RELIABLE DESIGN OF INTERDEPENDENT SERVICE FACILITY SYSTEMS UNDER CORRELATED DISRUPTION RISKS

BY

SIYANG XIE

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, 2018

Urbana, Illinois

Doctoral Committee:

Professor Yanfeng Ouyang, Chair
Assistant Professor Karthik Chandrasekaran
Professor Xin Chen
Assistant Professor Hadi Meidani

ABSTRACT

Facility location decisions lie at the center of planning many infrastructure systems. In many practice, public agencies (e.g., governments) and private companies (e.g., retailers) need to locate facilities to serve spatially distributed demands. For example, governments locate public facilities, e.g., hospitals, schools, fire stations, to provide public services; retail companies determine the locations of their warehouses and stores to provide business. The design of such facility systems involves considerations of investment of facility construction and transportation cost of serving demands, so as to maximize the system operational efficiency and profit.

Recently, devastating infrastructure damages observed in real world show that infrastructure facilities may be subject to disruptions that compromise individual facility functionality as well as overall system performance. This emphasizes the necessity of taking facility disruptions into consideration during planning to balance between system efficiency and reliability. Furthermore, facility systems often exhibit complex interdependence when: (1) facilities are spatially correlated due to physical connections/interrelations, and (2) facilities provide combinatorial service under cooperation, competition and/or restrictions. These further complicate the facility location design. Therefore, facility location models need to be extended to tackle all these challenges and design a reliable interdependent facility system.

This dissertation aims at investigating several important and challenging topics in the reliable facility location context, including facility correlations, facility combinations, and facility districting. The main work of this PhD research consist of: (1) establishing a new systematic methodological framework based on supporting stations and quasi-probabilities to describe and decompose facility correlations into succinct mathematical representations, which allows compact mathematical formulations to be developed for planning facility locations under correlated facility disruptions; (2) expanding the modeling framework to allow facilities to provide combinatorial service; e.g., in the context of sensor deployment problems, where sensors work in combinations to provide positioning/surveillance service via trilateration procedure; and (3) incorporating the concepts of spatial districting into the reliable facility location context, with the criteria of spatial contiguity, compactness, and demand balance being ensured.

First, in many real-world facility systems, facility disruptions exhibit spatial correlations, which

have strong impacts on the system performance, but are difficult to be described with succinct mathematical models. We first investigate facility systems with correlations caused by facilities' share of network access points (e.g., bridges, railway crossings), which are required to be passed through by customers to visit facilities. We incorporate these network access points and their probabilistic failures into a joint optimization framework. A layer of supporting stations are added to represent the network access points, and are connected to facilities to indicate their real-world relationships. We then develop a compact mixed-integer mathematical model to optimize the facility location and customer assignment decisions. Lagrangian relaxation based algorithms are designed to effectively solve the model. Multiple case studies are constructed to test the model and algorithm, and to demonstrate their performance and applicability.

Next, when there exists no real access points, facilities could also be correlated if they are exposed to shared hazards. We develop a virtual station structure framework to decompose these types of facility correlations. First, we define three probabilistic representations of correlated facility disruptions (i.e., with *scenario*, *marginal*, and *conditional* probabilities), derive pairwise transformations between them, and theoretically prove their equivalence. We then provide detailed formulas to transform these probabilistic representations into an equivalent virtual station structure, which enables the decomposition of any correlated facility disruptions into a compact network structure with only independent failures, and helps avoid enumerating an exponential number of disruption scenarios. Based on the augmented system, we propose a compact mixed-integer optimization program, and design several customized solution approaches based on lagrangian relaxation to efficiently solve the model. We demonstrate our methodology on a series of numerical examples involving different correlation patterns and varying network and parameter settings.

We then apply the reliable location modeling framework to sensor deployment problems, where multiple sensors work in combinations to provide combinatorial coverage service to customers via trilateration procedure. Since various sensor combinations may share common sensors, one combination is typically interrelated with some other combinations, which leads to internal correlations among the functionality of sensors and sensor combinations. We address the problem of where to deploy sensors, which sensor combinations are selected to use, and in what sequence and probability to use these combinations in case of disruptions. A compact mixed-integer mathematical model is developed to formulate the problem, by combining and extending the ideas of assigning

back-up sensors and correlation decomposition via supporting stations. A customized solution algorithm based on Lagrangian relaxation and branch-and-bound is developed, together with several embedded approximation subroutines for solving subproblems. A series of numerical examples are investigated to illustrate the performance of the proposed methodology and to draw managerial insights.

Finally, we develop an innovative reliable network districting framework to incorporate districting concepts into the reliable facility location context. Districting criteria including spatial contiguity, compactness, and demand balance are enforced for location design and extended in considerations of facility disruptions. The problem is modeled into a reliable network districting problem, in the form of a location-assignment based model. We develop customized solution approaches, including heuristics (i.e., constructive heuristic and neighborhood search) and set-cover based algorithms (e.g., district generation, lower bound estimation) to provide near-optimum solution with optimality gap. A series of hypothetical cases and an empirical full-scale application are presented to demonstrate the performance of our methodology for different network and parameter settings.

To my family and my loved

ACKNOWLEDGMENTS

First, I would like to express my deepest gratitude to my adviser, mentor and friend, Professor Yanfeng Ouyang, for his constant help, guidance, inspiration, and encouragement on all aspects of my life throughout my graduate studies at the University of Illinois at Urbana-Champaign. It was he who guided me into the areas of transportation and operations research, provided me with access to diverse challenging problems, and helped me conceive and carry out this dissertation work. His challenging mind, dedicated attitude, and endless passion shaped my determination toward excellence and in pursuit of my career.

I am also very grateful to the three other members of my committee, Professor Karthik Chandrasekaran, Professor Xin Chen and Professor Hadi Meidani. Their valuable comments and suggestions along the way have substantially helped me improve this dissertation. From their classes and seminars, I have also learned a great deal of fundamental knowledge and methodologies on which my research is based. This research has also much benefited from the courses I have taken from many departments at the University of Illinois. I want to express my sincere thanks to all the professors I have learned from and interacted with.

Next, I would like to particularly thank my collaborators Xiaopeng Li, Xi Chen, Zhaodong Wang, Kamalesh Somani, Jing Huang, Kun An, Chao Lei, Zhoutong Jiang, who collaborated with me on a number of research topics and helped me with modeling, programming, writing, etc. I also sincerely appreciate the friendship and generous help, both in research and in life, from all my former and current colleagues, including Fan Peng, Yun Bai, Taesung Huang, Seyed Mohammad Nourbakhsh, Leila Hajibabai, Xin Wang, Weijun Xie, Liqun Lu, Laura Ghosh, Antoine Petit, Han Liu.

Finally, I want to send my special thanks to my family, my loved, and my friends. Without their love and support, I could not have gone through the hard times to finish this dissertation.

I acknowledge financial support from the National Science Foundation, CSX Transportation, Inc., the Roadway Safety Institute (U.S. DOT Region V University Transportation Center), and the Association of American Railroads (AAR).

