %	0.01 %
27	7	0	5	0.17	2.92E+06	2.92E+06	2.92E+06	30	1	0 %	0 %
28	7	0.05	0	1.00	1.18E+06	1.24E+06	1.24E+06	124	1783	0 %	0.01 %
29	7	0.05	1	0.50	1.54E+06	1.54E+06	1.54E+06	56	599	0 %	0.01 %
30	7	0.05	5	0.17	2.78E+06	2.78E+06	2.78E+06	9	4	0 %	0 %
31	7	0.2	0	1.00	897554	936031	0	376	>1800	0 %	INF %
32	7	0.2	1	0.50	1.22E+06	1.22E+06	1.21E+06	111	>1800	0 %	1.04 %
33	7	0.2	5	0.17	2.36E+06	2.36E+06	2.36E+06	8	7	0 %	0 %
34	7	0.5	0	1.00	388425	411363	0	1800	>1800	26 %	INF %
35	7	0.5	1	0.50	622325	625738	0	576	>1800	0 %	INF %
36	7	0.5	5	0.17	1.48E+06	1.48E+06	1.48E+06	8	7	0 %	0 %

concave, indicating opposite sensitivity behaviors in different q value ranges.

Figure 4.4 shows the impact of α and q on the optimal sensor deployment. The link width illustrates flow volumes. The dark nodes are the optimal installation locations, which are generally at places with heavy traffic. The optimal deployment for path coverage ($\alpha = 1$) is more spread-out than that for flow coverage ($\alpha = 0$). This is intuitive because more scattered sensor pairs can cover longer paths. On the other hand, higher failure probability generally leads to a higher degree of sensor clustering.

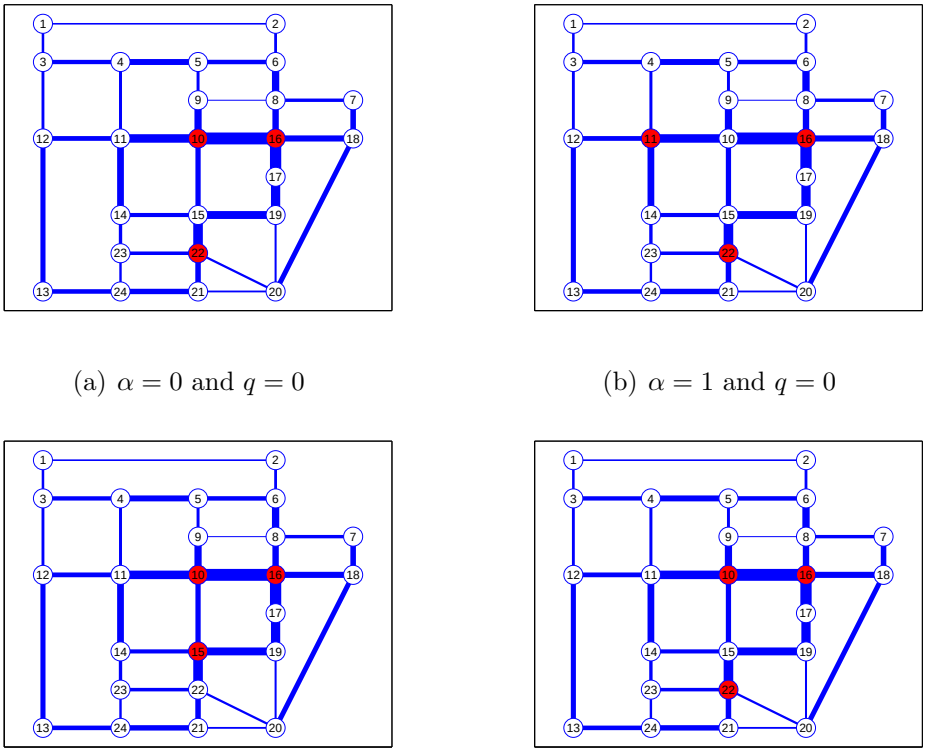
4.4.2 Chicago Intermodal Network

Figure 4.5 shows the geometry of Chicago interstate highway network, which contains 21 highway junctions and 17 railroad terminals, which are the railroad yards to upload and download intermodal freights. Traffic comes in and goes out of the network through 8 access points. Each highway junction is split into multiple candidate locations (Sheffi, 1985)



(a) (b)

Figure 4.3: Relationship between N, q and z^* for the Sioux-Falls network.



(a) $\alpha = 0$ and $q = 0$ (b) $\alpha = 1$ and $q = 0$

(c) $\alpha = 0$ and $q = 0.2$ (d) $\alpha = 1$ and $q = 0.2$

Figure 4.4: Optimal deployment of $N = 3$ installations in the Sioux-Falls network.

such that an installation at any candidate location can inspect all passing flows. The final network representation includes 89 candidate locations and 363 connecting links. The 2002 intermodal freight traffic[f] originated from or destined to Chicago is grouped into 1046 O-D paths on this network based on population distribution. Due to lack of detailed information, we again assume that all O-D flows follow their shortest distance paths.

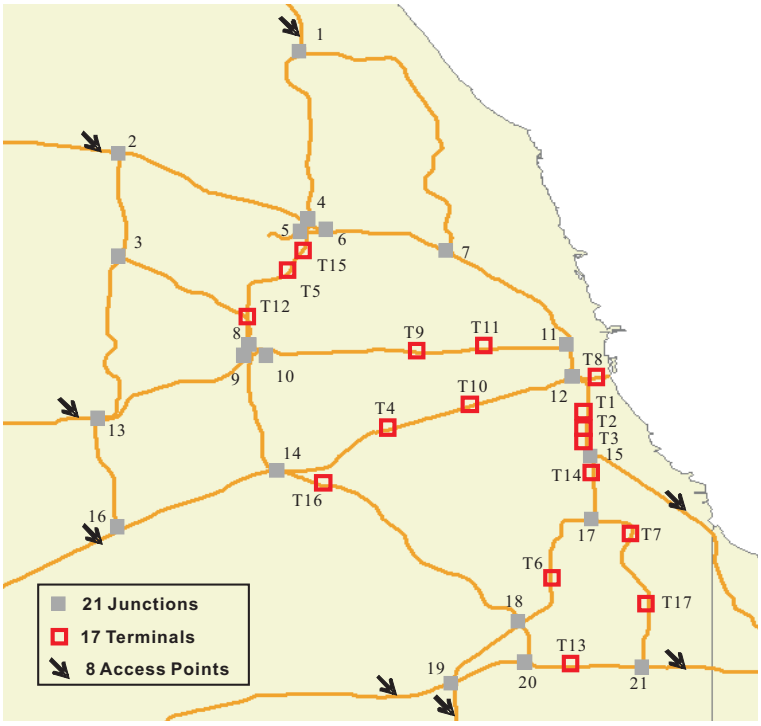
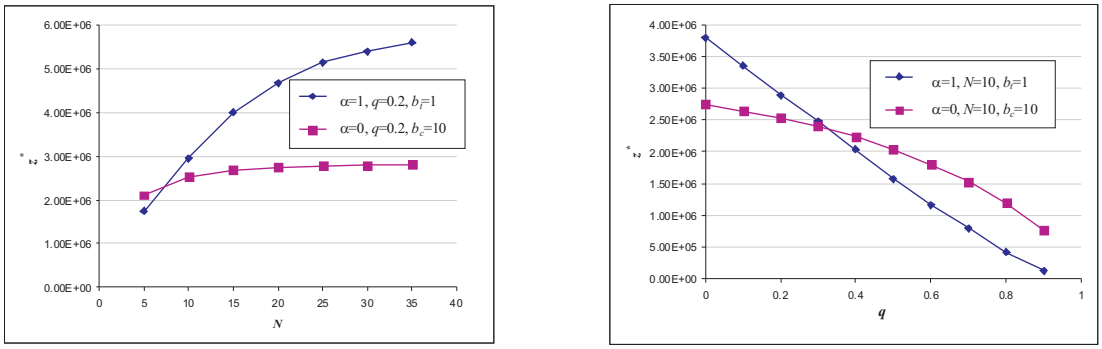


Figure 4.5: Chicago intermodal network.

A maximum solution time of 1200 seconds is enforced while the model is applied with a range of parameter values. Table 4.2 summarizes the results. Due to the increased problem size, CPLEX cannot even get a meaningful feasible solution for most instances. The LR-based algorithm always yields a near-optimum solution with a reasonable residual gap ($\leq$ 15%). From our experiments, the difference between the near-optimal solution and the optimum is often much smaller than the residual gap. Thus these solutions are suitable for engineering practice.

Figure 4.6 shows again that path coverage is much more sensitive to changes of N and q than flow coverage. Figure 4.7 illustrates how α and q affect the optimal sensor deployment. Again, the optimal deployment for path coverage tends to be more spread-out, as highlighted by the solid ellipses. For flow coverage, higher failure probability generally leads to a higher

[f]Data source: Bureau of Transportation Statistics, http://www.bts.gov/.



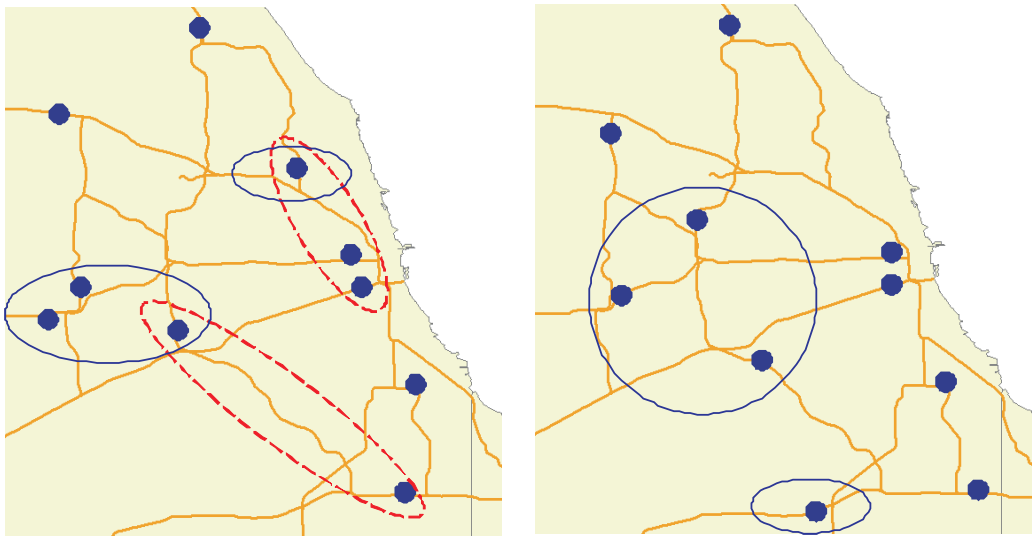
(a) (b)

Figure 4.6: Relationship between N, q and z^* for the Chicago intermodal network.

degree of sensor clustering, as highlighted by the dashed ellipses. Such clustering effect is not as obvious for path coverage probably due to the need for sensors to cover more path length.

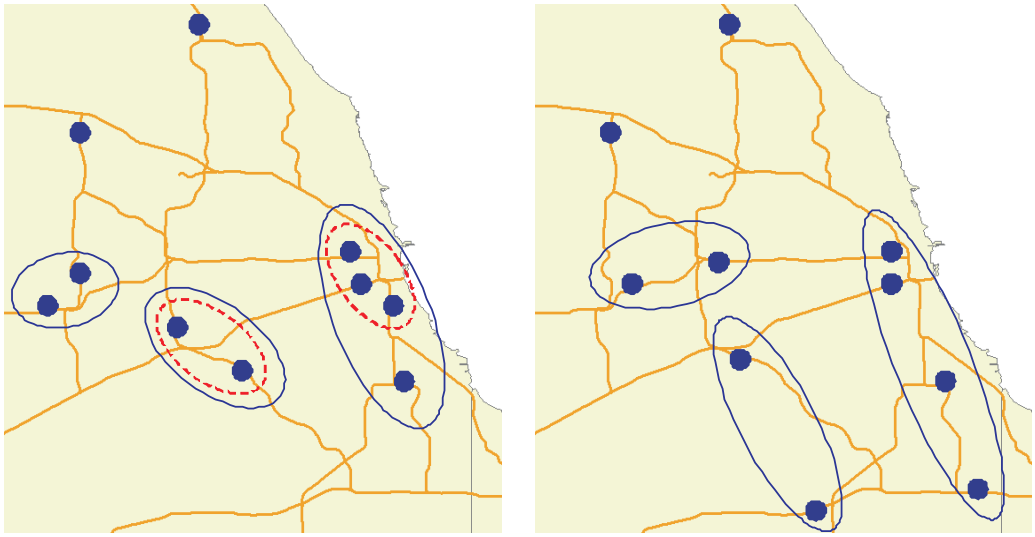
4.5 Full-Scale Implementation in Railroad Networks

The proposed model is applied to a full-scale real problem of railroad wayside defect detection installations. We obtained empirical data from a major U.S. railroad on its network topology and traffic information for 30-, 60-, and 90-day intervals. According to the railroad specifications, we consider individual sensor flow coverage of railcars ($b_t = 0$) only and ignore sensor failures ($q = 0$). Though model (4.1)-(4.7) now reduces to a maximal covering problem, the size of the problem is much larger. The original data contain more than 10,000 candidate locations in the network, about half a million distinct railcars conducting about 2 million shipments per month. Because of the large scale, preprocessing was conducted to eliminate dominated candidate locations and merge railcar flows with the same itinerary. Since only flow coverage is considered, if $\mathcal{I}_{j'} \subset \mathcal{I}_j$ for some $j, j' \in \mathcal{J}$, then location j' is dominated by location j and can be excluded from the optimal solution because all the railcars that can be potentially inspected by installing at location j' could have been equivalently inspected by installing at j. If $\mathcal{J}_{i'} = \mathcal{J}_i$ for some $i, i' \in \mathcal{I}$, then flow i and flow i' have exactly the same itinerary and can be merged into one new railcar flow whose volume equals $f_i + f_{i'}$. Also, the huge amount of data is stored in a sparse matrix format and integrated into the LR algorithm to save memory and increase processing speed. To further improve the efficiency of the LR algorithm, we temporarily store the values of the Lagrangian multipliers at convergence. These multiplier values can be used as the starting multiplier values for similar problem instances (e.g., after we slightly vary the installation budget).



(a) $\alpha = 0$ and $q = 0$

(b) $\alpha = 1$ and $q = 0$



(c) $\alpha = 0$ and $q = 0.2$

(d) $\alpha = 1$ and $q = 0.2$

Figure 4.7: Optimal deployment of $N = 10$ installations in the Chicago intermodal network.

Table 4.2: Results for Chicago intermodal network.

#	N	q	b_c	α	z^G	z^*	z^C	T/s	T^C/s	ϵ	ϵ^C
1	10	0	0	1	3.80E+06	4.22E+06	1.95E+06	1200	1200	5 %	120 %
2	10	0	1	0.5	4.12E+06	4.48E+06	0	1200	1200	4 %	INF %
3	10	0	4	0.2	4.84E+06	5.28E+06	0	1200	1200	3 %	INF %
4	10	0	1	0	274219	275462	275461	1200	1200	5 %	0 %
5	10	0.2	0	1	2.93E+06	3.00E+06	0	1200	1200	15 %	INF %
6	10	0.2	1	0.5	3.14E+06	3.25E+06	0	1200	1200	12 %	INF %
7	10	0.2	4	0.2	3.82E+06	3.99E+06	0	1200	1200	9 %	INF %
8	10	0.2	1	0	253215	253408	253408	1200	48	5 %	0 %
9	10	0.5	0	1	1.57E+06	1.69E+06	0	1200	1200	15 %	INF %
10	10	0.5	1	0.5	1.85E+06	1.85E+06	0	1200	1200	14 %	INF %
11	10	0.5	4	0.2	2.38E+06	2.41E+06	0	1200	1200	9 %	INF %
12	10	0.5	1	0	203567	203567	203567	1200	6	7 %	0 %
13	20	0	0	1	5.60E+06	5.78E+06	5.82E+06	1200	838	10 %	0 %
14	20	0	1	0.5	6.01E+06	6.08E+06	6.10E+06	1200	820	10 %	0 %
15	20	0	4	0.2	6.86E+06	6.91E+06	6.94E+06	1200	608	10 %	0 %
16	20	0	1	0	283361	283361	0	1200	1200	9 %	INF %
17	20	0.2	0	1	4.71E+06	4.75E+06	0	1200	1200	9 %	INF %
18	20	0.2	1	0.5	4.95E+06	5.02E+06	0	1200	1200	8 %	INF %
19	20	0.2	4	0.2	5.79E+06	5.84E+06	0	1200	1200	7 %	INF %
20	20	0.2	1	0	274962	275057	274480	1200	516	7 %	0 %
21	20	0.5	0	1	2.93E+06	2.99E+06	0	1200	1200	5 %	INF %
22	20	0.5	1	0.5	3.19E+06	3.22E+06	0	1200	1200	5 %	INF %
23	20	0.5	4	0.2	3.90E+06	3.93E+06	0	1200	1200	3 %	INF %
24	20	0.5	1	0	244680	244680	240456	1200	49	7 %	2 %

While the branch and bound procedure is no longer efficient due to the huge number of variables, the designed LR algorithm alone can yield very good results—on a PC with a 2.3 GHz CPU, the LR algorithm can yield near-optimal solutions (optimality gap 3%) in about 1 hour for all computed cases. The objective function values (i.e., the number of inspected distinct cars) are quite close for 30, 60, and 90 days of traffic. The optimality gap can be further reduced by increasing computational time, but the marginal computational effort needed increases dramatically as the gap itself gets closer to 0. For example, if we reduce the tolerable optimality gap from 3% to 2%, the extra computational time for each problem instance is about 1 hour on average.