TABLE OF CONTENTS

CHAPTER 1:

INTRODUCTION

1.1 Motivation

Facility location decisions lie at the center of planning many infrastructure systems. In many practice, public agencies (e.g., governments) and private companies (e.g., retailers) both need to locate their facilities to serve spatially distributed demands/customers. For example, governments locate various public facilities, such as hospitals, schools, fire stations, to provide public services; retail companies determine the locations of their facilities including warehouses, assembly plants, stores, etc, to sell goods and provide business. The design of all such facility systems generally involves considerations of fixed investment of facility construction and transportation cost of serving demands, so as to maximize the operational efficiency and service profit of the system.

Recently, observations of uncertainties in many real-world infrastructure systems have further complicated the facility location planning. There are two basic sources of uncertainties. First, demands could be stochastic and thus cannot be accurately identified beforehand, which introduces additional modeling difficulty compared to the case with deterministic demands. Plenty of researches has been done to study demand uncertainties in the past few decades. The second source of uncertainties is facility disruptions, revealed by recent devastating infrastructure damages and catastrophic system failures observed in natural and anthropogenic disasters. Facilities may become unavailable from time to time due to either exogenous or endogenous factors. When a facility is disrupted, its customers have to seek service from some other functioning alternatives or even completely give up their services. Therefore, ignoring the possibilities of facility disruptions often yields a suboptimal system design that is vulnerable to even infrequent facility disruptions. This emphasizes the necessity of taking real-world facility disruptions into consideration when planning

a facility system (Snyder and Daskin, 2005).

Under probabilistic facility disruptions, one has to deal with a huge number of disruption scenarios, each of which is a unique combination of realized functioning states of the facilities. If each facility can be at one of two possible states (i.e., operating or disrupted) at any time, it is easy to see that the total number of disruption scenarios is two to the power of the facility count, and thus it grows exponentially with the system size. A reliable design needs to evaluate (and then optimize) the expected system performance across all these disruption scenarios, which is apparently a very tedious task. To get around this issue, many studies assume that facility disruptions occur independently (Snyder and Daskin, 2005; Chen et al., 2011). This assumption enables each individual facility's performance to be evaluated separately in a small polynomial time, which results in a much less complexity of evaluating the expected system performance and further leads to fruitful developments of compact mathematical models and efficient solution algorithms for reliable facility location design (Cui et al., 2010; Daskin, 2011; Li and Ouyang, 2010).

However, in many real-world systems, facility disruptions exhibit spatial correlations (e.g., due to shared supporting infrastructure or simultaneous exposition to hazards). Disruption correlations tend to have a strong impact on the performance of a reliable facility location design. Consider a simple network where two facilities A and B jointly serve one unit of demand from a customer. The costs for serving the demand from these two facilities are 10 and 20 units, respectively, and the penalty for not serving the demand is 100 units. When both facilities are perfectly reliable, the demand will obviously be served by A with a total cost of 10 units. When the facilities are subject to disruption, the demand will be served by A as long as A is functioning (i.e., event A), or by B if A is disrupted but B is functioning (i.e., event $\overline{A}B$), or the customer will bear the penalty if both A and B fail (i.e., event $\overline{AB}$). In the case where A and B fail independently with an equal probability of 0.5, the expected service cost is $10 \times (0.5) + 20 \times (0.5 \times 0.5) + 100 \times (0.5 \times 0.5) = 35$ units. If the facility disruptions are positively correlated, say $P(AB) = P(\overline{AB}) = 0.4, P(\overline{A}B) = P(A\overline{B}) = 0.1$, the expected service cost becomes $10\times(0.1+0.4)+20\times0.1+100\times0.4 = 47$ units. If the facility disruptions are negatively correlated, say $P(AB) = P(\overline{AB}) = 0.1, P(\overline{A}B) = P(A\overline{B}) = 0.4$, the expected service cost becomes $10 \times (0.1 + 0.4) + 20 \times 0.4 + 100 \times 0.1 = 23$ units. Although the disruption probability of each facility remains 0.5 in all the three cases, we can see that the presence of disruption correlations (both positive and negative) significantly affects the expected system service cost, and thus such

factors should be carefully considered and incorporated.

In traditional facility location problems, each facility functions individually to serve customers. In some recent applications, however, facilities work in combinations to provide integrated combinatorial services/supplies. For example, in the supply chain context, downstream processes/services are typically in need of various types of products/materials from its upstream facilities, thus downstream customers seek services from multiple upstream facilities simultaneously. In the sensor deployment context, sensors (which are facilities) are working in combinations to provide sensory coverage, and the effectiveness of the sensor system highly depends on the quality (working range and precision level) and quantity of sensors installed in the area. For the sensor deployment problems, if we further consider the possibility of sensor disruptions, since various sensor combinations may share common sensors, disruption of one sensor combination could be directly related to that of another combination. This leads to internal correlation among the functionality of multiple sensors and sensor combinations. Therefore, where to deploy sensors, how to form sensor combinations, which sensor combinations to use, and in what sequence and probability to use them in case of sensor disruptions, are nontrivial questions, and should be carefully investigated.

Another challenging extension of reliable facility location problem is to incorporate the network districting concepts. Specifically, the systems/networks are first partitioned into multiple geographical districts following various operational criteria including spatial contiguity, compactness, and demand balance, then the customers within each partitioned district are assigned to a supplier/facility for service. An example is the call center design problem, which aims at efficiently handling the incoming calls by assigning the calls from different spatial regions to the best call responders. A good design typically has good characteristics such as (1) the expected workload is well balanced across call responders; and (2) the spatial district served by a responder is contiguous and compact in shape so as to satisfy some practical operational requirements. In the integrated framework with both districting and reliable location considerations, interrelations among facilities are introduced from the restrictions enforced on the demand assignment, which complicates the location-assignment decisions. In addition, new/adapted customized methods are needed to address the various districting criteria and to combine the reliable facility location and network districting modules.

1.2 Contributions

This Ph.D. research aims at investigating several important and challenging extended topics in the reliable facility location context, including: (i) facility correlations; (ii) facility combinations; and (iii) facility districting. All these extensions are widely recognized to be extremely complicated and challenging to tackle, due to the additional complexity associated with the new ingredients added to the traditional reliable facility location problems. The main contributions of this PhD research consist of: (1) establishing a new systematic methodology framework based on quasi-probabilities and supporting stations to describe and decompose facility correlations into succinct mathematical representations, which allows compact mathematical formulations to be developed for planning facility locations under correlated facility disruptions; (2) expanding the reliable facility location modeling framework to allow facilities to provide combinatorial service; e.g., in the context of sensor deployment problems, sensors work in combinations to provide positioning/surveillance service via trilateration procedure; and (3) incorporating reliability concepts into the spatial districting context (e.g., for political, school, service systems), where spatial contiguity, compactness, and demand balance must be ensured, and developing various types of new customized model formulations and solution approaches.

First, the dissertation relaxes the assumption on independent facility disruptions by planning facility systems in which facilities share external network access points and failures of the shared access points introduce complex facility correlation patterns. An additional layer of supporting stations representing the network access points are added to the orginal customer-facility system, and are connected to the facilities to indicate the real-world relationships between the facilities and access points. A compact mixed-integer mathematical model is built upon the stations to address the joint optimization of facility location and customer assignment decisions. Customized solution approaches based on Lagrangian relaxation are designed to solve the model efficiently.

Next, the dissertation develops a systematic station structure framework to decompose more general correlated facility disruptions. First, we define three probabilistic representations of correlated facility disruptions (i.e., with *scenario* probabilities, *marginal* probabilities and *conditional* probabilities), derive pairwise transformations between them, and theoretically prove their equivalence. We then provide detailed formulas to transform these probabilistic representations into an

equivalent adapted virtual station structure, which enables us to decompose any correlated facility disruptions into a compact network structure (consisting of customers, facilities, and stations) that can be efficiently modeled with only independent station failures. This in turn allows us to avoid enumerating an exponential number of disruption scenarios in the system performance evaluation. Based on the augmented customer-facility-station system, we extend the mixed-integer program proposed in the previous chapter, and design several new customized solution approaches to more efficiently solve the model. We demonstrate our methodology on a series of numerical examples involving different correlation patterns and varying network and parameter settings.

We then apply the reliable location modeling framework to sensor deployment problems, where multiple sensors work in combinations to provide combinatorial coverage service to customers via trilateration procedure. Since various sensor combinations may share common sensors, one combination is typically interrelated with some other combinations, which leads to internal correlations among the functionality of sensors and sensor combinations. We address the problem of where to deploy sensors, which sensor combinations are selected to use, and in what sequence and probability to use these combinations in case of disruptions. A compact mixed-integer mathematical model is developed to formulate the problem, by combining and extending the ideas of back-up assignments and correlation decomposition via supporting stations. A customized solution algorithm based on Lagrangian relaxation and branch-and-bound is developed, together with several embedded approximation subroutines for solving subproblems. A series of numerical examples are investigated to illustrate the performance of the proposed methodology and to draw managerial insights.

Finally, we develop an innovative reliable network districting framework to incorporate districting concepts into the reliable facility location context. The system/network is partitioned into multiple districts following various districting criteria including spatial contiguity, compactness, and demand balance, and the demands in each district are assigned to a particular facility. The districting criteria are further extended in considerations of facility reliability issues. The problem is modeled into a reliable network districting problem, in the form of a location-assignment based model. We develop customized solution approaches, including heuristics (i.e., constructive heuristic and neighborhood search) and set-cover based algorithms (e.g., district generation, error estimation) to provide near-optimum solutions and corresponding optimality gaps. A series of hypothetical test cases and an empirical full-scale application are presented to demonstrate the performance and

effectiveness of our methodology for different network sizes and parameter settings.

1.3 Outline

This dissertation is organized as follows. Chapter 2 first summarizes literatures in the reliable facility location context. Mixed-integer programming models are normally used to formulate the reliable location problems under probabilistic facility disruptions. The models are typically solved by solvers or implementing customized algorithms to obtain optimal or near-optimum solutions. We also review the literatures and methodologies in each of the three studied topics: facility correlations, sensor deployment, and network districting, and present the various models that have been proposed.

Established on previous work in reliable facility location context, Chpater 3 relaxes the assumption on independent facility disruptions and investigates facility systems with correlations caused by the influence from shared network access points. A layer of supporting stations are constructed to represent the real network access points, and a compact mixed-integer mathematical model is built upon the stations to optimize the location and assignment decisions for such facility systems. Customized solution approaches based on Lagrangian relaxation and branch and bound are designed to efficiently solve the model.

Facility correlations may be caused by facilities' simultaneous exposion to hazards even when there is no physical shared access points. Chapter 4 develops a systematic methodological framework to describe and analyze any type of correlated facility disruptions in succinct mathematical models. A decomposition scheme is designed to capture the effects of facility correlations by adding virtual supporting stations. We also discussed analytical properties of the decomposition scheme and the resulting station structure.

Based on the station structure provided by the decomposition scheme, Chapters 5 extends the optimization formulation in Chapter 3 to model any type of correlated facility disruptions. Additional constraints, and new customized algorithms with more accurate approximation subroutines are developed to improve the solution quality and efficiency. Multiple numerical case studies with various types of correlated facility disruptions are carried out for the purpose of demonstration.

Chapter 6 applies the reliable facility location modeling framework to sensor deployment context, in which sensors work in combinations to provide combinatorial service. The ideas of support-

ing stations are adopted with adjustment to represent the combinations of sensors. A mixed-integer mathematical model is formulated to determine the optimal sensor locations, the sensor combination plans, and the backup assignment decisions. Several approximation subroutines are carefully designed inside a Lagrangian relaxation framework to solve the model.

Next, Chapter 7 studies a reliable network districting problem to integrate both the reliable facility location and the network districting modules. A series of modeling techniques are adopted to address multiple districting criteria as well as the facility reliability considerations. Formulation based on location-assignment modeling techniques is proposed, with customized algorithms designed. The approaches are applied to solve several numerical examples including an empirical full-scale application, to demonstrate their performance and to draw managerial insights.

Finally, Chapter 8 concludes the dissertation and discusses future research directions.

CHAPTER 2:

LITERATURE REVIEW

This chapter reviews work in reliable facility location literature, as well as its several extensions. There have been a lot of studies done on the reliable facility location problems, and most of them formulated the problems into discrete integer linear programs. The ideas of probabilistic disruptions, backup facility assignments, and expected system evaluation are widely adopted in these programs and models. However, to the best of our knowledge, only a few efforts have been made to study the extensions/topics of reliable interdependent facility system design that are investigated in this research.

2.1 Reliable Facility Location

Facility location problems have been intensively studied in the past several decades, with the most original formulation dated back to 1909, since when there have been numerous studies and a large number of related models on facility location problems. (Drezner, 1995) reviewed a series of classic mathematical models for deterministic location problems including covering problems, center problems, median problems, etc. Most of these traditional studies considered deterministic infrastructure service where each built facility is assumed functioning and available for service all the time (Drezner, 1995; Daskin, 2011). Recently, researchers began to recognize that facilities may lose functionalities due to various external/internal factors such as natural disasters, adverse weather, human factors, etc. And a series of new reliable facility location models have been proposed to furnish a facility system with proper redundancy to take real-world facility disruptions into consideration (Snyder and Daskin, 2005; Li and Ouyang, 2010; Cui et al., 2010).