The railroad also provided information on its current wayside detector installations. Compared with the existing installations on this railroad network, the solution from the proposed model (with the same number of installations) will improve the inspection benefit by a relative amount ranging from 20% to 60%.

For this wayside defect detection location problem, a stand-alone computer program, Railroad Wayside Detector Location Solver (RWDLS), was developed to determine the best set of locations that inspect the maximum number of railcar flows. Figure 4.8 shows the interfaces. The left dialog box provides flexible input options for problem customization, the middle one determines the subset of railcars that are inspected by any given set of locations, and the right one graphically presents the result summary and statistics. For more information about this software, see Li and Ouyang (2007). Figure 4.9 shows the

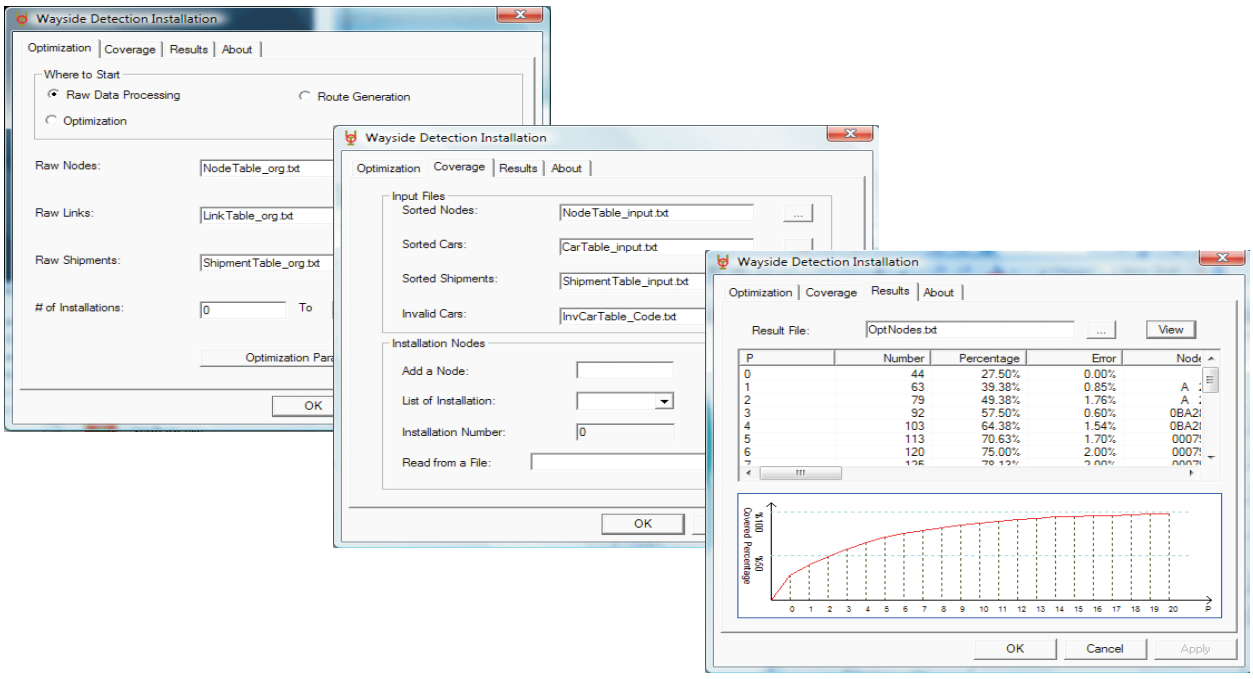


Figure 4.8: Software interface of railroad wayside detection installation locations.

actual railcar coverage for the railroad company under 7 and 12 installations. On the railroad network, the width of a green (red) segment illustrates the number of covered (uncovered) railcars passing this location. We see that 7 installations already cover about over 80% railcars and 12 installations further improve the coverage to over 90%.

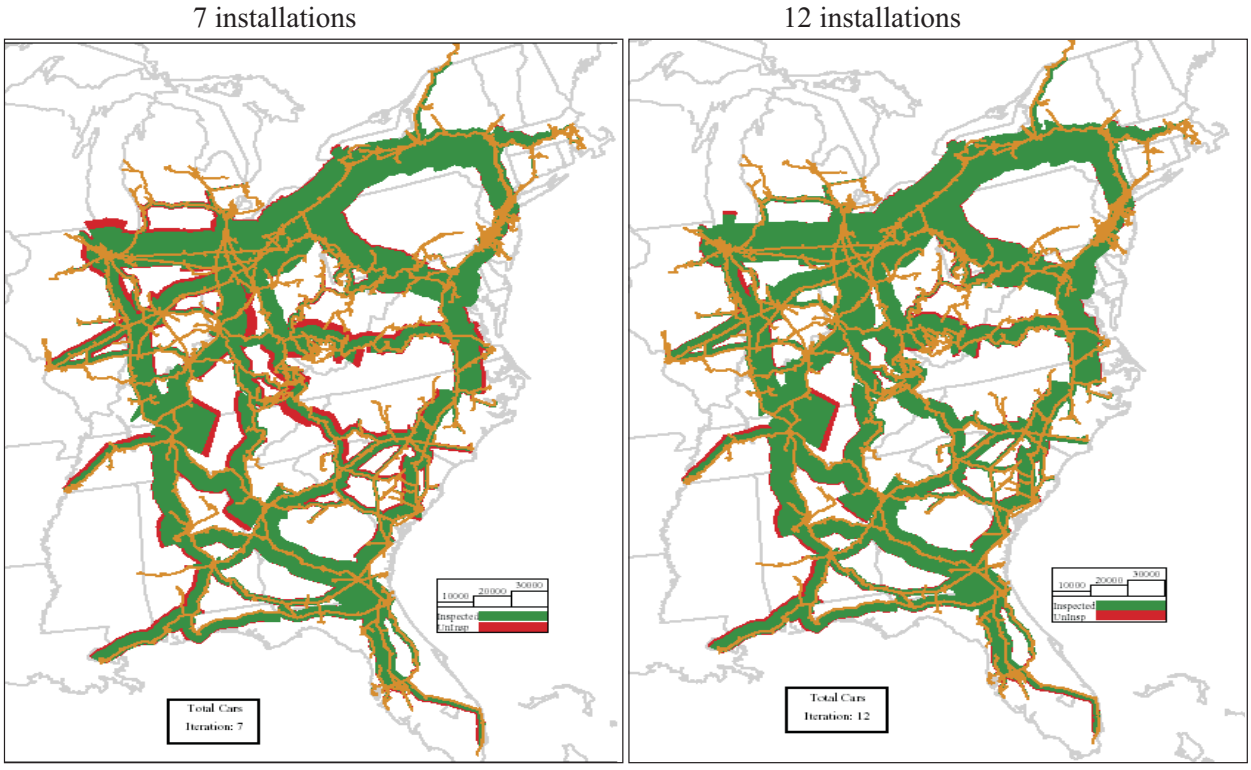


Figure 4.9: Optimal railcar coverage with $N = 7$ (left) and $N = 12$ (right) installations.

4.6 List of Symbols

b_c: Nonnegative coefficient for flow coverage

b_t: Nonnegative coefficient for path coverage

B^r: $q^r(1-q)f_i b_c$

$C(\mathcal{J}')$: Expected coverage benefit given sensor installation location set $\mathcal{J}'$

$\mathbf{e} = \{e_{ijr}\}$: $e_{ijr} = 1$ ($e_{ijr} = 0$) if a sensor is (not) installed at j and it is assigned to i as a level-r rear sensor

f_i: The traffic volume on path $i \in \mathcal{I}$

$G(d)$: Complete subgraph containing d nodes

$\mathbf{h} = \{h_{ijr}\}$: $h_{ijr} = 1$ ($h_{ijr} = 0$) if a sensor is (not) installed at j and it is assigned to i as a level-r head sensor

$\mathcal{I}$: Set of O-D paths on the network

$\mathcal{I}_j$: Set of paths that pass the same location $j \in \mathcal{J}$

j^h: Location for a head sensor

j^e: Location for a rear sensor

$\mathcal{J}$: Set of all candidate locations

$\mathcal{J}_i$: Set of candidate locations on path $i \in \mathcal{I}$

m_{ij}: Mileage of candidate location $j \in \mathcal{J}_i$ on path $i \in \mathcal{I}$

N: Maximum number of facilities that the budget allows to build

q: Site-independent sensor failure probability

$\mathcal{Q}$: Set of locations

R_i: Number of levels of possible head (or rear) assignment for path $i \in \mathcal{I}$

S_i: Number of sensors installed on path $i \in \mathcal{I}$

T: Solution time for the LR algorithm

T^C: Solution time for CPLEX

$\mathbf{x} := \{x_j\}$: $x_j = 1$ ($x_j = 0$) if a sensor is (not) installed at j

$z(\mathbf{x})$: Total coverage benefit for sensor deployment $\mathbf{x}$

z^C: Optimal CPLEX objective

z^G: Total coverage benefit from the greedy algorithm

z^*: Optimal total coverage benefit from the greedy algorithm

z_R^*: Optimal LR objective

z^{LB}: Objective value of the best-known feasible solution in LR

α: $b_t/(b_t + b_c)$

ϵ: Optimality gap for the LR algorithm

ϵ^C: Optimality gap for CPLEX

$\gamma = \{\gamma_{ir}\}$: Lagrangian multipliers for rear assignments

$\lambda = \{\lambda_{ir}\}$: Lagrangian multipliers for head assignments

μ^k: Control scaler in LR

ρ_n: Marginal benefit of the n^{th} installation in the greedy algorithm

Chapter 5

Sensor Deployment under Site-Dependent Failure and Generalized Surveillance Effectiveness Measures

This chapter aims to extend the sensor location model in Chapter 4 into a more generalized framework that (i) addresses an overarching surveillance effectiveness measure to unify existing measures; and (ii) allows sensors to fail with site-dependent probabilities. We define a novel surveillance effectiveness measure based on the reduction of estimation error that is capable of encompassing many well-known measures (e.g., flow coverage, path coverage and state estimation error). A compact model is formulated to minimize the total expected estimation error for all O-D paths on the transportation network across all possible sensor failure scenarios, subject to site-dependent sensor failures. A range of customized solution algorithms are investigated to solve this problem efficiently. Case studies are conducted to test the performance of proposed algorithms and draw useful insights on sensor deployment benefit.

5.1 Motivating Example

The lack of consideration on a unifying surveillance effectiveness measure and site-dependent sensor failures may lead to a dramatically different sensor deployment and significantly inferior surveillance effectiveness. Figure 5.1 shows a simple traffic network with two symmetric 100-mile O-D paths that share a 90-mile highway segment. The flow volume on each path

is equal to 1. Candidate sensor installation locations (marked as squares) are indexed by their mileposts. Sensors installed at locations $20, 25, \cdots 80$ (lighter squares) will be perfectly reliable (i.e., with zero failure probabilities), while those installed at all other locations (darker squares) fail independently with a 30% probability[a]. Table 1 compares the optimal surveillance effectiveness of three sensors under different effectiveness measures and different "perceptions of failure". Solution 1 is the optimal sensor location design when sensor failure is completely ignored, solution 2 assumes that all candidate locations are subject to an identical failure probability of 13% (which is about the average probability across all candidate locations), while solution 3 takes into account the true site-dependent failure probabilities. Under the vehicle-mile coverage measure, solution 1 will obviously deploy sensors at the three ends 0, 100_1 and 100_2 so as to cover all the vehicle-miles in this network. Solution 2 deploys all three sensors on the shared highway segment so that they can back up each other against potential failures. Solution 3 installs two sensors at perfectly reliable locations 20 and 80 in consideration of site-dependent senor failures. As a result of misperceptions of site-dependent sensor failure probabilities, the first two solutions only yield suboptimal benefits (or effectivenesses) that are respectively 33.8% and 22.5% lower than that from solution 3. Alternatively, we could measure surveillance effectiveness by the squared error of traffic state estimation (i.e., the smaller the square error, the better the effectiveness). In this simple illustrative example, we assume that the error is defined in the following way: for a path segment that is incident to two neighboring functioning sensors, the error equals the square of the segment length; for a segment that is incident to only one or zero sensor, the error is four times the squared segment length. Under this measure, sensors tend to be distributed in the middle of the paths to avoid large squared errors from long path segments. The actual surveillance effectivenesses for solution 1 and solution 2 are both 76% worse than that for solution 3.

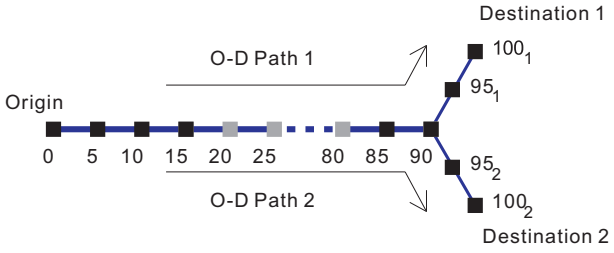


Figure 5.1: A motivating example.

Table 5.1 has revealed the drastic impact of effectiveness measures and site-dependent sensor failure probabilities on the optimal surveillance effectiveness and sensor deployment.

[a]It is not rare for loop detectors to have such a high failure probability; see Rajagopal and Varaiya (2007).

Table 5.1: Result summary for the motivating example.

Measure type	Solution #	Optimal sensor locations	Actual surveillance benefit/error	Percentage difference
Vehicle -mile coverage	1	$0, 100_1, 100_2$	98	33.8%
	2	$0, 5, 90$	114.66	22.5%
	3	$0, 20, 80$	148	0%
Squared error	1	$10, 50, 90$	17600	76%
	2	$10, 50, 90$	17600	76%
	3	$20, 50, 80$	10000	0%

This highlights the need for a network-level reliable sensor location design framework that (i) addresses an overarching surveillance effectiveness measure that encompasses most existing measures; and (ii) allows sensors to fail with site-dependent probabilities. Traffic surveillance effectiveness is defined as the reduction of "generalized estimation errors" on all highway segments between neighboring sensor pairs, such that the existing flow volume coverage, vehicle-mile coverage and squared estimation error measures can all be expressed as special cases. The objective of the proposed model is to minimize the total expected estimation error for all O-D paths on the transportation network across all possible sensor failure scenarios, subject to site-dependent sensor failures. Like many other location design problems, the deterministic version of the sensor location model is already complex; considering an exponential number of failure scenarios will further increase the difficulty—the computational burden will be prohibitive if we solve this problem with traditional approaches. In this work we develop an innovative compact mixed integer programming formulation for this problem and propose a range of customized solution algorithms to solve this problem efficiently. Case studies are conducted to test the proposed algorithms and to draw useful insights on how the surveillance measure definitions and various parameters (e.g., sensor failure probability and its spatial heterogeneity) impact optimal sensor deployment. We also present alternative problem formulations and algorithms, including a continuous approximation model for the sensor deployment problem on a highway corridor.

The remainder of the chapter is organized as follows. Section 2 introduces the overarching surveillance effectiveness measure and develops the compact mixed integer program (MIP) model for optimal sensor location design. Section 3 proposes a range of customized algorithms to solve this problem. Section 4 presents alternative models including a continuous approximation model and a fixed-charge location model. Section 5 conducts case studies to test the solution algorithms and draw managerial insights. Section 6 makes concluding remarks and briefly discusses future research directions.

5.2 Model Formulation

5.2.1 Generalized Surveillance Effectiveness

Let $\mathcal{I}$ be the set of O-D traffic flow paths on the network. Each path $i \in \mathcal{I}$ with traffic flow volume v_i passes a set of candidate locations $\bar{\mathcal{J}}_i$ for potential sensor installations. Set $\bar{\mathcal{J}} := \bigcup_{\forall i} \bar{\mathcal{J}}_i$ contains all candidate locations for sensor installations. Without loss of generality, we add two virtual locations u and d to the transportation network, each with an installed imaginary sensor that never fails. For every path i, connect u (and d) to the origin (and the destination) of the path with virtual links of zero length, such that under any sensor deployment design each segment on path i will be incident to exactly two sensors (including the imaginary sensors). Let $\mathcal{J}_i := \bar{\mathcal{J}}_i \bigcup \{u, d\}$ and $\mathcal{J} := \bar{\mathcal{J}} \bigcup \{u, d\} = \bigcup_{\forall i} \mathcal{J}_i$. For each $j \in \mathcal{J}_i$, let $\mathcal{J}_{ij+}$ denote the set of candidate locations downstream to j (not including j) on path i, and let $\mathcal{J}_{ij-}$ denote the set of locations upstream to j (i.e., $\mathcal{J}_{ij-} = \mathcal{J}_i \backslash (\mathcal{J}_{ij+} \bigcup \{j\})$). Define $\mathcal{J}_{ijk} := \mathcal{J}_{ij+} \backslash (\mathcal{J}_{ik+} \bigcup \{k\}), \forall k \in \mathcal{J}_{ij+}$, which denotes the candidate locations between j and k on path i. For convenience of notation, let $\mathcal{I}_j$ denote the set of paths that pass the same location j, where $\bigcup_{\forall j} \mathcal{I}_j = \mathcal{I}$.