In the reliable facility location literature, one stream of studies focused on design-related facility

disruptions that can be prevented by fortification. Interdiction models were often used to identify critical components in an infrastructure system, and cost-effective fortification strategies were sought during facility location design (Church et al., 2004; Scaparra and Church, 2008a; Liberatore et al., 2011; Scaparra and Church, 2008b). Another stream of research focused on modeling the expected consequences of location-specific facility disruptions (Snyder and Daskin, 2005; Li and Ouyang, 2010). A comprehensive review can be found in Snyder (2006). Among a rich variety of efforts, Snyder and Daskin (2005) and Berman et al. (2009) formulated discrete models where facilities are subject to site-independent disruptions with identical failure probabilities. More recently, a series of reliable location models were proposed to allow site-dependent disruption probabilities. Berman et al. (2007) provided a nonlinear mixed-integer programming formulation as well as an efficient heuristic solution approach. Cui et al. (2010) developed two distinct sets of models (discrete and continuous) and corresponding solution algorithms to allow the disruption probabilities to be site-dependent. Li and Ouyang (2012) proposed a reliable sensor location model to optimize traffic system surveillance effectiveness where sensors are subject to site-dependent probabilistic failures. Atamtürk et al. (2012) further presented reliable location-inventory models (which allowed facilities to be subject to failures due to inventory shortage) as well as an innovative conic programming solution approach. All these studies assumed independent facility disruptions and furnish a facility system with proper redundancy so as to balance its efficiency in the normal scenario and its reliability when disruptions happen.

2.1.1 Facility Correlations

Most reliable location models hold the assumption that facility disruptions are independent. However, in many real-world facility systems, the disruptions of facilities often exhibit complex correlations, and a straightforward modeling approach would need to enumerate or simulate an exponential number of scenarios; this makes it computationally difficult to even just evaluate the system performance under a given design. To the best of our knowledge, only a few efforts have been made to address correlated facility disruptions, either exactly or approximately (e.g., Liberatore et al. (2012); Li and Ouyang (2010); Lu et al. (2015)). Liberatore et al. (2012) considered the problem of optimally protecting a capacitated median system with a limited amount of protective resources subject to disruptions, a tri-level formulation of the problem and an exact solution algorithm

based on a tree-search procedure were proposed. Li and Ouyang (2010) developed a continuum approximation model for the reliable uncapacitated fixed charge location problem where facilities are subject to spatially correlated disruptions that occur with site-dependent probabilities. Lu et al. (2015) allowed facility disruptions to be correlated with an uncertain joint distribution, and applied distributionally robust optimization to minimize the expected cost under the worst-case distribution. In addition, Huang et al. (2010) addressed a variant of the p-center model in case of large-scale emergencies, where correlated disruption was introduced by allowing many facilities to become functionless simultaneously. Gueye and Menezes (2015) considered a two-stage stochastic program model for a median problem under correlated facility disruptions, and asymptotic results were presented based on a scenario-based model formulation. Berman and Krass (2011) and Berman et al. (2013) introduced analytical approaches to help understand the effects of correlated failures in simpler spatial settings, e.g., along a line segment. Li et al. (2013) proposed a virtual station structure that transforms a facility network with correlated disruptions into an equivalent one with added virtual supporting stations, and the virtual stations were assumed to be subject to independent disruptions. An optimization model was developed to handle cases where facilities are positively correlated and the station disruption probabilities are all identical.

2.1.2 Sensor Deployment

Sensor deployement problems are natural applications where facilities work in combinations to provide services. Extensive researches have been conducted to study the sensor deployment problems. Gentili and Mirchandani (2012) provided a comprehensive literature review on existing sensor location models in traffic networks. Many of those studies aim at maximizing sensor coverage or minimizing the error/cost of estimation. Mirchandani et al. (2010) addressed the problem of locating surveillance infrastructure to cover a target surface; possible barriers that may block sensing signals were considered. Erdemir et al. (2008) developed models to study a location covering problem with consideration of both nodal and path-specific demand. Geetla et al. (2014) studied the deployment of omni-directional audio sensors that can detect vehicle crashes on a roadway. Eisenman et al. (2006) proposed a sensor location problem based on a simulation-based real-time network traffic estimation and prediction system. Fei and Mahmassani (2011) presented a multi-objective model that deploys a minimal number of passive point sensors in a roadway network considering

link information gains and origin-destination demand coverage. Danczyk et al. (2016) developed a sensor location model to minimize the error of monitoring freeway traffic condition. Various customized solution methods for sensor location problems have also been developed. Among them, Wang et al. (2005) partitioned the sensing field into smaller sub-regions and deployed sensors in these sub-regions when the working range of a sensor forms an arbitrarily shaped region (i.e., polygon). Clouqueur et al. (2003) developed a sequential decision-making approach to maximize the exposure of network travel paths to a set of sensors. The overall goal was to minimize the system cost needed to achieve a desired exposure rate. Zou and Chakrabarty (2004) proposed a virtual force strategy for sensor deployment and a probabilistic target localization algorithm to enhance sensor coverage. He (2013) presented a graphical approach to find the smallest set of network links to locate sensors, so as to infer the traffic flow on all other links. Ouyang et al. (2009) and Peng et al. (2011) investigated ways to deploy wayside sensors in a railroad network to monitor railcar traffic. Studies on optimal sensor placement, especially those in the context of trilateration, are quite limited. While deploying directional sensors that collectively form regular convex polygons, Xie and Dai (2014) optimized the number of edges and length of these polygons so as to maximize coverage accuracy. As sensor deployment on a regular lattice is usually not optimal for trilateration, Roa et al. (2007) proposed a diversified local Tabu search method where omni-directional sensors can follow a non-regular configuration. De Stefano et al. (2015) investigated the placement of sensors on an engineering structure to detect the existence, location and extent of internal damage. These studies, however, assumed that the sensing targets are homogeneously distributed in a 2-dimensional plane; this is often unrealistic in the real world. Indeed, sensor locations are critical to the overall performance of the surveillance system. For example, Ahmed et al. (2014) demonstrated the significance of sensor location in influencing real-time traffic state prediction after traffic crashes.

2.1.3 Network Districting

Districting is a well-known problem in the operations research literature. It aims at partitioning a geographical space into sub-districts under various criteria and constraints. Depending on the specific application context, operational criteria may include the district contiguity, district compactness, workload balance, socio-economic homogeneity, etc. In the literature, probably the most

intensively studied problem is regarding political districting, which divides a jurisdiction area (e.g., a state or a region) into electoral constituencies such that the political candidates from each area are elected to a parliamentary assembly. The "one man-one vote" principle requires that all districts contain approximately the same number of candidates/voters to avoid benefiting a certain party or candidate. Hess et al. (1965) is among the first several that used mathematical programming techniques to model the political districting problem. An assignment formulation with additional planning constraints was developed and an iterative heuristic algorithm was proposed. But the convergence of the algorithm or the contiguity of districts is not guaranteed. Garfinkel and Nemhauser (1970), on the other hand, considered selecting districts from a set of predefined feasible ones. The various constraints (e.g., contiguity, compactness) were implemented while defining the set of feasible districts (before implementing the optimization model). Mehrotra et al. (1998) adopted a column generation method to solve a similar problem as the one in Garfinkel and Nemhauser (1970). Bozkaya et al. (2003) considered more operational criteria, which were evaluated and incorporated as soft constraints. Health services districting is another application context that aims at partitioning a health service territory into districts and assign a certain amount of medical resources to each district. Blais et al. (2003) studied the districting for a public health clinic where five districting criteria were sought under a tabu search algorithm. School districting problem assigns residential neighborhoods to existing schools, as different important criteria for planning must be taken into account, e.g., the capacity, the accessibility, and the racial balance of each school. Notably, Ferland and Guénette (1990) proposed a decision support system to solve the school districting problem; the system included a network-based mathematical model and used several heuristic procedures to assign network edges (with students located on it) to schools. Other examples of districting problems include sales/market districting (Hess and Samuels, 1971), police districting (Camacho-Collados et al., 2015), waste/garbage collection districting (Muyldermans et al., 2002), etc.