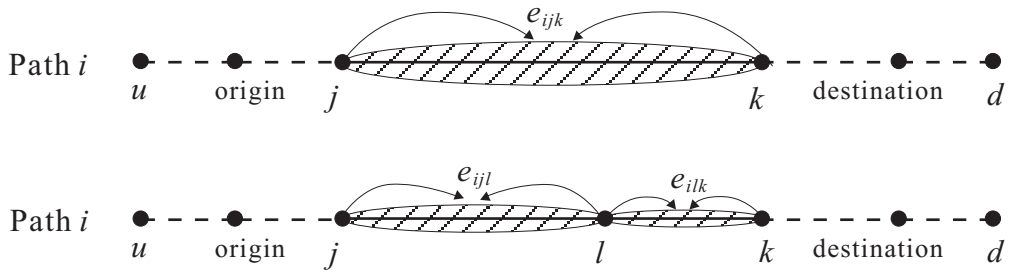


Figure 5.2: Neighboring sensor estimation error measure.

We define a general traffic state estimation error measure e_{ijk} for the segment on path i in between locations $j \in \mathcal{J}_{id-}$ and $k \in \mathcal{J}_{ij+}$, if the estimation is based on surveillance data from sensors at j and k. Widely used estimation approaches include simple interpolation and Newell's three detector method (Newell, 1993). Error e_{ijk} can be interpreted as the integral or summation of estimation inaccuracies from all neighborhoods on segment $j - k$ (or the negative value of coverage benefits, as shown with the examples at the end of this subsection). As illustrated in Figure 5.2, an additional sensor installation at the intermediate location l normally will not impair estimation accuracy on segment $j - k$ (in most cases, it helps improve estimation accuracy); i.e., it is reasonable to assume that $e_{ijl} + e_{ilk} \leq e_{ijk}, \forall i \in \mathcal{I}, j \in \mathcal{J}_{id-}, l \in \mathcal{J}_{ijd}, k \in \mathcal{J}_{il+}$. Note that the possible contribution of an "outsider" sensor at location k to error e_{ijl} is negligible if two functioning sensors are available at j, l, and

$k \notin \mathcal{J}_{ijl} \bigcup \{j, l\}$. In order to minimize the total estimation error along the entire path, only the immediate neighboring sensors should be used to estimate (or measure) the traffic state on the segments inbetween. Note that $\forall j \in \bar{\mathcal{J}}_i$, e_{iuj} (or e_{ijd}) actually represents the estimation error for the segment from the upstream end to location j (or from j to the downstream end) only with data from one real sensor at j, and e_{iud} is the benchmark estimation error for the entire path i without using any sensor data.[b] Suppose that there are S_i sensors (in addition to the two imaginary ones) installed on path i whose locations are $j_{i1}, \cdots, j_{iS_i} \in \bar{\mathcal{J}}_i$ ordered from upstream to downstream, and we further define $j_{i0} = u$ and $j_{i(S_i+1)} = d$. The surveillance effectiveness measure for path i is defined as $e_{iud} - \sum_{s=0}^{S_i} e_{ij_sj_{s+1}}$, i.e., the change of estimation errors with or without the S_i sensors. The network surveillance effectiveness can be expressed as

$$\sum_{i \in \mathcal{I}} e_{iud} - \sum_{i \in \mathcal{I}} \sum_{s=0}^{S_i} e_{ij_sj_{s+1}} \tag{5.1}$$

Since the first term is a constant, a sensor location problem of maximizing the network surveillance effectiveness can be equivalently solved by minimizing the total estimation errors $\sum_{i \in \mathcal{I}} \sum_{s=0}^{S_i} e_{ij_sj_{s+1}}$.

Now we will see how several existing surveillance effectiveness measures can be expressed in terms of $\{e_{ijk}\}$. For all $j \in \mathcal{J}_{id-}, k \in \mathcal{J}_{ij+}$, we let a_{ijk} denote the distance from j to k along path i.

Example 1 The flow volume coverage (FV) assumes that the surveillance benefit is proportional to the total path flow volume intercepted by all sensors (e.g., Yang and Zhou (1998), Li and Ouyang (2010)). If path flow i (with volume v_i) passes at least one installed sensor, then it contributes to the total benefit by $b_i^c v_i$, where b_i^c is the benefit coefficient. The network FV benefit measure is hence $\sum_{i \in \mathcal{I}, S_i \geq 1} b_i^c v_i$.

It can be easily verified that if the general error measure $\{e_{ijk}\}$ is defined as follows

$$\begin{aligned} &e_{iud} = b_i^c v_i, \ e_{ij_0j_1} = \frac{a_{ij_0j_1} - a_{iud}}{a_{iud}} b_i^c v_i, \\ &e_{ij_sj_{s+1}} = \frac{a_{ij_sj_{s+1}}}{a_{iud}} b_i^c v_i, \forall i \in \mathcal{I}, s = 1, 2, \cdots, S_i, \end{aligned} \tag{5.2}$$

then (5.1) is equivalent to the network FV coverage.

Example 2 The vehicle-mile coverage (VM) measures the total path flow-length that is covered by sensor pairs (Mirchandani et al., 2009; Li and Ouyang, 2010). The surveillance benefit for path i is assumed to be $b_i^t v_i a_{ij_1j_{S_i}}$, i.e., the product of coefficient

[b] This is possible when other data sources such as historical records are available.

b_i^t, traffic volume v_i, and segment length $a_{ij_1j_{S_i}}$. It can be shown that if we specify $\{e_{ijk}\}$ as follows,

$$e_{iud} = b_i^t v_i a_{iud},\ e_{ij_0j_1} = b_i^t v_i a_{ij_0j_1},\ e_{ij_{S_i}j_{S_i+1}} = b_i^t v_i a_{ij_{S_i}j_{S_i+1}},$$
$$e_{ij_sj_{s+1}} = 0, \forall i \in \mathcal{I}, s = 1, 2, \cdots, S_i - 1, \tag{5.3}$$

then (5.1) becomes the network VM coverage.

Example 3 The squared-error reduction (SER) measure computes the difference of the total squared error between (i) traffic state estimation without using sensor data and (ii) the estimation based on traffic state reconstruction from sensor data (Ban et al., 2009). Suppose that path i starts from mileage 0 and ends at mileage M_i and let M_{ij} denote its mileage at candidate location j. Each neighborhood $x \in [0, M_i]$ has a ground-truth traffic state $w(x)$, which is usually unknown. Let $\hat{w}(x)$ denote the estimated traffic state using data from either the closest sensor or sensor pair around x. The squared error of state estimation on path i is then specified as $\int_0^{M_i}(w(x)-\hat{w}(x))^2dx$. Before sensors are installed on path i, the estimation of $w(x)$, which is denoted by $\bar{w}(x)$, has to be obtained from offline data only and shall be less accurate than $\hat{w}(x)$. Hence, the SER measure for the network is $\sum_{i\in\mathcal{I}}\int_0^{M_i}\left[(w(x)-\bar{w}(x))^2-(w(x)-\hat{w}(x))^2\right]dx$, which is exactly equal to (5.1) by setting

$$e_{iud} = \int_0^{M_i}(w(x)-\bar{w}(x))^2dx,$$
$$e_{ij_sj_{s+1}} = \int_{M_{ij_s}}^{M_{ij_{s+1}}}(w(x)-\hat{w}(x))^2dx, \forall i \in \mathcal{I}, s = 1, \cdots, S_i. \tag{5.4}$$

5.2.2 Formulation

In the long run, sensors may be disrupted or malfunctional from time to time. We assume that failures of different sensors are independent, and a sensor installed at location $j \in \mathcal{J}$ has a site-dependent failure probability $0 \le q_j < 1$. Recall that both imaginary sensors are always functioning, i.e. $q_u = q_d = 0$, so that under any failure scenario every location along a path always has functioning sensors upstream and downstream. Given a sensor deployment on a path i, in any sensor failure scenario, a functioning sensor $j \in \mathcal{J}_{id-}$ will always be paired up with its nearest downstream functioning neighbor in $\mathcal{J}_{ij+}$ (which may be at different locations under different failure scenarios) to estimate the traffic state on the path segment inbetween. We rank sensor locations in $\mathcal{J}_{ij+}$ into different levels (starting with 0) according to the priority for them to pair up with j; i.e., in any scenario when the

sensor at j is functioning, the sensor at the lowest level location (among all those locations in $\mathcal{J}_{ij+}$ with functioning sensors) will be paired up with this sensor. In Figure 5.3, the installation locations on path i are again given as $\{j_{i0}, j_{i1}, .., j_{i(S_i+1)}\}$ ordered from upstream to downstream with $j_{i0} = u$ and $j_{i(S_i+1)} = d$. A sensor at j_{is} and its first downstream neighbor at $j_{i(s+1)}$ will always work together whenever they are both functioning, and we say that they are paired up at level 0. If the sensor at $j_{i(s+1)}$ fails, the next downstream sensor at $j_{i(s+2)}$ takes over and pairs up with the sensor at j_{is} at level 1. This process can be repeated so that each downstream sensor is assigned a unique level to pair up with the sensor at j_{is}, which is described by the following simple rule.

Definition 2. *(Valid pairing-up rule)* A sensor at $j_{is} \in \mathcal{J}_{id-}$ pairs up with a sensor at $j_{i(s+r+1)} \in \mathcal{J}_{ij_{is}+}$ at level $r, \forall r = 0, 1, \cdots S_i - s$.

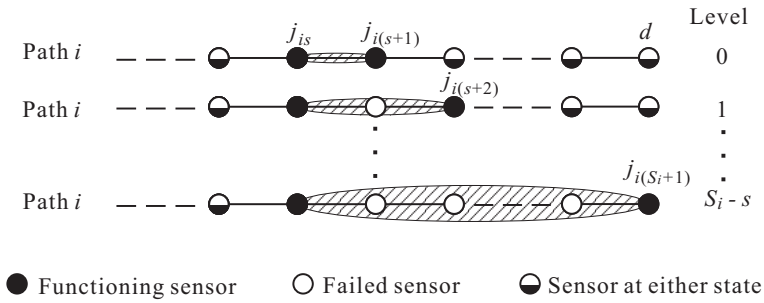


Figure 5.3: Pairing-up levels between the sensor at j_{is} and its downstream sensors on path i.

Due to the budget constraint, no more than N sensors can be installed in the network. The primary decision variables $\mathbf{X} := \{X_j\}_{j \in \mathcal{J}}$ determine sensor locations, where

$$X_j = \begin{cases} 1, & \text{if a sensor is installed at location } j; \\ 0, & \text{otherwise.} \end{cases} \tag{5.5}$$

Based on the valid pairing-up rule, the maximum possible pairing-up level for two sensors at $j \in \mathcal{J}_{id-}$ and $k \in \mathcal{J}_{ij+}$ is $R_{ijk} := \min\{|\mathcal{J}_{ij+}| - |\mathcal{J}_{ik+}| - 1, N\}$. Also define $R_{ij} := \max_{k \in \mathcal{J}_{ij+}} R_{ijk} = \min\{|\mathcal{J}_{ij+}| - 1, N\}$. Given $\mathbf{X}$, the first set of auxiliary variables $\mathbf{Y} = \{Y_{ijkr} | i \in \mathcal{I}, j \in \mathcal{J}_{id-}, k \in \mathcal{J}_{ij+}, r = 0, \cdots, R_{ijk}\}$ decide how sensors are paired up at each level;

$$Y_{ijkr} = \begin{cases} 1, & \text{if sensors at } j \text{ and } k \text{ are paired up at level } r \text{ on path } i; \\ 0, & \text{otherwise.} \end{cases} \tag{5.6}$$

The second set of auxiliary variables $\mathbf{P} = \{P_{ijkr} | i \in \mathcal{I}, j \in \mathcal{J}_{id-}, k \in \mathcal{J}_{ij+}, r = 0, \cdots, R_{ijk}\}$ specify the probability that sensors at j and k are paired up at level r on path i given sensor deployment $\mathbf{X}$.

The objective of this reliable neighboring-sensor-pair-covering problem (RNSPC) is to determine the optimal sensor deployment that minimizes the expected total estimation errors for the whole network across all sensor failure scenarios. However, this objective is difficult to quantify even for a given sensor deployment $\mathbf{X}$ because of the exponential number (i.e., 2^N) of possible sensor failure scenarios (combinations of all sensors' binary states). To address this challenge, we propose a methodology below to consolidate the failure scenarios such that we only need to deal with a polynomial number of scenarios.

As illustrated in Figure 5.3, for any $r = 0, \cdots, S_i - s$, we can consolidate all scenarios in which (i) sensors at j_{is} and $j_{i(s+r+1)}$ are functioning and (ii) all r sensors inbetween (if any) have failed, regardless of the states of all other sensors. The probability of this consolidated scenario to occur equals $(1 - q_{j_{is}})(1 - q_{j_{i(s+r+1)}})\Pi_{r'=1}^{r} q_{j_{i(s+r')}}$, and the expected error for the segment $j_{is} - j_{i(s+r+1)}$ equals $e_{ij_{is}j_{i(s+r+1)}}$ times this probability. For simplicity of notation, we just associate all these errors between j_{is} and $j_{i(s+r+1)}, \forall r$ to the sensor at j_{is} only. As such, the total expected error associated with the sensor at j_{is} across all scenarios is $\sum_{r=0}^{S_i - s} e_{ij_{is}j_{i(s+r+1)}}(1 - q_{j_{is}})(1 - q_{j_{i(s+r+1)}})\Pi_{r'=1}^{r} q_{j_{i(s+r')}}$; i.e., the sum of errors when the sensor at j pairs up with all its downstream sensors.

The total estimation error for the entire network can be written as a polynomial expression

$$\sum_{i \in \mathcal{I}} \sum_{s=0}^{s=S_i} \sum_{r=0}^{S_i - s} e_{ij_{is}j_{i(s+r+1)}}(1 - q_{j_{is}})(1 - q_{j_{i(s+r+1)}})\Pi_{r'=1}^{r} q_{j_{i(s+r')}}, \tag{5.7}$$

and the sensor location model for RNSPC can be formulated as follows,

$$\textbf{(RNSPC)} \quad \min_{\mathbf{X}} \Phi(\mathbf{X}) := \min_{\mathbf{Y},\mathbf{P}} \sum_{i \in \mathcal{I}} \sum_{j \in \mathcal{J}_{id-}} \sum_{k \in \mathcal{J}_{ij+}} \sum_{r=0}^{R_{ijk}} P_{ijkr} Y_{ijkr} e_{ijk} \tag{5.8a}$$

subject to

$$\sum_{j \in \bar{\mathcal{J}}} X_j \leq N \tag{5.8b}$$

$$X_u = X_d = 1 \tag{5.8c}$$

$$\sum_{k \in \mathcal{J}_{ijd} | R_{ijk} \geq r} Y_{ijkr} + \sum_{r'=0}^{r} Y_{ijdr'} = X_j, \forall i \in \mathcal{I}, j \in \mathcal{J}_{id-}, r = 0, \cdots, R_{ij} \tag{5.8d}$$

$$\sum_{j \in \mathcal{J}_{ik-} | R_{ijk} \geq r} Y_{ijkr} \leq X_k, \forall i \in \mathcal{I}, k \in \mathcal{J}_{iu+}, r = 0, \cdots, R_{iuk} \tag{5.8e}$$

$$\sum_{r=0}^{R_{ijk}} Y_{ijkr} \leq X_k, \forall i \in \mathcal{I}, j \in \mathcal{J}_i, k \in \mathcal{J}_{ijd} \tag{5.8f}$$

$$P_{ijk0} = (1-q_j)(1-q_k), \forall i \in \mathcal{I}, j \in \mathcal{J}_{id-}, k \in \mathcal{J}_{ij+} \tag{5.8g}$$

$$\begin{aligned} P_{ijkr} =& (1-q_k) \sum_{l \in \mathcal{J}_{ij+} | R_{ijl} \geq r-1} \frac{q_l}{1-q_l} P_{ijl(r-1)} Y_{ijl(r-1)}, \\ & \forall i \in \mathcal{I}, j \in \mathcal{J}_{-id}, k \in \mathcal{J}_{ij+}, r = 0, \cdots, R_{ijk} \end{aligned} \tag{5.8h}$$

$$X_j \in \{0,1\}, \forall j \in \mathcal{J}' \tag{5.8i}$$

$$Y_{ijkr} \in \{0,1\}, \forall i \in \mathcal{I}, j \in \mathcal{J}_{id-}, k \in \mathcal{J}_{ij+}, r = 0, \cdots, R_{ijk} \tag{5.8j}$$

$$0 \leq P_{ijkr} \leq 1, \forall i \in \mathcal{I}, j \in \mathcal{J}_{id-}, k \in \mathcal{J}_{ij+}, r = 0, \cdots, R_{ijk}. \tag{5.8k}$$

Constraint (5.8b) enforces the budget. Constraint (5.8c) postulates that the imaginary sensors are pre-installed. Constraints (5.8d) make sure that a sensor has to pair up with one and only one downstream sensor at each level until the imaginary sensor d is used. Constraints (5.8e) and (5.8f) respectively exclude the possibilities that (i) more than one upstream sensors pair up with this sensor at the same level and (ii) more than one levels are assigned to the same downstream sensor. Constraints (5.8g) and (5.8h) formulate the conditional probabilities for two sensors to pair up at different levels. Constraints (5.8i)-(5.8k) postulate binary and continuous decision variables. Note that constraints (5.8e) are redundant given (5.8d) and (5.8f), but we still keep them in the formulation because they are useful to some of the solution techniques in the next section. The following proposition reveals the relationship between the above formulation and the valid pairing-up rule.