2.2 Classic Models

2.2.1 Reliable Facility Location

Snyder and Daskin (2005) formulated discrete mathematical models for facility location design in which facilities are subject to site-independent disruptions with identical failure probabilities. Let

$\mathcal{I}$ be the set of customers, indexed by i, and $\mathcal{J}$ be the set of candidate facility locations, indexed by j. Each customer has μ_i units of demand, and the cost for serving per unit demand of customer i by facility j is denoted as d_{ij}. In addition, we associate with each customer i a cost π_i that represents the penalty cost of not serving its demand (per unit). To model the penalty cost, we add an "emergency" facility indexed by 0 that has transportation cost $d_{i0} = \pi_i, \forall i \in \mathcal{I}$. Normally, a customer i visits its nearest facility for service. Once the facility is disrupted, the customer seeks service from the next nearest functioning facility, until no functioning facility can provide service at a cost less than π_i, and the penalty cost is incurred. Let X_j denote whether a facility is open at location j and Y_{ijr} represent whether customer i visits facility j as its rth backup option. The mathematical model formulation for the reliable facility location problem with i.i.d. facility disruptions can be written as

$$\text{(RFL-IID)} \quad \min \quad \sum_{j\in\mathcal{J}} f_j X_j + \sum_{i\in\mathcal{I}} \sum_{j\in\mathcal{J}\cup\{0\}} \sum_{r=1}^{R+1} \mu_i d_{ij}(1-q)q^r Y_{ijr} \tag{2.1a}$$

$$\text{s.t.} \quad \sum_{r=1}^{R} Y_{ijr} \le X_j, \ \forall i \in \mathcal{I}, j \in \mathcal{J}, \tag{2.1b}$$

$$\sum_{j\in\mathcal{J}} Y_{ijr} + \sum_{s=1}^{r} Y_{i0s} = 1, \ \forall i \in \mathcal{I}, r = 1, 2, \cdots, R+1, \tag{2.1c}$$

$$\sum_{r=1}^{R+1} Y_{i0r} = 1, \ \forall i \in \mathcal{I}, \tag{2.1d}$$

$$X_j, Y_{ijr} \in \{0,1\}, \ \forall i \in \mathcal{I}, j \in \mathcal{J}\cup\{0\}, r = 1, 2, \cdots, R+1. \tag{2.1e}$$

The objective function (2.1a) is the summation of fixed facility open cost and expected system transportation cost. Constraints (2.1b) ensure that customers only visit open facilities. Constraints (2.1c) enforce that each customer i either visits a regular facility j as its rth option or has chosen to pay the penalty cost as its sth option, $s \le r$. Constraint (2.1d) indicate that each customer has to pay the penalty cost at some point.

When facility disruptions are heterogeneous, i.e., their disruption probabilities are site-dependent, Cui et al. (2010) extends the above formulation (RFL-IID) into the following formulation

$$\text{(RFL-HETER)} \quad \min \quad \sum_{j\in\mathcal{J}} f_j X_j + \sum_{i\in\mathcal{I}} \sum_{j\in\mathcal{J}\cup\{0\}} \sum_{r=1}^{R+1} \mu_i d_{ij} Z_{ijr} Y_{ijr} \tag{2.2a}$$

$$\text{s.t.} \quad \sum_{r=1}^{R} Y_{ijr} \le X_j, \ \forall i \in \mathcal{I}, j \in \mathcal{J}, \tag{2.2b}$$

$$\sum_{j \in \mathcal{J}} Y_{ijr} + \sum_{s=1}^{r} Y_{i0s} = 1, \ \forall i \in \mathcal{I}, r = 1, 2, \cdots, R+1, \tag{2.2c}$$

$$\sum_{r=1}^{R+1} Y_{i0r} = 1, \ \forall i \in \mathcal{I}, \tag{2.2d}$$

$$Z_{ij1} = 1 - q_j, \ \forall i \in \mathcal{I}, j \in \mathcal{J} \cup \{0\}, \tag{2.2e}$$

$$Z_{ijr} = (1 - q_k) \cdot \sum_{j' \in J} \frac{q_{j'}}{1 - q_{j'}} Z_{ij'(r-1)} Y_{ij'(r-1)},$$

$$\forall i \in \mathcal{I}, j \in \mathcal{J} \cup \{0\}, r = 2, 3, \cdots, R+1, \tag{2.2f}$$

$$X_j, Y_{ijr} \in \{0, 1\}, \ \forall i \in \mathcal{I}, j \in \mathcal{J} \cup \{0\}, r = 1, 2, \cdots, R+1. \tag{2.2g}$$

(2.2a)–(2.2d) are similar to the corresponding constraints in (RFL-IID). Given the site-dependent facility disruption probabilities, Constraints (2.2e)–(2.2f) iteratively calculate the probability that facility j served customer i at level r. Note that this formulation (RFL-HETER) is nonlinear due to the existence of terms $Z_{ijr}Y_{ijr}$.

2.2.2 Network Districting

Districting problems are typically modeled as a graph partitioning problem, which aims at providing a partitioning of the nodes in a graph (the nodes in each partition induce a connected subgraph) as a districting plan solution. Let $\mathcal{G} = (\mathcal{I}, \mathcal{E})$ denote a graph with node set $\mathcal{I}$ and edge set $\mathcal{E}$. Let $\mathcal{M}$ be the set of all possible subgraph of a graph, and a_{im} indicate whether node $i \in \mathcal{I}$ is contained in subgraph $m \in \mathcal{M}$ or not. If Z_m represent whether subgraph m is selected or not in the districting plan, the network districting problem can be formulated as the following set-cover based formulation

$$\text{(ND)} \quad \min \quad \sum_{m \in \mathcal{M}} F(z_m) \tag{2.3a}$$

$$\text{s.t.} \quad \sum_{m \in \mathcal{M}} z_m = M, \tag{2.3b}$$

$$\sum_{m \in \mathcal{M}} a_{im} z_m = 1, \ \forall i \in \mathcal{I}, \tag{2.3c}$$

$$z_m \in \{0, 1\}, \ \forall m \in \mathcal{M}, \tag{2.3d}$$

The objective (2.3a) is formulated as a function of $\{z_m\}_{m \in \mathcal{M}}$, with details customized in different applications. Constraints (2.3b) enforce that M districts are selected, and constraints (2.3c) ensure that each node i is contained in exactly one district.