Proposition 7. *At least one optimal solution to problem NSPC* (5.8a)-(5.8k) *satisfies the valid pairing-up rule. Furthermore, if* $q_j > 0, \forall j \in \bar{\mathcal{J}}$ *and* $e_{ijk} < e_{ijl}, \forall i \in \mathcal{I}, j \in \mathcal{J}_i, k \in \mathcal{J}_{ij+}, l \in \mathcal{J}_{ik+}$*, then this rule must be satisfied by all optimal solutions.*

Proof. Proof: For the simplicity of notation, in an optimal solution, locations with sensors installed on each path i are indexed with $\{0, 1, .., S_i + 1\}$ from upstream to downstream, where $u = 0$ and $d = S_i + 1$. Constraints (5.8d)-(5.8f) enforce that a sensor $0 \leq j \leq S_i$ pairs up with one and only one downstream sensor at each level until d is used. Let r_{ijd} denote the level for j to pair up d and j_r denote the downstream sensor paired up with j at a valid level $r \leq r_{ijd}$. Then the expected estimation error associated with the sensor at j is $e_{ij}^{r_{ijd}} := \sum_{r=0}^{r_{ijd}} (1-q_j)(1-q_{j_r}) \prod_{r'=0}^{r-1} q_{j_{r'}} e_{ijj_r}$. Note that the objective value (5.8a) equals $\sum_{i \in \mathcal{I}} \sum_{j \in \mathcal{J}_{id-}} e_{ij}^{r_{ijd}}$. We can prove that optimality indicates that there exists an optimal solution such that $e_{ijj_r} \leq e_{ijj_{r+1}}$ for any $0 \leq r < r_{ijd}$; otherwise if $e_{ijj_r} > e_{ijj_{r+1}}$ holds for any

optimal solutions, exchanging the levels of j_r and j_r+1 is supposed to result in a suboptimal solution. However, this exchange reduces $e_{ij}^{r_{ijd}}$ by

$$(1-q_j)\Bigg[e_{ijj_r}(1-q_{j_r})\prod_{r'=0}^{r-1}q_{j_{r'}}+e_{ijj_{r+1}}(1-q_{j_{r+1}})\prod_{r'=0}^{r}q_{j_{r'}} \\ -e_{ijj_{r+1}}(1-q_{j_{r+1}})\prod_{r'=0}^{r-1}q_{j_{r'}}-e_{ijj_r}(1-q_{j_r})q_{j_{r+1}}\prod_{r'=0}^{r-1}q_{j_{r'}}\Bigg] \\ =(1-q_j)(e_{ijj_r}-e_{ijj_{r+1}})(1-q_{j_r})(1-q_{j_{r+1}})\prod_{r'=0}^{r-1}q_{j_{r'}}\geq 0. \quad (5.9)$$

This means that the new solution at least preserves optimality, which contradicts the suboptimality of this solution. Since $e_{ijj'}\leq e_{ijj''}$, for any $0\leq j<j'<j''\leq S_i+1$, in this optimal solution, we can let the paring-up level for j and j' be no greater than that for j and j''. This implies that $j_r=j+r+1,\forall r<r_{ijr}$.

If level $r_{ijd}+1$ is also a feasible level to pair up j and d, then

$$e_{ij}^{r_{ijd}}-e_{ij}^{r_{ijd}+1}=(1-q_j)(1-q_{j+r_{ijd}})\prod_{r'=0}^{r_{ijd}-1}q_{j_{j+r'+1}}(e_{ijd}-e_{ij(j+r_{ijd})})\geq 0. \quad (5.10)$$

Thus there exists an optimal solution such that $r_{ijd}=J_i-2-j$ or every sensor downstream of j is paired up with j, which is consistent with the valid pairing-up rule.

Note that if $q_j>0,\forall j\in\bar{\mathcal{J}}$ and $e_{ijk}<e_{ijl},\forall i\in\mathcal{I},j\in\mathcal{J}_i,k\in\mathcal{J}_{ij+},l\in\mathcal{J}_{ik+}$, the inequalities (5.9) and (5.10) become strict. Then the proposed rule must be satisfied for any optimal solution. This completes the proof. □

It shall be noted that the RNSPC model can be easily adapted to accommodate existing sensor installations: We simply enforce $X_j=1$ if a sensor is already installed at location j; the model still has the same structure and complexity.

5.3 Solution Algorithms

The nonlinear mixed-integer program RNSPC is NP-hard since the well-known maximum covering problem is an obvious special case. It is often difficult to find its exact optimal solution when the problem size is large. Instead, heuristics and neighborhood search algorithms are usually adopted to obtain near-optimal feasible solutions. In order to estimate the quality of these solutions, relaxation techniques can be used to estimate the optimality residual

gaps between near-optimal feasible solutions and their dual bounds. The section proposes a variety of ways to obtain near-optimal feasible solutions and dual bounds to RNSPC.

5.3.1 Greedy and Interchange Heuristics

Greedy heuristic is widely applied to many practical problems not only because of its simplicity but also due to its reasonable practical performance. The greedy algorithm for RNSPC simply selects sensor locations sequentially based on the best marginal decrease of objective (5.8a), until all N installation locations have been selected. The exact steps are as follows.

Step G0: Initialization. Let the set of selected location indices $\mathcal{Q} := \emptyset$ and the iteration index $n := 1$. Define $\mathbf{X}^G := \{X_j^G\}_{j\in\mathcal{J}}$ and set

$$X_j^G = \begin{cases} 0, & \text{if } j \in \bar{\mathcal{J}} \\ 1, & \text{otherwise.} \end{cases} \tag{5.11}$$

Step G1: Search for the n^{th} location in $\bar{\mathcal{J}}\backslash\mathcal{Q}$ that will bring the maximum marginal decrease of objective (5.8a); i.e., select

$$j^* = \arg\min_{j\in\bar{\mathcal{J}}\backslash\mathcal{Q}} \{\Phi(\mathbf{X}) : X_k = 1, \text{ iff } k \in \mathcal{Q}\bigcup\{j\}\}. \tag{5.12}$$

Let $X_{j^*}^G = 1$ and $\mathcal{Q} = \mathcal{Q}\bigcup\{j^*\}$.

Setp G2: If $n = N$, stop and return $\mathbf{X}^G$; otherwise, $n = n + 1$, and go to step G1.

For the classic maximum covering problem (Feige, 1998) and the reliable maximal covering problem (Li and Ouyang, 2010), it has been showed that the greedy solution is no smaller than $(1 - 1/e)$ of the true optimum. The following paragraph presents a similar bound analysis for the optimality ratio of the RNSPC problem.

We slightly abuse the notation to let $\Phi(\mathcal{Q}) = \Phi(\mathbf{X}|X_j = 1, \text{iff} j \in \mathcal{Q}\bigcup\{u, d\}), \forall\mathcal{Q} \subseteq \bar{\mathcal{J}}$. Let $B(\mathcal{Q}) := \Phi(\emptyset) - \Phi(\mathcal{Q})$ denote the surveillance benefit from sensor installations at $\mathcal{Q}$. Let $B_i(\mathcal{Q})$ represent the benefit on path i from sensor installations at $\mathcal{Q}$. Note that $B(\mathcal{Q}) = \sum_{i\in\mathcal{I}} B_i(\mathcal{Q})$. Suppose that $B_i(\mathcal{Q}), \forall\mathcal{Q} \subseteq \mathcal{Q}$ is bounded from below by $L_i(|\mathcal{Q}\bigcap\bar{\mathcal{J}}_i|)$ and from above by $U_i(|\mathcal{Q}\bigcap\bar{\mathcal{J}}_i|)$, i.e., $L_i(|\mathcal{Q}\bigcap\bar{\mathcal{J}}_i|) \leq B_i(\mathcal{Q}) \leq U_i(|\mathcal{Q}\bigcap\bar{\mathcal{J}}_i|)$. Let $\mathcal{Q}^*$ represent the set of optimal sensor installation locations. Let $\mathcal{Q}^{G,n}$ be the first n locations chosen by greedy solution for n installations. Then the following equation holds

$$B(\mathcal{Q}^{G,n}) - B(\mathcal{Q}^{G,n-1}) \geq \frac{\sum_{i\in\mathcal{I}}\left[B_i(\mathcal{Q}^*) - \left(U_i(N) - L_i(|\mathcal{Q}^{G,n-1}\bigcap\bar{\mathcal{J}}_i| + 1)\right) - B_i(\mathcal{Q}^{G,n-1})\right]}{N}. \tag{5.13}$$

This yields

$$B(\mathcal{Q}^{G,n}) - B(\mathcal{Q}^{G,n-1}) \geq \frac{B(\mathcal{Q}^*) - C_{n-1} - B(\mathcal{Q}^{G,n-1})}{N} \tag{5.14}$$

where $C_{n-1} = \sum_{i\in\mathcal{I}} \left(U_i(N) - L_i(|\mathcal{Q}^{G,n-1} \bigcap \bar{\mathcal{J}}_i| + 1)\right)$. This leads to

$$B(\mathcal{Q}^{G,n}) \geq \left[1 - \left(\frac{N-1}{N}\right)^n\right] B(\mathcal{Q}^*) - \frac{1}{N}\sum_{k=0}^{n-1}\left(\frac{N-1}{N}\right)^{n-k-1} C_k; \tag{5.15}$$

Any given feasible solution, e.g. $\mathbf{X}^G$, can be further improved by interchange heuristics. The exact steps are as follows.

Step I0: Initialization. Set the local search step size η to be a small positive integer (usually $\eta \leq 2$). Let $\mathbf{X}^I := \mathbf{X}^G$.

Step I1: Search for a feasible $\mathbf{X}'$ within η distance from $\mathbf{X}^I$ (in the solution space $\{0,1\}^N$) that minimizes the objective (5.8a) ; i.e.,

$$\mathbf{X}' = \arg\min_{\mathbf{X}}\{\Phi(\mathbf{X}) : (5.8b), (5.8c), (5.8i), \sum_{j\in\bar{\mathcal{J}}} |X_j - X_j^I| \leq \eta\}. \tag{5.16}$$

Setp I2: If $\mathbf{X}' = \mathbf{X}^I$, stop and return $\mathbf{X}^I$; otherwise, set $\mathbf{X}^I = \mathbf{X}'$, and go to step I1.

5.3.2 Linear Programming Based Algorithm

Although the greedy and interchange algorithms normally only require a short solution time and are simple to implement, they do not yield any information on solution quality. Thus we propose additional algorithms that not only yield near-optimal solutions but also provide optimality gaps.

In the RNSPC model, equations (5.8a) and (5.8h) are nonlinear due to the existence of $\{P_{ijkr}Y_{ijkr}\}$, each of which is the product of a continuous variable and a binary variable. We can linearize the formulation by the technique introduced in Sherali and Alameddine (1992). For each $i \in \mathcal{I}, j \in I_i, k \in I_{ij}, r = 0, 1, \cdots, R_{ijk}$, we replace each $P_{ijkr}Y_{ijkr}$ by a new variable W_{ijkr} and add the following set of new constraints to enforce $W_{ijkr} = P_{ijkr}Y_{ijkr}$:

$$W_{ijkr} \leq P_{ijkr} \tag{5.17a}$$

$$W_{ijkr} \leq Y_{ijkr} \tag{5.17b}$$

$$W_{ijkr} \geq 0 \tag{5.17c}$$

$$W_{ijkr} \geq P_{ijkr} + Y_{ijkr} - 1. \tag{5.17d}$$

The linearized formulation (LRNSPC) becomes the following:

$$\textbf{(LRNSPC)} \quad \min_{\mathbf{X}} \Phi^L(\mathbf{X}) := \min_{\mathbf{Y},\mathbf{P},\mathbf{W}} \sum_{i\in\mathcal{I}} \sum_{j\in\mathcal{J}_{id-}} \sum_{k\in\mathcal{J}_{ij+}} \sum_{r=0}^{R_{ijk}} W_{ijkr} e_{ijk} \tag{5.18a}$$

subject to

$$\begin{aligned} P_{ijkr} =& (1-q_k) \sum_{l\in\mathcal{J}_{ij+}|R_{ijl}\geq r-1} \frac{q_l}{1-q_l} W_{ijl(r-1)}, \\ & \forall i \in \mathcal{I}, j \in \mathcal{J}_{id-}, k \in \mathcal{J}_{ij+}, r = 0, \cdots, R_{ijk} \end{aligned} \tag{5.18b}$$

$$(5.8b) - (5.8g), (5.8i) - (5.8k), (5.17a) - (5.17d).$$

For small-size instances, commercial software such as CPLEX may be able to solve the linear mixed-integer program LRNSPC. But in general, such an approach demands an excessively long time even for moderate-size instances. Thus we also propose a faster approximation approach based on linear relaxation.

If the integer constraints (5.8i) and (5.8j) are relaxed and replaced by

$$\begin{aligned} &0 \leq X_j \leq 1, && \forall j \in \mathcal{J}; \\ &0 \leq Y_{ijkr} \leq 1, && \forall i \in \mathcal{I}, j \in \mathcal{J}_{id-}, k \in \mathcal{J}_{ij+}, r = 0, \cdots, R_{ijk}, \end{aligned} \tag{5.19}$$

then LRNSPC becomes a linear program and can be solved in polynomial time. The solution to the relaxed problem $\mathbf{X}^L$ provides a lower bound to RNSPC, which however may be far from the true optimum, and $\mathbf{X}^L$ may be infeasible (i.e., containing fractional variables). We adopt a simple heuristic method in Ageev and Sviridenko (1999) to round $\mathbf{X}^L$ into a feasible integer solution $\mathbf{X}^R$ as follows.

Step 0: $\mathbf{X}^R = \mathbf{X}^L$

Step 1: If $\mathbf{X}^R$ is an integer solution, stop and return $\mathbf{X}^R$. Otherwise, choose $j, k \in \mathcal{J}, j \neq k$ such that X_j^R and X_k^R are the two elements closest to 0.5 among all fractional elements of $\mathbf{X}^R$.

Step 2: Let $\mathbf{X}' = \{X_1^R, \cdots, X_j^R + \epsilon \cdots, X_k^R - \epsilon \cdots, X_{|\mathcal{J}|}^R\}$ where ϵ equals either $-\min\{X_j^R, 1 - X_k^R\}$ or $\min\{X_k^R, 1 - X_j^R\}$, whichever yields a smaller value of $\Phi^L(\mathbf{X}')$; Update $\mathbf{X}^R = \mathbf{X}'$ and go to Step 1.

In Step 2, it is tedious to evaluate the value of $\Phi^L(\mathbf{X}')$ due to the existence of many auxiliary variables and constraints. Actually, function $\Phi^L(\mathbf{X}')$ could be replaced by a much simpler function $F(\mathbf{X}')$ derived from (5.7) as follows,

$$F(\mathbf{X}') := \sum_{i\in\mathcal{I}} \sum_{j\in\mathcal{J}_{id-}} \sum_{k\in\mathcal{J}_{ij+}} e_{ijk}(1-q_i)X'_i(1-q_k)X'_k \prod_{l\in\mathcal{J}_{ijk}} [1-(1-q_l)X'_l] \tag{5.20}$$

Note that $F(\mathbf{X}') = \Phi^L(\mathbf{X}') = \Phi(\mathbf{X}')$ for all integer $\mathbf{X}'$. Further more, from our experience, $F(\mathbf{X}')$ is likely to be smaller than $F(\mathbf{X}^R)$ for both ϵ values, i.e., $-\min\{X^R_j, 1-X^R_k\}$ and $\min\{X^R_k, 1-X^R_j\}$. Thus function F is a reasonable heuristic function to guide the rounding direction.

The above steps may not always yield the true optimum. The solution may be potentially improved by meta heuristics or neighborhood search methods.

5.3.3 Lagrangian Relaxation (LR) Based Algorithm

The linear-relaxation based solutions, especially the lower bounds, may be far from optima. This section presents a Lagrangian relaxation approach that will always yield better lower bounds.

We relax constraints (5.8f) and (5.8e), and add them to the objective function (5.8a) with nonnegative Lagrangian multipliers $\lambda = \{\lambda_{ijk}\}$ and $\gamma = \{\gamma_{ikr}\}$, respectively. The relaxed problem becomes

$$\begin{aligned}\textbf{(RRNSPC)} \quad & \max_{\lambda,\gamma\geq 0} \Delta(\lambda,\gamma) := \min_{\mathbf{X},\mathbf{Y},\mathbf{P}} \Gamma(\lambda,\gamma,\mathbf{X},\mathbf{Y},\mathbf{P}) \\ := & \sum_{j\in\mathcal{J}\setminus d} \left(\sum_{i\in\mathcal{I}_j} \sum_{k\in\mathcal{J}_{ij+}} \sum_{r=0}^{R_{ijk}} (P_{ijkr}e_{ijk} + \lambda_{ijk} + \gamma_{ikr})Y_{ijkr} - X_j \sum_{i\in\mathcal{I}_j} \left(\sum_{k\in\mathcal{J}_{ij-}} \lambda_{ikj} + \sum_{r=0}^{R_{iuj}} \gamma_{ijr} \right) \right)\end{aligned} \tag{5.21}$$

subject to (5.8b)-(5.8d), (5.8i),(5.8j).

The optimal solution of RRNSPC provides a lower bound to the original RNSPC problem (5.8). However, it is not easy to calculate $\Delta(\lambda,\gamma)$ even for given λ and γ. Thus we propose an approximate algorithm that bounds RRNSPC from below.

This algorithm is inspired by ideas in Cui et al. (2009). Let $j_1, j_2, \cdots j_{|\mathcal{J}_{ijk}|}$ be an ordering of the candidate locations in $\mathcal{J}_{ijk}$ such that $q_{j_0} \leq q_{j_1} \leq \cdots \leq q_{j_{|\mathcal{J}_{ijk}|}}$. Then let $p_{ijkr} = (1-q_j)(1-q_k)\prod_{l=0}^{r-1} q_{j_l}$. Note that for any feasible solution of the original problem, the probability variables $\mathbf{P}$ satisfy $P_{ijkr} \geq p_{ijkr}, \forall i \in \mathcal{I}, j \in \mathcal{J}_i, k \in \mathcal{J}_{ij+}, r \in 0, 1, \cdots, R_{ijk}$. We replace the probability variables $\mathbf{P}$ with fixed values $\mathbf{p} := \{p_{ijkr}\}$, and RRNSPC can be

approximated by the following

$$\textbf{(ARRNSPC)} \quad \max_{\lambda,\gamma\geq 0} \Delta^A(\lambda,\gamma) := \min_{\mathbf{X},\mathbf{Y}} \Gamma(\lambda,\gamma,\mathbf{X},\mathbf{Y},\mathbf{p}) \tag{5.22}$$

subject to (5.8b)-(5.8d), (5.8i),(5.8j).

The following proposition states the bounding relationship between ARRNSPC and RNSPC.

Proposition 8. *The solution to ARRNSPC* (5.22) *yields a lower bound of RNSPC objective* (5.8a).

Proof. Proof: Let $\mathbf{X}^*$, $\mathbf{Y}^*$ and $\mathbf{P}^*$ be the the optimal solution to (5.8). We construct a new model from (5.8) by replacing $\mathbf{P}$ with $\mathbf{p}$ and removing constraints (5.8g), (5.8h) and (5.8k). Due to the relaxation of these constraints, $\mathbf{X}^*$ and $\mathbf{Y}^*$ shall be also feasible to this new model. Furthermore, $p_{ijkr}Y^*_{ijkr}$ shall be no greater than $P^*_{ijkr}Y^*_{ijkr}, \forall i,j,k,r$ since p_{ijkr} is a lower bound of any non-trivial (i.e., when $Y_{ijkr}=1$) P_{ijkr} that satisfies the valid pairing-up rule. This implies that the optimal objective for the new model is a lower bound of (5.8a). Note that ARRNSPC (5.22) is actually the Lagrangian relaxed problem of the new model and thus yields a lower bound of it. Hence, the optimal objective for ARRNSPC (5.22) bounds (5.8a) from below. This completes the proof. □

Note that if q_j values for all $j \in \bar{\mathcal{J}}$ are identical, ARRNSPC is the same as the RRNSPC. Hence, when the spatial heterogeneity of q_j is not too dramatic, ARRNSPC should be a good approximation.

Given feasible λ and γ values, $\Delta^A(\lambda,,\gamma)$ can be simplified as follows.

$$\Delta^A(\lambda,\gamma) = \min_{\mathbf{X}} \sum_{j\in\mathcal{J}\setminus\{d\}} X_j \Delta^A_j(\lambda,\gamma) \tag{5.23}$$

subject to (5.8b) and (5.8c), where

$$\Delta^A_j(\lambda,\gamma) := \min_{\mathbf{Y}} \sum_{i\in\mathcal{I}_j} \left(\sum_{r=0}^{R_{ij}} \sum_{k\in\mathcal{J}_{ij+}|R_{ijk}\geq r} (p_{ijkr}e_{ijk} + \lambda_{ijk} + \gamma_{ikr}) Y_{ijkr} - \sum_{k\in\bar{\mathcal{J}}_{ij+}} \lambda_{ikj} - \sum_{r=0}^{R_{iuj}} \gamma_{ijr} \right)$$

subject to

$$\sum_{k\in\mathcal{J}_{ij+}\setminus\{d\}|R_{ijk}\geq r} Y_{ijkr} + \sum_{r'=0}^{r} Y_{ijdr'} = 1, \forall i \in \mathcal{I}_j, r = 0,\cdots,R_{ij}$$

Function $\Delta_j^A(\lambda,\gamma)$ can be simplified as $\sum_{i\in\mathcal{I}_j}\min_{0\le r\le R_{ij}}\Delta_{ijr}^A$ where

$$\Delta_{ijr}^A=\sum_{r'=0}^{r-1}\min_{k\in\mathcal{J}_{ijd}|R_{ijk}\ge r'}(p_{ijks}e_{ijk}+\lambda_{ijk}+\gamma_{iks})+(p_{ijdr}e_{ijd}+\lambda_{ijd}+\gamma_{idr})-\sum_{k\in\bar{\mathcal{J}}_{ij+}}\lambda_{ikj}-\sum_{r=0}^{R_{iuj}}\gamma_{ijr}$$

Given λ and γ, $\Delta^A(\lambda,\gamma)$ can be easily solved: Select up to N smallest negative $\Delta_j^A(\lambda,\gamma)$'s with $j\in\bar{\mathcal{J}}$ and set the corresponding X_j's to be 1; then Y_{ijkr} is set to be one if and only if $X_j=1$ and $M(p_{ijkr},\lambda_{ijk})\le M(p_{ijk'r},\lambda_{ijk'}),\forall k'\in\mathcal{J}_{ij+},R_{ijk'}\ge r$.

Proposition 9. *Function* $\Delta^A(\lambda,\gamma)$ (5.23) *is concave.*

Proof. Proof: Since component $\Delta_j^A(\lambda,\gamma)$ is linear except for minimization operations, it is a concave function. Let $(\lambda^1,\gamma)\ge 0$, $(\lambda^2,\gamma^2)\ge 0$ and $(\lambda^3,\gamma^3)=\alpha(\lambda^1,\gamma^1)+(1-\alpha)(\lambda^2,\gamma^2)\ge 0$ where scalar $0\le\alpha\le 1$. Let $\mathbf{X}^1$, $\mathbf{X}^2$ and $\mathbf{X}^3$ be the optimal minimizers for $\Delta^A(\lambda^1,\gamma^1)$, $\Delta^A(\lambda^2,\gamma^2)$ and $\Delta^A(\lambda^3,\gamma^3)$, respectively. Then

$$\begin{aligned}\alpha\Delta^A(\lambda^1,\gamma^1)+(1-\alpha)\Delta^A(\lambda^2,\gamma^2)&=\alpha\sum_{j\in\mathcal{J}}X_j^1\Delta_j^A(\lambda^1,\gamma^1)+(1-\alpha)\sum_{j\in\mathcal{J}}X_j^2\Delta_j^A(\lambda^2,\gamma^2)\\&\le\alpha\sum_{j\in\mathcal{J}}X_j^3\Delta_j^A(\lambda^1,\gamma^1)+(1-\alpha)\sum_{j\in\mathcal{J}}X_j^3\Delta_j^A(\lambda^2,\gamma^2)\\&=\sum_{j\in\mathcal{J}}X_j^3\left(\alpha\Delta_j^A(\lambda^1,\gamma^1)+(1-\alpha)\Delta_j^A(\lambda^2,\gamma^2)\right)\\&\le\sum_{j\in\mathcal{J}}X_j^3\Delta_j^A(\lambda^3,\gamma^3)=\Delta^A(\lambda^3,\gamma^3).\end{aligned}$$

Thus $\Delta^A(\lambda,\gamma)$ is a concave function. This completes the proof. □

The concavity of $\Delta^A(\lambda,\gamma)$ allows us to solve ARRNSPC with an iterative subgradient search. We update λ and γ iteratively to find the tightest upper bound $\min_{\lambda,\gamma}\Delta_A(\lambda,\gamma)$, while superscript m is added to distinguish variables in iteration m. The initial values λ^0 and γ^0 are set to zero or obtained from heuristics (e.g., the dual solution to LRNSPC). At the end of each iteration m, multipliers are updated as follows.

$$\lambda_{ijk}^{m+1}=\max(0,\lambda_{ijk}^m+t^m\delta_{ijk}^m),\forall i\in\mathcal{I},j\in\mathcal{J}_{id-},k\in\mathcal{J}_{ij+}.$$

$$\gamma_{ikr}^{m+1}=\max(0,\gamma_{ikr}^m+t^m\sigma_{ikr}^m),\forall i\in\mathcal{I},k\in\mathcal{J}_{iu+},r=0,\cdots,R_{iuk}.$$

where the subgradients $\delta_{ijk}^m:=\sum_{r=0}^{R_{ijk}}Y_{ijkr}^m-X_k^m$, $\sigma_{ikr}^m:=\sum_{j\in\mathcal{J}_{ik-}|R_{ijk}\ge r}Y_{ijkr}^m-X_k^m$ and

$\mathbf{X}^m, \mathbf{Y}^m$ are solutions to $\Delta^A(\lambda^m, \gamma^m)$. Step size t^m is usually set to

$$t^m = \frac{\mu^m(\Delta^A(\lambda^m, \gamma^m) - Z^{LB})}{\sum_{i\in\mathcal{I}}\sum_{j\in\mathcal{J}_{id-}}\sum_{k\in\mathcal{J}_{ij}}\left(\delta^m_{ijk}\right)^2 + \sum_{i\in\mathcal{I}}\sum_{k\in\mathcal{J}_{iu+}}\sum_{r=0}^{R_{iuk}}\left(\sigma^m_{ikr}\right)^2},$$

where μ^m is a control scalar, which generally decreases over iterations and can be updated in different ways (Fisher, 1981; Caprara et al., 1999). Z^{LB} is the objective value (5.8a) of the best (or smallest) feasible solution among all known ones. An feasible solution can be obtained from other algorithms or the following heuristic: Given a LR solution $\mathbf{X}$, determine $\mathbf{Y}$ and $\mathbf{P}$ based on the valid pairing-up rule.

If the LR algorithm ends up having a non-zero optimality gap, we embed the LR algorithm into a branch and bound (BB) framework to further reduce or close the gap. We branch on variables $\mathbf{X}$ to construct a binary tree where a greedy heuristic is used to expand children branches for each node: the next variable to branch is X_j if an installation at j brings in the greatest decrease of the objective value given the variables that have already been branched. We branch each variable first to 1 (enforcing installation) and then to 0 (forbidding installation). At each node, we run the LR algorithm to determine its feasible solution and lower bound, while extra constraints for already-branched variables are exerted. The multipliers of a node are passed down to next node as the initial multipliers. We record the best feasible solution from all the nodes traversed so far. If the lower bound for the current node is no smaller than the best feasible solution, the entire subtree rooted at this node no longer has potential and is trimmed. If the current node has already had N enforced or $|\mathcal{J}| - N$ forbidden installations, only one non-trivial feasible solution exists and is returned as both the lower and the upper bounds. After finishing both branches of a node, the lower bounds and upper bounds for the branches can be used to update those for this node. For moderate-size instances, the tree is traversed in a depth-first manner so as to rapidly trim branches and close the residual gap. For large-size instances where it is difficult to completely close the gap, we traverse the tree with a breadth-first search in the hope to obtain a smaller gap even without traversing the entire tree.

5.4 Alternative Formulations

5.4.1 A Continuous Approximation Approach for a Single Corridor

In many cases, practitioners are faced with the problem of deploying sensors on a freeway corridor rather than on a complex network (Bartin et al., 2007; Ban et al., 2009); i.e., the

path set $\mathcal{I}$ degrades to a singleton. The proposed RNSPC model and all solution algorithms are still applicable. However, with the problem scale increases, the computational efficiency of discrete models in general decreases significantly. An alternative approach with superior computational tractability is appealing for large scale instances. The single path structure allows us to adopt a continuum approximation (CA) solution approach that has attractive computational properties. The CA approach was originally proposed for the fixed-charge facility location problem in the supply chain design context (Newell, 1971, 1973; Daganzo, 1984a,b; Daganzo and Newell, 1986; Ouyang and Daganzo, 2006). See Langevin et al. (1996) and Daganzo (2005) for reviews. Recently, Cui et al. (2009) extended the CA method to address the reliable fixed-charge location problem and compared its performance with its discrete counterpart. So far, most existing CA models do not involve any explicit budget constraint. Now we will adapt the CA framework to solve the single corridor RNSPC problem that has an explicit budget constraint.

We consider a corridor between mileposts 0 and M. We first suppose that sensors can be installed anywhere on $[0, M]$ and the sensor installed at any $x \in [0, M]$ has a failure probability $q(x)$ that satisfies $q(0) = q(M) = 0$. We allow $q(x)$ to slowly vary along x. Define $A(x) : [0, M] \to \mathbb{R}_+$ to approximate the spacing between two neighboring sensors near x. Note that the inverse of $A(x)$ indicates the sensor density in the neighborhood of x. The estimation error of a segment of length a centered at $x \in [0, M]$ is now expressed as a function $e(x, a)$. We assume that $e(x, a), \forall x$ is a strongly super-linear (but sub-exponential) function increasing with a, and its structure only slowly varies with $x \in [0, M]$.

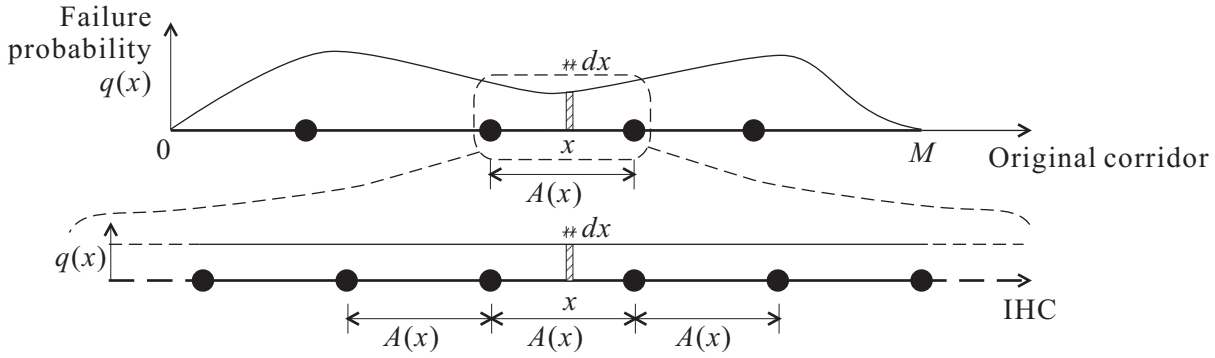


Figure 5.4: IHC for neighborhood x.

We convert sensor installation investment and disbenefits from surveillance errors along the corridor into unified cost units. The key to the CA approach is that the unit-length cost at each neighborhood of x is approximated by that of an infinite homogeneous corridor (IHC) with a similar parameter configuration; see Figure 5.4. On this IHC, sensors are distributed evenly with spacing $A(x)$, the failure probability is equal to $q(x)$ everywhere, and the error

measure function is identical to $e(x,a)$ everywhere. Figure 5.5 illustrates all (consolidated) scenarios on the IHC whenever x is covered by a level r neighboring sensor pair. We see that on the IHC there are $r+1$ exclusive and transitionally symmetric scenarios with two functioning neighboring sensors shifted from left to right. In each scenario, since the distance between two functioning neighboring sensors is always $(r+1)A(x)$, the error measure for the segment inbetween is $e(x,(r+1)A(x))$ and then the unit-length error near x is $\frac{e(x,(r+1)A(x))}{(r+1)A(x)}$. Since each scenario has two functioning sensors and r failed sensors, the probability for this scenario to occur shall be $q(x)^r(1-q(x))^2$. Thus the total expected unit-length error for x to be covered by a level r sensor pair is the summation across all the scenarios in Figure 5.5:

$$\sum_{s=1}^{r+1} q(x)^r(1-q(x))^2 \cdot \frac{e(x,(r+1)A(x))}{(r+1)A(x)} = \frac{1}{A(x)} q(x)^r(1-q(x))^2 e(x,(r+1)A(x)). \quad (5.24)$$

Then the total expected unit-length error for x to be covered by all levels of sensor pairs is

$$C(x,A(x)) := \frac{1}{A(x)} \sum_{r=0}^{\infty} q(x)^r(1-q(x))^2 e(x,(r+1)A(x)), \forall x \in [0,M], \quad (5.25)$$

which shall be a finite value since $e(x,a)$ is sub-exponential.

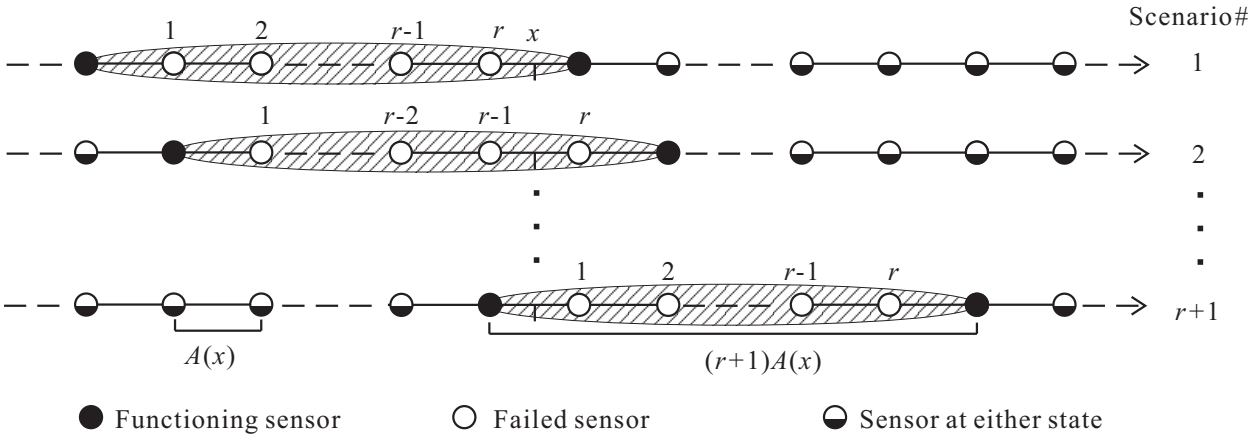


Figure 5.5: Scenarios for level r neighboring sensor coverage on the IHC for x.