CHAPTER 3:

RELIABLE FACILITY LOCATION UNDER THE RISK OF NETWORK ACCESS FAILURES

Most traditional studies on reliable facility location problems hold the assumption on independent facility disruptions. However, in many real-world facility systems, the facilities exhibit complex correlations due to various types of interdependence and connections among them. Specifically, in service systems with natural or anthropogenic barriers (e.g., rivers, railroads), customers who intend to visit facilities for service must first pass through certain network access points (e.g., bridges, railway crossings). Possible blockage or disruptions of these access points could change the customer-facility assignments or even affect reachability of various facilities, and thus introduce facility correlation issues.[a]

In this chapter, we incorporate network access points and their probabilistic failures into a joint optimization framework by generalizing the supporting station structure approach from Li et al. (2013). A layer of network access points (we call them stations in the remainder of this chapter) are added and connected to facilities to imply the real-world connections between facilities and access points. The stations are assumed to be subject to disruptions with site-dependent probabilities. We then develop a compact mixed-integer mathematical model to optimize the facility location and customer assignment decisions for the facility systems design. Lagrangian relaxation based algorithms are designed to effectively solve the proposed model. Multiple case studies are constructed to test the model and the algorithm, and to demonstrate their performance and applicability.

[a]This chapter is based on a submitted paper, Xie and Ouyang (2018).

3.1 Introduction

In many real-world facility systems, customers have to pass through certain network access points to visit facilities for service, and the access points are subject to possible blockage or disruptions. If an access point is blocked/disrupted, some customers may have to change their paths to the assigned facilities, or they may even change their facility choices – some facilities may become unreachable if all associated access points are blocked or disrupted. The unreachability of a facility can be interpreted as the facility's failure to provide service, although the facility itself may always be functioning. For example, in coastal areas or cities where rivers or lakes exist and partition the area into sub-regions, bridges, as the possible only access points to enter/leave the sub-regions, link the partitioned regions as an interconnected network (one distinct example would be the city of Venice, Italy). In such a network, if a bridge is blocked or disrupted due to external factors (e.g., structure or material damages, traffic accidents, congestions), customers originally intend to go through the bridge to visit service facilities like hospitals have to make a detour with a longer transportation distance and/or time, which may lead to significant deterioration of the service quality.

Similar situations could happen in many other contexts. For example, in a ground transportation network consisting of intersecting highway and railway corridors, the highway paths between customers (e.g., residential neighborhoods) and facilities (e.g., fire stations) may cross by railways tracks. If railroad incidents happen and cause railroad blockage, the paths between customers and facilities may be cut off as well. The customers are no longer able to receive service (e.g., emergency response) from its preferable facility in time, which could lead to catastrophic losses (e.g., imagine the fire trucks are blocked at railroad crossings). Recent years have witnessed a series of rail crashes and derailments that have led to major oil spills, tanker fire or explosions. The train carriages are forced to stop at rail track, blocking all crossings to enter or leave the affected region by the explosions. Examples include the recent catastrophic railroad incidents in Casselton, N.D., and in Quebec, Canada (Crummy, 2013; NBC News, 2013). In the U.S. Midwest, the State of Minnesota also expressed strong concerns over the close proximity of hazardous material trains to densely-populated urban areas[b], and the long blockage of rail crossing by high train traffic volumes

[b]See http://www.dot.state.mn.us/newsrels/15/03/19oiltrains.html

which may disrupt emergency response efforts[c]. As a result, the U.S. federal and local regulators have issued a number of orders on railroad incidents prevention and emergency responses deployment so as to enhance rail crossing safety and reliability (Xie et al., 2018b; Gold and Stevens, 2014). This calls for careful design of emergency service facility locations (e.g., fire stations, hospitals) as well as the adjacent network characteristics such that critical resources can be delivered and the customers can be serviced efficiently even under emergency situations.

The planning of such systems under the risk of network access disruptions involves facility location decisions and customer path design under different realizations of the functioning states of access points. As the disruptions to the systems occur at the network access points, which lead to the unreachability/unfunctionality of multiple facilities, there is no existing literature explicitly capturing these types of facility disruptions. Therefore, incorporating network access points with site-dependent failures into the facility system design calls for a new methodological framework. In light of these challenges, we extend the concept of station structure in Li et al. (2013) to address the real-world situations with unreliable network access points. Network acces points are represented as supporting stations, and are connected to candidate facility locations based on their real-world relationships/connections. A customer who intends to visit a facility must first pass through one of the facility's connected stations. The facilities are functioning all the time, but the stations are subject to probabilistic disruptions. In particular, we allow the disruption probabilities of access points to be site-dependent. With the augmented structure with additional supporting stations, we develop a compact mixed-integer mathematical model to address the joint optimization for both facility location and customer access/assignment decisions. This model is considerably very complex, hence we develop several customized solution approaches based on Lagrangian relaxation algorithm to effectively solve it. Multiple case studies are presented to not only test the model and algorithm but also to draw managerial insights.

The remainder of this chapter is organized as follows. Section 3.2 introduces the supporting station structure and presents the mixed-integer mathematical model. Section 3.3 shows the customized lagrangian relaxation based solution approaches. In Section 3.4, several case studies involving various problem and parameter settings are presented.

[c]See http://www.startribune.com/politics/statelocal/286633141.html

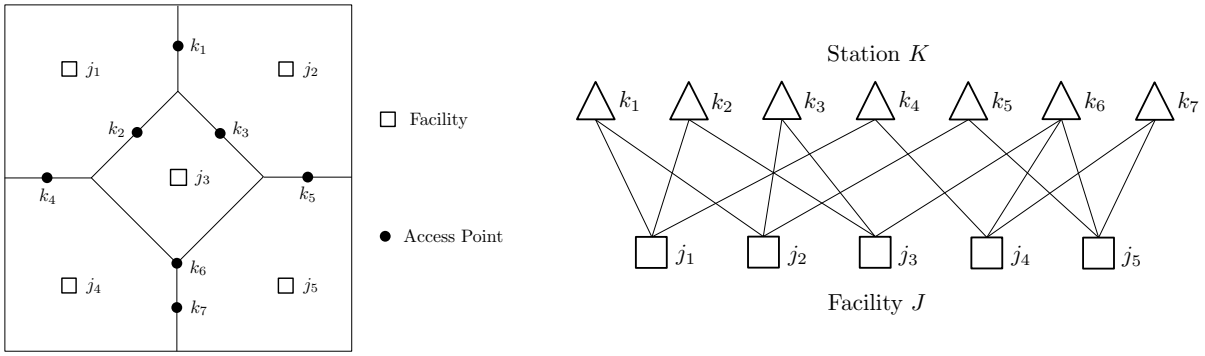
3.2 Model Formulation

In this section, we propose a new mixed-integer model formulation for the reliable facility location problem under the risk of network access failures.