We use formula (5.25) to approximate the actual unit-length error in the neighborhood of x of the original corridor. By integrating (5.25) for all neighborhoods $x \in [0,M]$, we can approximate the total expected error on the original corridor. Then the problem of determining the optimal sensor locations reduces to the minimization of the total expected

error, i.e.

$$\min_{A(x)>0} \int_{x\in[0,M]} C(x, A(x))dx, \tag{5.26a}$$

subject to the budget constraint

$$\int_{x\in[0,M]} A(x)^{-1}dx \leq N+1, \forall x \in [0, M], \tag{5.26b}$$

Solution Technique

Model (5.26) is a simple constraint nonlinear optimization problem which can be solved by Lagrangian relaxation. Relaxing constraint (5.26b) and adding it to the objective (5.25) with a scalar multiplier ω, (5.26) becomes

$$\max_{\omega\geq 0} \min_{A(x)>0} \int_{x\in[0,M]} \hat{C}(x, A(x), \omega)dx, \tag{5.27a}$$

where

$$\hat{C}(x, A(x), \omega) := -\frac{\omega(N+1)}{M} + \frac{\omega}{A(x)} + C(x, A(x)), \forall x \in [0, M], \tag{5.27b}$$

The relaxed model (5.27) has the same solution as the original model (5.26), as stated in the following proposition.

Proposition 10. *The optimal solutions to models* (5.26) *and* (5.27) *are always identical.*

Proof. Proof: Since (5.27) is a Lagrangian relaxation, it always bounds (5.26) from below. Since $e(x, a)$ is a strongly super-linear and increasing function over a, the optimal $A(x)$ for (5.27) is finite and increases with ω continuously from 0 to ∞. This implies $\int_{x\in[0,M]} A(x)^{-1}dx$ also increases with ω continuously from 0 to ∞. Therefore exists a feasible ω value in $(0, \infty)$ that corresponds an optimal solution with $\int_{x\in[0,M]} A(x)^{-1}dx = N+1$; i.e., the complementary condition holds. Hence, optimality gap is zero; i.e., the optimal solutions from (5.26) and (5.27) are identical. This completes the proof. □

Model (5.27) can be solved iteratively. Note that the case with $\omega = 0$ is trivial since it yields an obviously suboptimal objective of ∞. For $\omega > 0$, minimizing (5.27a) is equivalent

to minimizing its integrand at each x independently, i.e., $\min_{A(x)} \hat{C}(x, A(x), \omega), \forall x \in [0, M]$. Function $\hat{C}(x, A(x), \omega)$ is usually a unimodal function over $A(x)$; i.e., $\hat{C}(x, A(x), \omega)$ only has one stationary point over $A(x) \in (0, \infty)$ and that point is the optimal solution. This allows us to use a bisecting search to find the optimum. In some special cases (e.g., $\hat{C}(x, A(x), \omega)$ is an economic order quantity type function), we can even solve this optimum analytically. Then given ω, (5.27) is solved by numerically integrating the solutions of $\hat{C}(x, A(x), \omega)$ across all $x \in [0, M]$. By examining whether this solution violates constraint (5.26b) we can obtain the subgradient direction and improve ω accordingly. Starting with an arbitrary positive ω value [c], we repeatedly search for the optimal ω in a similar bisecting manner. For each x, this CA method requires only a squared logarithmic number of iterations.

The solution to (5.26) takes continuous input and yields continuous optimal sensor density at each neighborhood. In many real-world sensor location design problems, only discrete input is available and the expected output must be a discrete sensor location design that can be practically implemented. The interpolation based method proposed by Peng and Ouyang (2010) can be used to generate continuous input from discrete data. Suppose that the locations in $\mathcal{J}$ are $\{0, 1, \cdots |\mathcal{J}| - 1\}$ ordered from upstream to downstream and each location $j \in \mathcal{J}$ is at milepost M_j. Function $q(x)$, for example, can be obtained by interpolation based on $\{q_j\}$ (i.e., $q(x) = \frac{M_{j+1}-x}{M_{j+1}-M_j} q_j + \frac{x-M_j}{M_{j-1}-M_j} q_{j+1}, \forall M_j \leq x < M_{j+1}, j \in \mathcal{J}$). The specification of function $e(x, a)$ will be determined by the values of $\{e_{jk}\}$[d]. One possible method is to let $e(\frac{M_j+M_k}{2}, M_k - M_j) = e_{jk}, \forall 0 \leq j < k \leq |\mathcal{J}| - 1$, and then interpolate function $e(x, a), \forall x, a$. Once we solve the CA problem (5.26), the discretization method in Daganzo (2005) can be used to convert its continuous solution to a discrete sensor location design. If the candidate locations are a finite set of discrete points, sensor installation locations in the discrete solution are often rounded to their closest candidate locations.

Lower Bound Analysis

Under certain conditions, the solution to model (5.26) is a lower bound of that to the optimal discrete sensor deployment, and thus can help evaluate the residual gap of this discrete solution. When sensor failure probability is negligible, the relationship between the CA solution to (5.26) and the optimal discrete solution is discussed in the following proposition.

Proposition 11. *For the deterministic version (i.e., $q(x) = 0, \forall x \in [0, M]$) of the single corridor RNSPC problem where sensors can be installed anywhere along the corridor, if*

[c] Based on the fact that the optimal solution of (5.26) always activates constraint (5.26b), an initial ω can be roughly estimated from the magnitudes of $q(x)$ and $e(x, a)$ values.

[d] We omit subscript i since $\mathcal{I}$ is a singleton.

$e(x,a), \forall x \in [0,M]$ is concave over x for any given a, the optimal objective value of (5.26a) *is a lower bound of the optimal discrete solution.*

Proof. Proof: Suppose that in the optimal discrete solution, sensors are located at locations $x_1^* < x_2^* < \cdots < x_N^*$, and let $x_0^* = 0$ and $x_{N+1}^* = M$ be the imaginary sensor locations. Then the objective value of discrete solution is

$$\sum_{n=0}^{N} e\left(\frac{x_n^* + x_{n+1}^*}{2}, x_{n+1}^* - x_n^*\right). \tag{5.28}$$

We construct a feasible CA solution $A(x) = x_{n+1}^* - x_n^*, \forall x \in [x_n^* x_{n+1}^*), n = 0, \cdots, N$. Then,

$$(5.26a) \leq \int_0^M \frac{e(x, A(x))}{A(x)} dx = \sum_{n=0}^{N} \int_{x_n^*}^{x_{n+1}^*} \frac{e(x, x_{n+1}^* - x_n^*)}{x_{n+1}^* - x_n^*} dx \leq (5.28).$$

The first inequality holds because the CA optimal objective is no larger than that for the feasible solution $\{A(x)\}$. The second inequality comes from the concavity of $e(x,a)$ over x and the Jensen's inequality. This completes the proof. □

Proposition 11 can be easily adapted for problems where candidate sensor locations are a finite set of discrete points on the corridor, as stated below,

Corollary 1. *For the deterministic version (i.e., $q_j = 0, \forall j = 0, 1, \cdots, |\mathcal{J}| - 1$) of the single corridor RNSPC problem where sensors can only be installed at a finite number of candidate locations $0, 1, \cdots |\mathcal{J}| - 1$ (ordered from upstream to downstream), if $e(x,a)$ is constructed in a way such that $\int_{M_j}^{M_k} \frac{e(x, M_k - M_j)}{M_k - M_j} dx \geq e_{jk}, \forall 0 \leq j < k \leq |\mathcal{J}| - 1$, the optimal objective value of* (5.26a) *is a lower bound of the optimal discrete solution.*

Note that the total error under zero sensor failure probability shall be always no larger than that under positive probability (due to the loss of estimation accuracy from possible sensor failures). Then the CA solution for the deterministic case will also be a lower bound of the optimal discrete solution under non-zero sensor failure probabilities, as summarized below,

Corollary 2. *When the condition in Proposition 11 (or Corollary 1) holds, the optimal objective value of* (5.26a) *for a deterministic continuous problem (i.e., when $q(x) = 0, \forall x \in [0, M]$) is a lower bound of that for the corresponding discrete reliable problem (i.e. when $q_j \geq 0, \forall j = 0, 1, \cdots, |\mathcal{J}| - 1$).*

5.4.2 Fixed Charge Location Models

In some applications, rather than imposing an explicit budget constraint, the objective is to minimize the overall disbenifits from estimation errors and sensor infrastructure investment. These problems can be modeled by a slight variation of RNSPC where the budget constraint (5.8b) is removed and the facility construction cost is added to objective (5.8a). Assume installing a sensor at location $j \in \bar{\mathcal{J}}$ costs the same as f_j units of estimation error. Then we can formulate this Reliable Fixed Charge Neighboring Sensor Location model (RFCNSL) as follows.

$$\textbf{(RFCNSL)} \quad \min_{\mathbf{X},\mathbf{Y},\mathbf{P}} \sum_{i\in\mathcal{I}} \sum_{j\in\mathcal{J}_{id-}} \sum_{k\in\mathcal{J}_{ij+}} \sum_{r=0}^{R_{ijk}} P_{ijkr} Y_{ijkr} e_{ijk} + \sum_{j\in\mathcal{J}} f_j X_j \tag{5.29}$$

subject to (5.8c)-(5.8k).

The structure and complexity of the model are largely unchanged and all the solution techniques proposed in Section 3 can be easily adapted for this RFCNSL problem.

For the single corridor RFCNSL problem, we can also apply the CA approach in a similar manner. We use all the continuous settings in Section 5.4.1, and let $f(x)$ denote a slowly-varying installation cost function at location $x \in [0, M]$. The CA version of the single corridor RFCNSL model can be written as follows.

$$\min_{A(x)>0} \int_{x\in[0,M]} \left[C(x, A(x)) + \frac{f(x)}{A(x)} \right] dx, \tag{5.30}$$

where $\frac{f(x)}{A(x)}$ represents the facility investment cost per unit distance and the integrand is the total cost per unit length in the neighborhood of x. We can obtain the optimal solution to (5.30) by independently minimizing the integrand at each x with a similar bisecting method. Similar lower bound analysis can be conducted to show the relationship between the CA solution and the discrete solution.

5.5 Case Studies

This section presents several numerical examples of the RNSPC model. All solution algorithms are implemented on a PC with 2.0 GHz CPU and 2 GB memory, and we set the solution time limit to be 1800 seconds. In the the Sioux-Falls network example, we will test all the proposed algorithms under a variety of effectiveness measures and parameter settings in order to draw insights on how these settings affect the optimal objective value and sensor deployment. Then we will solve a Chicago intermodal network example. We also test the

proposed CA approach on a hypothetical highway corridor and compare its performance with those of the discrete algorithms.

The Sioux-Falls network is shown in Figure 4.2. Again, $|\mathcal{J}| = 24$, $|\mathcal{I}| = 528$. Assume too that the sensor at a vertex can detect all passing traffic from all directions. Based on passing traffic volumes, we group the 24 vertices into three sets (as marked with different colors), $\mathcal{J}^h = \{8, 10, 11, 15, 16, 17, 19, 22\}$ with the heaviest traffic, $\mathcal{J}^l = \{1, 2, 3, 7, 9, 13, 20, 23\}$ with the lightest traffic and $\mathcal{J}^m = \{4, 5, 6, 12, 14, 17, 19, 24\}$ with medium traffic. Assume that sensors installed at locations with heavier traffic are subject to higher failure probabilities. We define sensor failure probabilities as follows

$$q_j = \begin{cases} \bar{q} - \hat{q} & \text{if } j \in \mathcal{J}^l \\ \bar{q} & \text{if } j \in \mathcal{J}^m \\ \bar{q} + \hat{q} & \text{if } j \in \mathcal{J}^h \end{cases} \tag{5.31}$$

where scalar $\bar{q}$ is the average probability and scalar $\hat{q}$ indicates spatial variation.

For the FV measure (5.2), we set $b_i^c = a_{iud}, \forall i \in \mathcal{I}$; for the VM measure (5.3), we set $b_i^t = 1, \forall i \in \mathcal{I}$. Since no relevant empirical data are available to specify the exact SER measure, we assume that it follows a simple convex form

$$e_{iuj} = 2(a_{iuj})^\beta, e_{ijd} = 2(a_{ijd})^\beta, e_{iud} = 4(a_{ijd})^\beta, \text{ and } e_{ijk} = (a_{ijk})^\beta, \forall i \in \mathcal{I}, j = \bar{\mathcal{J}}_i, k \in \mathcal{J}_{ijd} \tag{5.32}$$

where scalar $\beta > 1$.

Table 5.2 compares the results from different algorithms under the three measures when $\bar{q} = 0.15, \hat{q} = 0, N = 10$; $\beta = 2$. The results include solution objective values, residual gaps (i.e., the percentage difference between the feasible solution and the estimated lower bound[e]), the true optimality gap (i.e., the percentage difference from the true optimum) and solution times. We see that only the LR algorithm can solve the instances for all measures to optimality. The greedy and interchange (with $\eta = 2$) algorithms can obtain solutions very fast. Although these two heuristics cannot provide lower bounds by themselves, their solutions (especially those from the interchange algorithm) are actually already very close to the true optima. This suggests that these two algorithms can be efficient tools for engineering practice. In general, CPLEX can neither provide good estimates of the lower bounds nor yield good near-optimal solutions. Compared with CPLEX, the linear programming algorithm can yield comparable feasible solutions in shorter times, but the lower bounds from the linear relaxation are far from optima. Overall, the solution quality is consistent across different

[e]We do not use the estimated lower bound as the denominator since it may be negative in some extreme cases.

measures. Thus we will only focus on the SER measure in the following analysis.

Table 5.2: Result for different error measures.

Algorithm	Measure	Objective	Residual gap	True optimality gap	Solution time (sec)
LR based	FV	168881	0 %	0.0 %	192
	VM	1650870	0 %	0.0 %	74
	SER	34521100	0 %	0.0 %	60
Greedy	FV	178041	-	5.1 %	1.73
	VM	1650870	-	0.0 %	0.04
	SER	34985300	-	1.3 %	0.05
Interchange	FV	171537	-	1.5 %	4
	VM	1650870	-	0.0 %	4
	SER	34581700	-	0.2 %	5
CPLEX	FV	225888	167 %	25.2 %	1800
	VM	1650870	27 %	0 %	1800
	SER	37603900	35 %	8.2 %	1800
LP based	FV	208295	1332 %	18.9 %	71
	VM	1731920	78 %	4.7 %	60
	SER	37006100	86 %	6.7 %	63

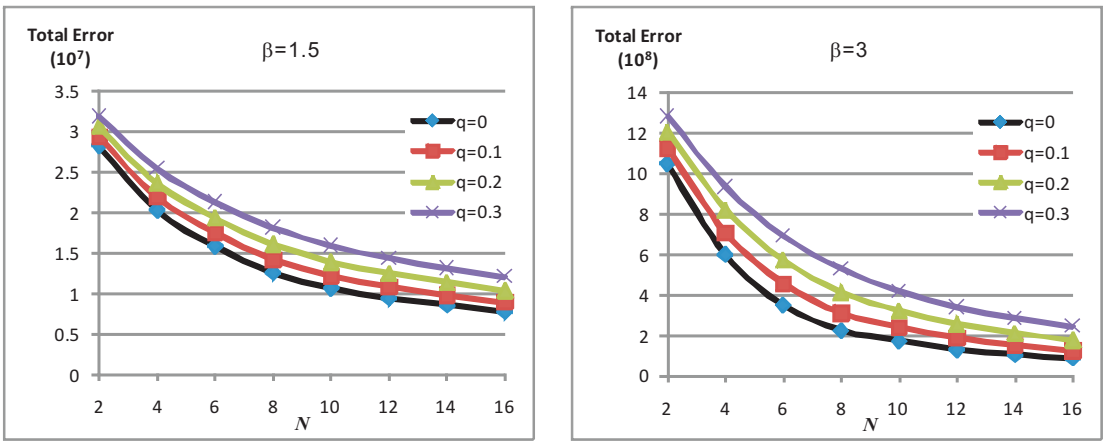
Table 5.3 shows the solutions with different parameter settings under the SER measure. We see that the algorithm performances are consistent with those in Table 5.2. The LR based algorithm solves the instances with $\hat{q} = 0$ (i.e., spatially homogeneous probabilities) more efficiently that those with $\hat{q} > 0$ (i.e., spatially heterogeneous probabilities), which is probably due to the approximation gap in the ARRNSPC model. The objective value significantly increases as $\bar{q}$ gets higher while it is somehow less sensitive to the value of $\hat{q}$. A larger installation number N yields a smaller objective value, which is intuitive since more sensors will generally bring in more benefit.

Figures 5.6 and 5.7 show the impacts of N and $\bar{q}$ on the SER measure under different β values. We see that the total error decreases with the installation number while the decreasing trend flattens out. The total error is more sensitive to the value of N for a larger β; i.e., the SER measure with larger convexity tends to have more improvement potential from additional sensor installations. While a larger β implies a higher sensitivity as well, the total error increases almost linearly with $\bar{q} \in [0, 0.4]$.

Figure 5.8 shows the optimal sensor deployment for different sensor failure probabilitie. The sensor installation locations are marked by circles. By comparing Figures 5.8(a) and 5.8(b), we see that sensors tend to cluster and back up each other when they are subject to failure. By comparing Figure 5.8(b) and Figure 5.8(c), we see that under a higher $\hat{q}$ the

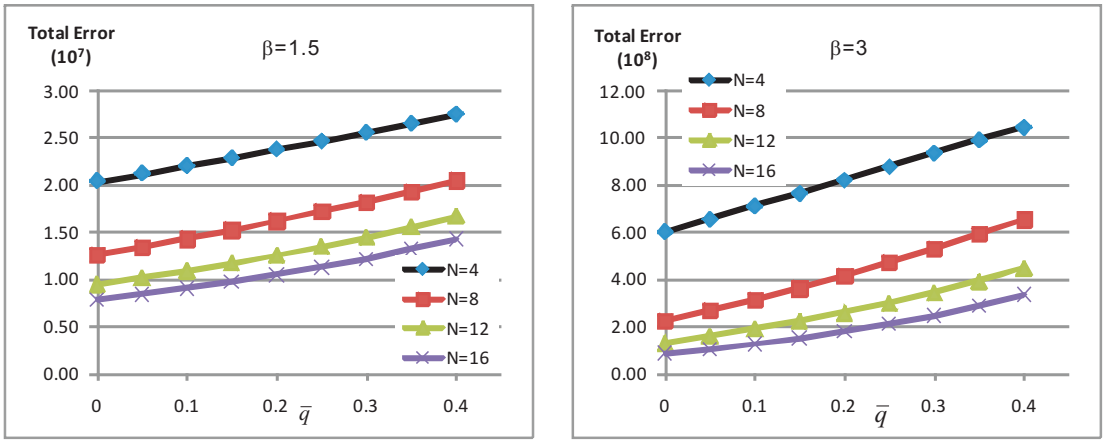
Table 5.3: Algorithm comparison (under the SER measure with $\beta = 2$).

Algorithm	$\bar{q}$	$\hat{q}$	N	Objective	Residual gap	True optimality gap	Solution time (sec)
LR based	0	0	6	4.27E+07	0 %	0.0 %	31
	0.15	0	6	5.37E+07	0 %	0.0 %	46
	0.15	0.05	6	5.60E+07	0 %	0.0 %	1040
	0	0	8	3.15E+07	0 %	0.0 %	8
	0.15	0	8	4.16E+07	0 %	0.0 %	41
	0.15	0.05	8	4.30E+07	0 %	0.0 %	1585
Greedy	0	0	6	4.37E+07	-	2.3 %	0.03
	0.15	0	6	5.40E+07	-	0.5 %	0.03
	0.15	0.05	6	5.64E+07	-	0.6 %	0.03
	0	0	8	3.25E+07	-	3.1 %	0.04
	0.15	0	8	4.23E+07	-	1.5 %	0.04
	0.15	0.05	8	4.41E+07	-	2.5 %	0.04
Interchange	0	0	6	4.27E+07	-	0.0 %	2
	0.15	0	6	5.37E+07	-	0.0 %	2
	0.15	0.05	6	5.60E+07	-	0.0 %	2
	0	0	8	3.15E+07	-	0.0 %	3
	0.15	0	8	4.16E+07	-	0.0 %	3
	0.15	0.05	8	4.30E+07	-	0.0 %	3
CPLEX	0	0	6	4.27E+07	0 %	0.0 %	39
	0.15	0	6	5.51E+07	23 %	2.4 %	1810
	0.15	0.05	6	6.02E+07	35 %	7.0 %	1801
	0	0	8	3.15E+07	0 %	0.0 %	34
	0.15	0	8	4.55E+07	37 %	8.6 %	1801
	0.15	0.05	8	4.37E+07	33 %	1.7 %	1801
LP based	0	0	6	4.58E+07	7 %	6.7 %	50
	0.15	0	6	5.56E+07	73 %	3.4 %	53
	0.15	0.05	6	5.81E+07	76 %	3.5 %	54
	0	0	8	3.15E+07	0 %	0.0 %	49
	0.15	0	8	4.21E+07	81 %	1.1 %	52
	0.15	0.05	8	4.42E+07	83 %	2.7 %	44



(a) (b)

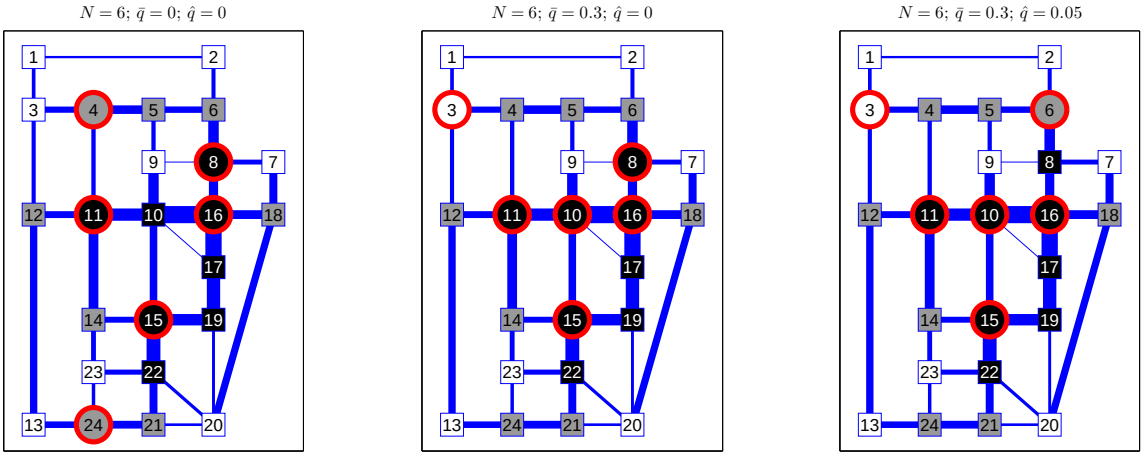
Figure 5.6: Relationship between the total error and N ($\hat{q} = 0$, under the SER measure).



(a) (b)

Figure 5.7: Relationship between the total error and $\bar{q}$ ($\hat{q} = 0$, under the SER measure).

sensor at vertex 8 has been relocated to vertex 6 where the failure probability is smaller. This implies that optimal sensor deployment seeks more reliable sensor installation locations when failure probabilities vary across space.



(a) (b) (c)

Figure 5.8: Optimal sensor deployment for $N = 6$ installations under the SER measure with $\beta = 2$: (a) $\bar{q} = \hat{q} = 0$; (b) $\bar{q} = 0.3, \hat{q} = 0$; (c) $\bar{q} = 0.3, \hat{q} = 0.05$.

5.5.1 Chicago Intermodal Network

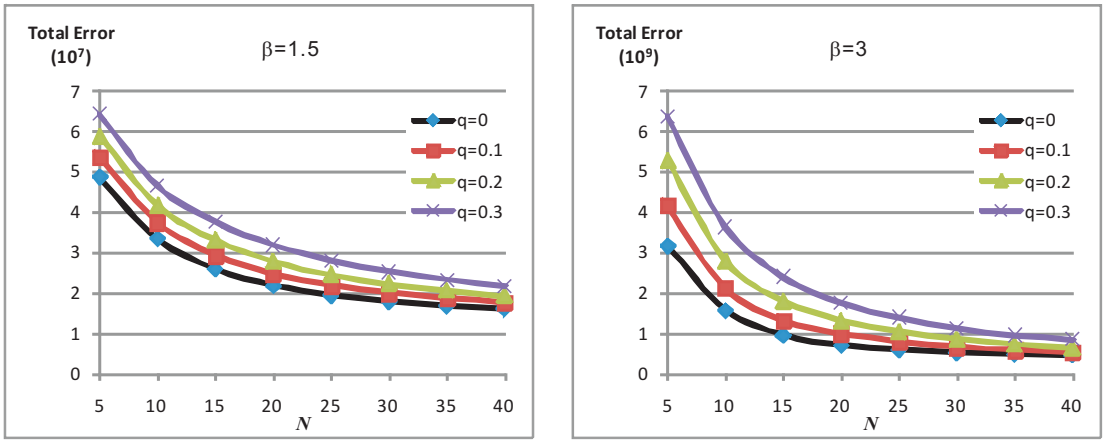
Figure 4.5 shows the geometry of the Chicago interstate highway network, which contains 21 highway junctions and 17 railroad terminals (i.e., the railroad yards for intermodal freights). Highway traffic comes in and goes out of the network through 8 access points. Since most sensor technologies have a limited effectiveness range, a sensor installation at a highway junction may not be able to inspect passing traffic from all directions. Thus, each highway junction is split into multiple candidate locations (Sheffi, 1985) such that an installation at any candidate location can inspect all passing flows. The final network representation includes 89 candidate locations and 363 (directed) connecting links. The 2002 intermodal freight traffic[f] originated from or destined to Chicago is grouped into 1046 O-D paths on this network based on population distribution. Due to lack of detailed information, we again assume that all O-D flows follow their shortest distance paths.

We test different algorithms with the SER measure (5.32). Let $\{q_j\}$ follow (5.31) where sets $\mathcal{J}^l$, $\mathcal{J}^m$ and $\mathcal{J}^h$ are specified similarly based on the passing traffic volumes; we have $|\mathcal{J}^l| = |\mathcal{J}^m| = 30$ and $|\mathcal{J}^h| = 29$. Due to the increased problem size, CPLEX ran out of

[f]Data source: Bureau of Transportation Statistics, http://www.bts.gov/.

memory even for the linear relaxed problem (LRNSPC). The LR based algorithm always yields a near-optimal solution with a reasonable residual gap ($\leq 15\%$) if $\hat{q}$ is not too large (e.g., $\leq 25\%\bar{q}$). The solutions from the greedy and interchange algorithms, though not as good, are close to those from the LR based algorithm. From our experiments, the differences between these near-optimal solutions and the true optima are often much smaller than the residual gaps. Thus these solutions are suitable for engineering practice.

Figures 5.9 and 5.10 show again that the objective value is more sensitive to changes of N and $\bar{q}$ under the SER measure when β is larger. The marginal change of the objective over N gradually diminishes while that over q remains almost the same $\forall \bar{q} \in [0, 0.4]$. Figure 5.11 illustrates how $\bar{q}$ and $\hat{q}$ affect the optimal sensor deployment. Again, a higher $\bar{q}$ generally leads to higher sensor concentration (to back up locations with heavier traffic) while a higher $\hat{q}$ forces sensors to seek more reliable substitute locations (as highlighted in Figures 5.11(b) and 5.11(c)).

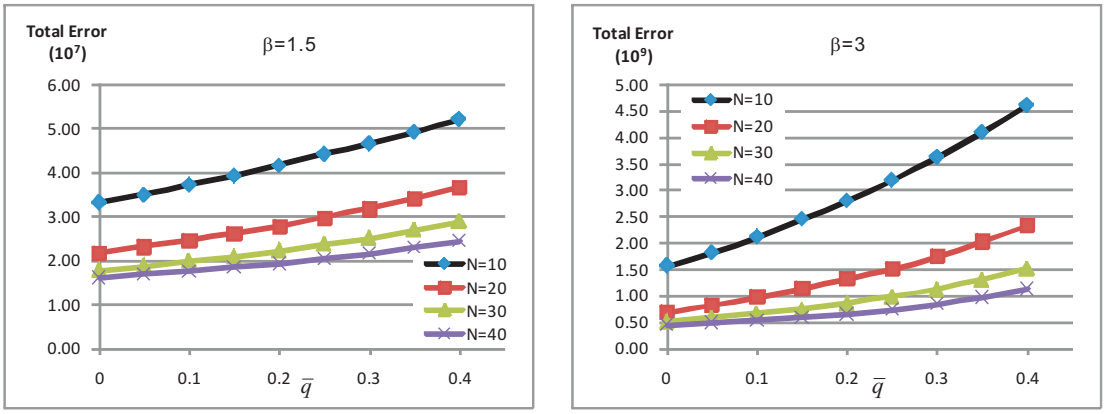


(a) (b)

Figure 5.9: Relationship between the total error and N ($\hat{q} = 0$, under the SER measure).

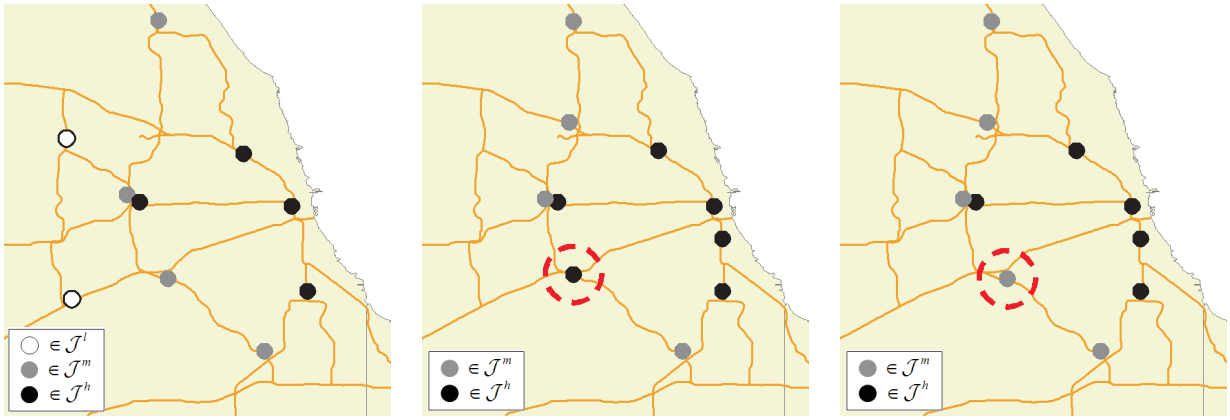
5.5.2 Single Corridor

Consider a hypothetical single corridor $[0, M]$ where $M = 27$. Since the path set $\mathcal{I}$ is a singleton, we omit subscript i in the notation. Let candidate locations be $\mathcal{J} = \{0, 1, \cdots, M\}$ with $u = 0, d = M$ and they are evenly distributed across the corridor; i.e., location j's mileage equals j. Define the a SER measure as $e_{jk} = c(\frac{k+j}{2})(k-j)^2, \forall 0 \leq j < k \leq M$ where function $c(x)$ is either a constant (e.g., $c(x) = 1, \forall x$) or slowly varying over $x \in [0, M]$ (e.g., $c(x) = 0.5 + h(x), \forall x$ where $h(x) = 1 - \frac{|x-M/2|}{M/2}$). Note that if the error measure $\{e_{jk}\}$ is weighted by the traffic volume, the variation of $c(x)$ can capture the traffic volume change



(a) (b)

Figure 5.10: Relationship between the total error and q ($\hat{q} = 0$, under the SER measure).



(a) (b) (c)

Figure 5.11: Optimal sensor deployment for $N = 10$ installations under the SER measure with $\beta = 2$: (a)$\bar{q} = \hat{q} = 0$; (b)$\bar{q} = 0.3, \hat{q} = 0$; (c)$\bar{q} = 0.3, \hat{q} = 0.05$.

along the corridor. Similarly, let $q_j, \forall j \in \bar{\mathcal{J}}$ be either a constant (e.g., 0.2) or a spatially varying value (e.g., $q_j = 0.4h(j)$). Let the installation budget be $N = 8$. In the CA model, define $e(x, a) = c(x)a^2$ and $q(x)$ as a piecewise linear function by interpolating $q_j, \forall j \in \mathcal{J}$, i.e., $q(x) = (\lfloor x + 1 \rfloor - x)q_{\lfloor x \rfloor} + (x - \lfloor x \rfloor)q_{\lfloor x+1 \rfloor}$.

Table 5.4 shows the test results for both the LR based algorithm and the CA approach for different failure probabilities and error measures. The LR based algorithm can solve instances # 1 and # 3 (where all q_j values are identical) to the exact optima. For instances # 2 and # 4 where the spatial heterogeneity of failure probabilities is significant, the LR based algorithm can not estimate lower bounds as effectively (the solutions end up with residual gaps around 30%). For all these instances, the CA appoach always very quickly yields approximate objective values that are very close to those from the LR based algorithm. These continuous solutions from the CA approach are discretized into sensor installation locations among $\bar{\mathcal{J}}$. Interestingly, these discrete solutions are almost identical to those from the LR based algorithm even under the spatial heterogeneity from sensor failure probabilities (instance # 2), error measures (instance # 3) or both of them (instance # 4). This suggests that the CA approach is able to not only efficiently estimate the optimal objective value but also yield very good discrete location design.

Table 5.4: Result summary.

Algorithm	#	$c(x)$	$q(x)$	Objective	Difference from the LR solution	Solution time (sec)
LR based	1	1	0.2	115.875	-	945
	2	1	$0.4h(x)$	123.456	-	1800
	3	$0.5 + h(x)$	0.2	114.929	-	675
	4	$0.5 + h(x)$	$0.4h(x)$	127.372	-	1800
CA estimate (5.26a)	1	1	0.2	119.787	3 %	≈ 0
	2	1	$0.4h(x)$	123.947	≈ 0 %	≈ 0
	3	$0.5 + h(x)$	0.2	117.623	2 %	≈ 0
	4	$0.5 + h(x)$	$0.4h(x)$	125.625	-1 %	≈ 0
Discrete CA Solution	1	1	0.2	115.875	≈ 0 %	≈ 0
	2	1	$0.4h(x)$	123.549	≈ 0 %	≈ 0
	3	$0.5 + h(x)$	0.2	115.261	≈ 0 %	≈ 0
	4	$0.5 + h(x)$	$0.4h(x)$	127.974	≈ 0 %	≈ 0

Figure 5.12 compares the discrete deployments from both the LR based algorithm and the CA approach. We see that in instance # 1 (with almost homogeneous space), the deployments from the two methods are exactly the same. In other three instances, their deployments are also consistent: Sensors are more concentrated in areas with larger $c(x)$ and $q(x)$.