3.2.1 Augmented Facility-Station Structure

In some facility systems, customers visit facilities by passing through certain network access points. For example, in Figure 3.1(a), to visit facility j_1, a customer may pass through one of the three access points k_1, k_2, k_4 located in the boundaries. We introduce a set K of supporting stations to represent the network access points, and denote I and J as the sets of customers and candidate facility locations, respectively. As shown in Figure 3.1(b), the stations are connected to facilities to imply the real-world relationships between the facilities and the corresponding access points, i.e., a facility is connected to a station if the facility can be reached by customers through the corresponding access point. For example, in Figure 3.1(b), station k_1 is connected to facilities j_1 and j_2 as both j_1 and j_2 can be reached by passing through k_1. As such, a station $k \in K$ could be connected to multiple facilities, and a facility $j \in J$ could be connected to multiple stations as well. The original facility system is consequently augmented into an integrated facility-station structure. We use a binary parameter $l_{kj} = 1$ to indicate that facility j is connected to station k, or $l_{kj} = 0$ otherwise. Since the network access points are subject to possible disruptions, we further assume that each $k \in K$ is associated with a site-dependent failure probability q_k. The basic mechanism of the augmented structure is defined as follows: a facility remains operational if and only if at least one of its connected stations is functioning (i.e., at least one access point is available to be passed through to reach the facility). Hence, the operating states of the facilities are determined collectively by the states of all stations. For example, in Figure 3.1(b), j_1, j_2, j_3 are unreachable while j_4, j_5 are available if and only if stations $k_1, k_2, k_3, k_4, k_5, k_6$ are all disrupted while k_7 is functioning.

In the facility system, each customer $i \in I$ has a demand μ_i and each candidate location $j \in J$ is associated with a fixed setup cost f_j. Normally, each customer i visits its most preferred facility j for service by passing through station k (defined as an station-facility pair (k, j) in the rest of the paper). The transportation cost for station-facility pair (k, j) to serve one unit of demand from



(a) Facility network with access points

(b) Augmented facility-station structure

Figure 3.1: Conceptual illustration of the augmented facility-station structure.

customer i is denoted as d_{ikj}. Moreover, a penalty cost π_i per unit demand will be incurred if customer i do not receive any service. This situation occurs if no facility is reachable/available, or if the cost of serving customer i by the nearest available facility already exceeds π_i.

Since the access points are possible to be disrupted, each customer may choose multiple station-facility pairs as backups in case her preferable choices are unavailable, whereas each station-facility pair corresponds to one unique backup level. As such, we assume that each customer can select at most a number R of station-facility pairs and pass through the associated access points to visit the corresponding facilities for service, and a customer will pass through its level-r station if all its level-1, $\cdots$, level-$(r-1)$ stations have been disrupted. Note that a customer passes through a station at no more than one backup level, but may visit a facility through different stations at multiple backup levels. We further add a dummy *emergency* station-facility pair with indices $k = 0, j = 0$ to allow the "penalty assignment", i.e., when a customer losses service. Note that $l_{00} = 1$ and $q_0 = 0$, and the corresponding transportation cost is set to be the penalty cost, i.e., $d_{ikj}|_{k=0,j=0} = \pi_i, \forall i \in I$. Typically, a customer shall be assigned to the pair $(0,0)$ at level $R+1$ as long as regular station-facility pairs are available for backup levels $1, 2, \cdots, R$. However, as long as customer i can no longer receive service from any station-facility pair (k, j) at a unit cost less than π_i at any backup level $s \in \{1, 2, \cdots, R\}$, it will choose the emergency pair at level s.

3.2.2 Notation and Formulation

We first define several sets of decision variables. First, variables $\mathbf{X} := \{X_j\}_{j\in J}$ determine the facility locations as follows

$$X_j = \begin{cases} 1 & \text{if a facility is build at } j; \\ 0 & \text{otherwise.} \end{cases}$$

Next, the assignment of customers to station-facility pairs at multiple backup levels is specified by $\mathbf{Y} := \{Y_{ikjr}\}_{i\in I,k\in K\cup\{0\},j\in J\cup\{0\},r\in\{1,2,\cdots,R+1\}}$ where

$$Y_{ikjr} = \begin{cases} 1 & \text{if customer } i \text{ visits facility } j \text{ through station } k \text{ at backup level } r; \\ 0 & \text{otherwise.} \end{cases}$$

Finally, we define $\mathbf{Z} := \{Z_{ikjr}\}_{i\in I,k\in K\cup\{0\},j\in J\cup\{0\},r\in\{1,2,\cdots,R+1\}}$, where $Z_{ikjr} \in [0,1]$ denotes the probability for customer i to visit facility j through station k at backup level r. The reliable facility location problem (RFL) under the risk of network access failures is formulated into the following mixed-integer mathematical program:

$$\text{(RFL)} \quad \min \quad \sum_{j\in J} f_j X_j + \sum_{i\in I}\sum_{k\in K\cup\{0\}}\sum_{j\in J\cup\{0\}}\sum_{r=1}^{R+1} \mu_i d_{ikj} Z_{ikjr} Y_{ikjr} \tag{3.1a}$$

$$\text{s.t.} \quad \sum_{r=1}^{R} Y_{ikjr} \le X_j, \quad \forall i \in I, k \in K, j \in J, \tag{3.1b}$$

$$Y_{ikjr} \le l_{kj}, \quad \forall i \in I, k \in K\cup\{0\}, j \in J\cup\{0\}, r = 1,2,\cdots,R+1, \tag{3.1c}$$

$$\sum_{j\in J}\sum_{r=1}^{R} Y_{ikjr} \le 1, \quad \forall i \in I, k \in K, \tag{3.1d}$$

$$\sum_{k\in K}\sum_{j\in J} Y_{ikjr} + \sum_{s=1}^{r} Y_{i00s} = 1, \quad \forall i \in I, r = 1,2,\cdots,R+1, \tag{3.1e}$$

$$\sum_{r=1}^{R+1} Y_{i00r} = 1, \quad \forall i \in I, \tag{3.1f}$$

$$Z_{ikj1} = l_{kj}(1-q_k), \quad \forall i \in I, k \in K\cup\{0\}, j \in J\cup\{0\}, \tag{3.1g}$$

$$Z_{ikjr} = l_{kj}(1-q_k)\sum_{k'\in K}\sum_{j'\in J}\frac{q_{k'}}{1-q_{k'}} Z_{ik'j'(r-1)} Y_{ik'j'(r-1)}, \quad \forall i \in I, k \in K\cup\{0\}, j \in J\cup\{0\}, r = 2,3,\cdots,R+1, \tag{3.1h}$$

$$X_j, Y_{ikjr} \in \{0,1\}, \quad \forall i \in I, k \in K \cup \{0\}, j \in J \cup \{0\}, r = 1,2,\cdots,R+1. \tag{3.1i}$$

The objective function (3.1a) presents the expected system cost as the summation of the fixed facility cost, the expected total transportation cost, and the expected penalty cost. Constraints (3.1b) and (3.1c) require that customers can only visit open facilities through stations that are connected to the facilities. Constraints (3.1d) ensure that for each customer, any station is selected to be passed through at no more than one backup level. Constraints (3.1e) enforce that at each level r, any customer $i \in I$ is either assigned to a regular station-facility pair, or assigned to the dummy station-facility pair at an earlier level $s \leq r$, while constraints (3.1f) postulate that each customer is assigned to the dummy station-facility pair at a certain backup level $r \in \{1,2,\cdots,R+1\}$. Constraints (3.1g)–(3.1h) recursively define the assignment probabilities Z_{ikjr}: at level $r = 1$, the probability Z_{ikjr} is simply the probability for station k to function; at level $r > 1$, the probability Z_{ikjr} equals $(1-q_k)q_{k'}/(1-q_{k'})Z_{ik'j'r-1}$ if that customer i is assigned to j' through k' at level $r-1$. Constraints (3.1i) are integrality constraints.