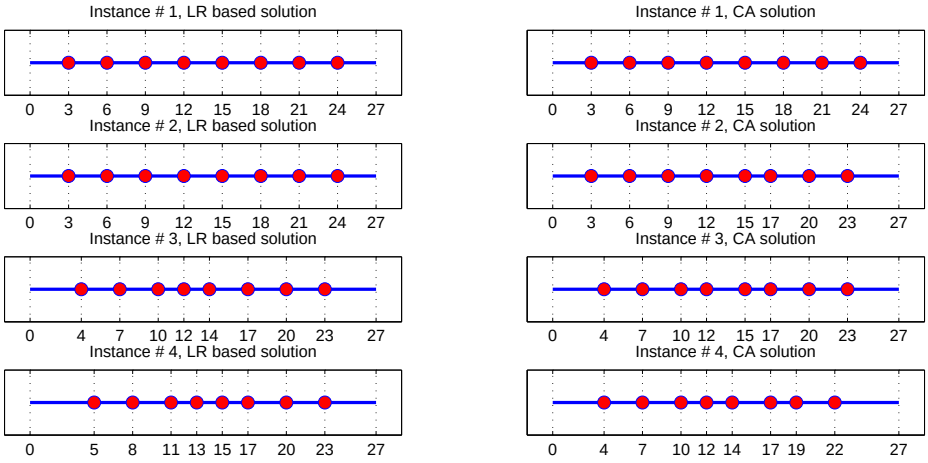


Figure 5.12: Optimal sensor deployment for solutions in Table 5.4.

The advantage of the CA model over its discrete counterparts becomes more significant as the problem size continues increasing. Table 5.5 shows the test results for different instance sizes, where $c(x) = 0.5 + h(x)$, $q(x) = 0.4h(x)$, $M \in \{27, 54, 81, 108, 135\}$ and $N \in \{8, 17, 26, 35, 44\}$. We see that with the instance size increases, the LR residual gap keeps increasing, which indicates that the solution quality from the discrete model deteriorates with the problem size. While the CA solutions obtained within negligible solution times are consistently better than those from LR.

Table 5.5: Sensitivity of solution quality over the problem instance size.

Algorithm	#	M	N	Objective	LR residual gap	Difference from the LR solution	Solution time (sec)
LR based	1	27	8	127.372	41 %	-	1800
	2	54	17	259.345	57 %	-	1800
	3	81	26	383.392	80 %	-	1800
	4	108	35	518.485	93 %	-	1800
	5	135	44	643.298	98 %	-	1800
Discrete CA Solution	1	27	8	127.974	-	≈ 0 %	≈ 0
	2	54	17	253.953	-	-2 %	≈ 0
	3	81	26	380.37	-	-1 %	≈ 0
	4	108	35	506.846	-	-2 %	≈ 0
	5	135	44	636.008	-	-1 %	≈ 0

5.6 List of Symbols

$A(x)$: Spacing between two neighboring sensors

a_{ijk}: DisList of variable definitionsto k along path i

b_i^c: Benefit coefficient for flow coverage on path i

b_i^t: Benefit coefficient for path coverage on path i

$A(x)$: Spacing between two neighboring sensors

a_{ijk}: Distance from j to k along path i

b_i^c: Benefit coefficient for flow coverage on path i

b_i^t: Benefit coefficient for path coverage on path i

$B(\mathcal{Q})$: Surveillance benefit from sensor installations at $\mathcal{Q}$

$B_i(\mathcal{Q})$: Benefit on path i from sensor installations at $\mathcal{Q}$

$C(x, A(x))$: Unit-length continuum approximation cost

$\hat{C}(x, A(x), \omega)$: Unit-length continuum approximation cost after Lagrangian relaxation

d: Downstream virtual location

$e(x, a)$: Estimation error of a segment of length a centered at $x \in [0, M]$

e_{ijk}: State estimation error measure for the segment on path i in between locations $j \in \mathcal{J}_{id-}$ and $k \in \mathcal{J}_{ij+}$

f_j: Sensor installation cost at location j

$F(\mathbf{X}')$: Auxiliary function for the linear relaxation based algorithm

$\mathcal{I}$: Set of O-D paths on the network

$\mathcal{I}_j$: Set of paths that pass the same location $j \in \mathcal{J}$

v_i: The traffic volume on path $i \in \mathcal{I}$

$\mathcal{J}$: $\bar{\mathcal{J}} \bigcup \{u, d\} = \bigcup_{\forall i} \mathcal{J}_i$

$\mathcal{J}^h$: Location set with a high failure probability

$\mathcal{J}^m$: Location set with a medium failure probability

$\mathcal{J}^l$: Location set with a low failure probability

$\mathcal{J}_{ij+}$: Set of candidate locations downstream to j on path i

$\mathcal{J}_{ij-}$: Set of candidate locations upstream to j on path i

$\mathcal{J}_{ijk}$: $\mathcal{J}_{ij+} \setminus (\mathcal{J}_{ik+} \bigcup \{k\})$

$\bar{\mathcal{J}}$: Set of all candidate locations

$\mathcal{J}_i$: $\bar{\mathcal{J}}_i \bigcup \{u, d\}$

$\bar{\mathcal{J}}_i$: Set of candidate locations on path $i \in \mathcal{I}$

N: Maximum number of facilities that the budget allows to build

M: Ending mileage on a corridor

M_i: Ending mileage on path $i \in \mathcal{I}$

M_{ij}: Mileage of location $j \in \mathcal{J}_i$ on path $i \in \mathcal{I}$

$\mathbf{P} = \{P_{ijkr}\}$: Probability that sensors at j and k are paired up at level r on path i

$q(x)$: Sensor failure probability at x

q_j: Sensor failure probability at location j

$\bar{q}$: Average sensor failure probability

$\bar{q}$: Scalar to capture sensor failure probability variation

$\mathcal{Q}$: Set of locations

r_{ijd}: Level for j to pair up d

R_{ij}: $\max_{k \in \mathcal{J}_{ij+}} R_{ijk} = \min\{|\mathcal{J}_{ij+}| - 1, N\}$

R_{ijk}: maximum possible pairing-up level for two sensors at $j \in \mathcal{J}_{id-}$ and $k \in \mathcal{J}_{ij+}$

S_i: Number of sensors installed on path $i \in \mathcal{I}$

$w(x)$: Ground-truth traffic state at x

$\hat{w}(x)$: Online estimation of $w(x)$ with sensor data

$\bar{w}(x)$: Offline estimation of $w(x)$

u: Upstream virtual location

$\mathbf{X} = \{X_j\}_{j \in \mathcal{J}}$: $X_j = 1$ ($X_j = 0$) if a sensor is (not) installed at j

$\mathbf{X}^G$: Greedy algorithm solution

$\mathbf{X}^I$: Interchange algorithm solution

$\mathbf{X}^L$: Linear relaxation based algorithm solution

$\mathbf{X}^R$: Rounded integral solution

$\mathbf{Y} = \{Y_{ijkr}\}$: $Y_{ijkr} = 1$ ($Y_{ijkr} = 0$) if sensors at j and k are (not) paired up at level r on path i

z^{LB}: Objective value of the best-known feasible solution in LR

$\Delta(\lambda, \gamma)$: Lagrangian relaxation objective

$\Delta^A(\lambda, \gamma)$: Approximated Lagrangian relaxation objective

$\Phi(\mathbf{X})$: Total error for sensor deployment $\mathbf{X}$

$\Phi^L(\mathbf{X})$: Linear relaxed objective for sensor deployment $\mathbf{X}$

$\gamma = \{\gamma_{ikr}\}$: Lagrangian multipliers for the Lagrangian relaxation algorithm

$\lambda = \{\lambda_{ijk}\}$: Lagrangian multipliers for the Lagrangian relaxation algorithm

μ^m: Control scalar in the Lagrangian relaxation algorithm

t^m: Step size in the Lagrangian relaxation algorithm

ω: Lagrangian multiplier for continuum approximation

References

Ageev, A. A., Sviridenko, M. I., 1999. Approximation algorithms for maximum coverage and max cut with given sizes of parts. In: Integer Programming and Combinatorial Optimization: 7th International IPCO Conference, Graz, Austria. Vol. 1610. Springer Berlin, pp. 17–30.

Bakkaloglu, M., Wylie, J. J., Wang, C., Ganger, G. R., 2002. On correlated failures in survivable storage systems. Technical Report CMU-CS-02-129. Carnegie Mellon University.

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

Ban, X. J., Herring, R., Margulici, J., Bayen, A. M., 2009. Optimal sensor placement for freeway travel time estimation. Presented at Transportation Research Board 88th Annual Meeting, Jan 2009.

Bartin, B., Ozbay, K., Iyigun, C., 2007. A clustered based methodology for determining the optimal roadway configuration of detectors for travel time estimation. Transportation Research Record 2000, 98–105.

Batta, R., Dolan, J., Krishnamurthy, N., 1989. The maximal expected covering location problem - revisited. Transportaion Science 23 (4), 277–287.

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), 332350.

Bianco, L., Confessore, G., Gentili, M., 2006. Combinatorial aspects of the sensor location problem. Annals of Operation Research 144 (1), 201–234.

Bianco, L., Confessore, G., Reverberi, P., 2001. A network based model for traffic sensor location with implications on O/D matrix estimates. Transportation Science 35 (1), 50–60.

Campbell, J. F., 1993a. Continuous and discrete demand hub location problems. Transportation Research Part B 27 (6), 473–482.

Campbell, J. F., 1993b. One-to-many distribution with transshipments: An analytic model. Transportation Science 27 (4), 330–340.

Caprara, A., Fischetti, M., Toth, P., 1999. A heuristic method for the set covering problem. Operations Research 47 (5), 730–743.

Carbunar, B., Ramanathan, M., Koyuturk, M., Grama, A., Hoffmann, C., 2005. Redundant-reader elimination in RFID systems. In: Proceedings of the 2nd IEEE International Conference on Sensor and Ad Hoc Communications and Networks (SECON). Santa Clara, September 2005.

Castillo, E., Menendez, J. M., Jimenez, P., 2008. Trip matrix and path flow reconstruction and estimation based on plate scanning and link observations. Transportation Research Part B 42 (5), 455–481.

Christofides, N., 1975. Graph theory: An algorithmic approach (Computer science and applied mathematics). Academic Press, Inc., Orlando, FL, USA.

Church, R., ReVelle, C., 1974. The maximal covering location problem. Papers in Regional Science 32 (1), 101–118.

Cornuejols, G., Fisher, M. L., Nemhauser, G. L., 1977. Location of bank accounts to optimize float: An analytic study of exact and approximate algorithms. Management Science 23 (8), 789–810.

Cui, T., Ouyang, Y., Shen, Z. M., 2009. Reliable facility location under the risk of disruptions. Operations Research, in press.

Daganzo, C., 1984a. The distance traveled to visit N points with a maximum of C stops per point: A manual tour-building strategy and case study. Transportation Science 18 (4), 331–350.

Daganzo, C., 1984b. The length of tours in zones of different shapes. Transportation Research Part B 18 (2), 135–146.

Daganzo, C., 1999. On planning and design of logistics systems for uncertain environments. In: New Trends in Distribution Logistic. Springer, Berlin, Germany, pp. 3–21.

Daganzo, C., 2005. Logistics System Analysis (4th Edition). Springer, Berlin, Germany.

Daganzo, C., Newell, G., 1986. Configuration of physical distribution networks. Networks 16 (2), 113–132.

D'Amico, E., 2002. West coast port lockout creates problems for chemical shippers. Chemical Week 164 (40), 10–10.

Dasci, A., Verter, V., 2001. A continuous model for production-distribution system design. European Journal of Operational Research 129 (2), 287–298.

Daskin, M., 1982. Application of an expected covering model to emergency medical service system design. Decision Science 13 (3), 416–439.

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

Daskin, M., 1995. Network and Discrete Location: Models, Algorithms, and Applications. John Wiley, New York.

Drezner, Z. (Ed.), 1995. Facility Location: A Survey of Applications and Methods. Springer, New York.

Ehlert, A., Bell, M. G., Grosso, S., 2006. The optimisation of traffic count locations in road networks. Transportation Research Part B 40 (6), 460 – 479.

Fei, X., Mahmassani, H. S., 2008. Two-stage stochastic model for sensor location problem in a large-scale network. Presented at Transportation Research Board 87th Annual Meeting, Jan 2008.

Fei, X., Mahmassani, H. S., Eisenman, S. M., 2007. Sensor coverage and location for real-time traffic prediction in large-scale networks. Transportation Research Record 2039 (1), 1–15.

Feige, U., 1998. A threshold of ln n for approximating set cover. Journal of the ACM 45 (4), 314–318.

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

Gentili, M., Mirchandani, P. B., 2005. Locating active sensors on traffic network. Annals of Operations Research 136 (1), 229–257.

Godoy, L. A., 2007. Performance of storage tanks in oil facilities damaged by hurricanes katrina and rita. Journal of Performance of Constructed Facilities 21 (6), 441–449.

Goyal, A., Nicola, V. F., 1990. Modeling of correlated failures and community error recovery in multi-version software. IEEE Transaction on Software Engineering 16 (3), 350 – 359.

Griffiths, D. A., 1973. Maximum likelihood estimation for the beta-binomial distribution and an application to the household distribution of the total number of cases of a disease. Biometrics 29 (4), 637–648.

Hakimi, S. L., 1964. Optimum locations of switching centers and the absolute centers and medians of a graph. OPERATIONS RESEARCH 12 (3), 450–459.

Hall, R., 1984. Travel distance through transportation terminals on a rectangular grid. Journal of the Operational Research Society 35 (12), 1067–1078.

Hall, R., 1986. Discrete models / continuous models. OMEGA - International Journal of Management Science 14 (3), 213–220.

Hall, R., 1989. Configuration of an overnight package air network. Transportation Research Part A 23 (2), 139–149.

Hu, S.-R., Peeta, S., Chu, C.-H., 2009. Identification of vehicle sensor locations for link-based network traffic applications. Transportation Research Part B 43 (8-9), 873–894.

Kerner, B., Rehborn, H., 1996. Experimental properties of complexity in traffic flow. Physical Review E 53 (5), R4275–R4278.

Kleindorfer, P., Saad, G., 2005. Managing disruption risks in supply chains. Production and Operations Management 14 (1), 53–68.

Lam, W., Lo, H., 1990. Accuracy of O-D estimates from traffic counts. Traffic Engineering and Control 31, 358–367.

Langevin, A., Mbaraga, P., Campbell, J., 1996. Continuous approximation models in freight distribution: An overview. Transportation Research Part B 30 (3), 163–188.

Li, X., Ouyang, Y., 2007. Railroad wayside detector location solver (RWDLS) software: User's guide.

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

Li, X., Peng, F., Ouyang, Y., 2010. Measurement and estimation of traffic oscillation properties. Transportation Research Part B 44 (1), 1–14.

Mirchandani, P., Gentili, M., He, Y., 2009. Location of vehicle identification sensors to monitor travel-time performance. Intelligent Transport Systems 3 (3), 289–303.

Mirzain, A., 1985. Lagrangian relaxation for the starstar concentrator location problem: approximation algorithm and bounds. Networks 15 (1), 1 – 20.

Newell, G., 1971. Dispatching policies for a transportation route. Transportation Science 5 (1), 91–105.

Newell, G., 1973. Scheduling, location, transportation and continuum mechanics: some simple approximations to optimization problems. SIAM Journal on Applied Mathematics 25 (3), 346–360.

Newell, G. F., 1993. A simplified theory of kinematic waves in highway traffic, part iii: Multi-destination flows. Transportation Research Part B 27 (4), 305–313.

Ouyang, Y., 2007. Design of vehicle routing zones for large-scale distribution systems. Transportation Research Part B 41 (10), 1079–1093.

Ouyang, Y., Daganzo, C., 2006. Discretization and validation of the continuum approximation scheme for terminal system design. Transportation Science 40 (1), 89–98.

Ouyang, Y., Li, X., Barkan, C. P. L., Kawprasert, A., Lai, Y.-C., 2009. Optimal locations of railroad wayside defect detection installations. Computer-Aided Civil and Infrastructure Engineering 24 (5), 309–319.

Peng, F., Ouyang, Y., 2010. A continuum approximation approach to discrete facility location problems. Working paper, University of Illinois at Urbana Champaign.

Rajagopal, R., Varaiya, P., 2007. Health of californias loop detector system. California PATH Research Report,UCB-ITS-PRR-2007-13.

Revelle, C., Hogan, K., 1989. The maximum availability location problem. Transportation Science 23 (3), 192 – 200.

Schewe, P. F., 2004. The massive Northeast blackout. Physics Today 57 (10), 9–9.

Sheffi, Y., 1985. Urban Transportation Networks: Equilibrium Analysis with Mathematical Programming Methods. Prentice-Hall, Inc., Englewood Cliffs, N.J.

Sherali, H., Desai, J., Rakha, H., El-Shawarby, I., 2006. A discrete optimization approach for locating automatic vehicle identification readers for the provision of roadway travel times. Transportation Research Part B 40 (10), 857–871.

Sherali, H. D., Alameddine, A., 1992. A new reformulation-linearization technique for bilinear programming problems. Journal of Global Optimization 2 (4), 379–410.

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

Tang, D., Iyer, R. K., 1992. Analysis and modeling of correlated failures in multicomputer systems. IEEE Transactions on Computers 41 (5), 567–577.

Toth, L. F., 1959. Sur la representation dune population infinie par un nombre fini delements. Acta Mathematica Hungarica 10 (3-4), 299–304.

Weber, A., 1957. Theory of the Location of Industries. University of Chicago Press, IL.

Yang, H., Iida, Y., Sasaki, T., 1991. An analysis of the reliability of an origin-destination trip matrix estimated from traffic counts. Transportation Research Part B 25 (5), 351 – 363.

Yang, H., Yang, C., Gan, L., 2006. Models and algorithms for the screen line-based traffic-counting location problems. Computers & Operations Research 33 (3), 836 – 858.

Yang, H., Zhou, J., 1998. Optimal traffic counting locations for origin-destination matrix estimation. Transportation Research Part B 32 (2), 109 – 126.