The above formulation is nonlinear because of the nonlinear terms $\{Z_{ikjr}Y_{ikjr}\}$ contained in the objective function and constraints (3.1h). Therefore, we linearize each $Z_{ikjr}Y_{ikjr}$, which is a product of a bounded continuous variable and a binary variable, by applying the technique introduced by Sherali and Alameddine (1992). Specifically, a new continuous variable W_{ikjr} is introduced to equivalently replace $Z_{ikjr}Y_{ikjr}$ by enforcing four additional sets of constraints (i.e., (3.2d)-(3.2g)), and to transform (RFL) into the following mixed-integer linear program

$$\text{(LRFL)} \quad \min \quad \sum_{j\in J} f_j X_j + \sum_{i\in I}\sum_{k\in K\cup\{0\}}\sum_{j\in J\cup\{0\}}\sum_{r=1}^{R+1} \mu_i d_{ijk} W_{ikjr} \tag{3.2a}$$

$$\text{s.t.} \quad \text{(3.1b)} - \text{(3.1g)}, \tag{3.2b}$$

$$Z_{ikjr} = (1-q_k)\cdot \sum_{k'\in K}\sum_{j'\in J} \frac{q_{k'}}{1-q_{k'}} W_{ik'j'(r-1)'}$$

$$\forall i \in I, k \in K \cup \{0\}, j \in J \cup \{0\}, r = 2,3,\cdots,R+1, \tag{3.2c}$$

$$W_{ikjr} \leq Z_{ikjr} + 1 - Y_{ikjr}, \quad \forall i \in I, k \in K \cup \{0\}, j \in J \cup \{0\}, r = 1,2,\cdots,R+1, \tag{3.2d}$$

$$W_{ikjr} \geq Z_{ikjr} + Y_{ikjr} - 1, \quad \forall i \in I, k \in K \cup \{0\}, j \in J \cup \{0\}, r = 1, 2, \cdots, R+1, \tag{3.2e}$$

$$W_{ikjr} \leq Y_{ikjr}, \quad \forall i \in I, k \in K \cup \{0\}, j \in J \cup \{0\}, r = 1, 2, \cdots, R+1, \tag{3.2f}$$

$$W_{ikjr} \geq -Y_{ikjr}, \quad \forall i \in I, k \in K \cup \{0\}, j \in J \cup \{0\}, r = 1, 2, \cdots, R+1, \tag{3.2g}$$

$$X_j, Y_{ikjr} \in \{0, 1\}, \quad \forall i \in I, k \in K \cup \{0\}, j \in J \cup \{0\}, r = 1, 2, \cdots, R+1, \tag{3.2h}$$

$$0 \leq W_{ikjr} \leq 1, \quad \forall i \in I, k \in K \cup \{0\}, j \in J \cup \{0\}, r = 1, 2, \cdots, R+1. \tag{3.2i}$$

In theory, this mixed-integer linear program (LRFL) is compact and polynomial in size (which is much smaller compared to the exponential size of the scenario-based model), and thus could potentially be solved by commercial solvers such as CPLEX and Gurobi. However, considering the combinatorial nature of the problem as well as the difficulty associated with site-dependent probabilities, existing solvers generally take an excessively long computation time for even moderately sized instances, as we will show with numerical examples in Section 3.4. Therefore, in the next section, we develop customized solution approaches to efficiently solve (LRFL).

3.3 Solution Approach

3.3.1 Lagrangian Relaxation

We relax constraints (3.1b) in (LRFL) with Lagrangian multipliers $\{\lambda_{ikj}\}_{\forall i \in I, \forall k \in K, \forall j \in J}$ and move them as penalty terms, which yields the following objective function

$$\sum_{j \in J} \left(f_j - \sum_{i \in I} \sum_{k \in K} \lambda_{ikj} \right) X_j + \sum_{i \in I} \sum_{k \in K \cup \{0\}} \sum_{j \in J \cup \{0\}} \sum_{r=1}^{R+1} \mu_i d_{ikj} W_{ikjr} + \sum_{i \in I} \sum_{k \in K} \sum_{j \in J} \lambda_{ikj} \sum_{r=1}^{R} Y_{ikjr}.$$

With this relaxation, the original model (LRFL) is decomposed into two unrelated parts, involving the location and assignment variables $\mathbf{X}$ and $\mathbf{Y}$, respectively. The part involving $\mathbf{X}$,

$$\min_{X_j \in \{0,1\}, \forall j} \sum_{j \in J} \left(f_j - \sum_{i \in I} \sum_{k \in K} \lambda_{ikj} \right) X_j,$$

can be solved easily by simple inspection; i.e., given any $\{\lambda_{ikj}\}$,

$$X_j = \begin{cases} 1 & \text{if } f_j - \sum_{i\in I}\sum_{k\in K}\lambda_{ikj} < 0; \\ 0 & \text{otherwise.} \end{cases}$$

For the part involving $\mathbf{Y}$, we observe that it can be further separated into individual subproblems, one for each customer. The subproblem (RFL-SP$_i$) with respect to customer $i \in I$ is

$$\text{(RFL-SP}_i\text{)} \quad \min \quad \sum_{k\in K\cup\{0\}}\sum_{j\in J\cup\{0\}}\sum_{r=1}^{R+1}\mu_i d_{ikj} W_{kjr} + \sum_{k\in K}\sum_{j\in J}\lambda_{kj}\sum_{r=1}^{R} Y_{kjr} \tag{3.3a}$$

$$\text{s.t.} \quad Y_{kjr} \le l_{kj}, \quad \forall k \in K\cup\{0\}, j \in J\cup\{0\}, r = 1,2,\cdots,R+1, \tag{3.3b}$$

$$\sum_{j\in J}\sum_{r=1}^{R} Y_{kjr} \le 1, \quad \forall k \in K, \tag{3.3c}$$

$$\sum_{k\in K}\sum_{j\in J} Y_{kjr} + \sum_{s=1}^{r} Y_{00s} = 1, \quad \forall r = 1,2,\cdots,R+1, \tag{3.3d}$$

$$\sum_{r=1}^{R+1} Y_{00r} = 1, \tag{3.3e}$$

$$Z_{kj1} = l_{kj}(1-q_k), \quad \forall k \in K\cup\{0\}, j \in J\cup\{0\}, \tag{3.3f}$$

$$Z_{kjr} = l_{kj}(1-q_k)\sum_{k'\in K}\sum_{j'\in J}\frac{q_{k'}}{1-q_{k'}} W_{k'j'(r-1)}, \quad \forall k \in K\cup\{0\}, j \in J\cup\{0\}, r = 2,3,\cdots,R+1, \tag{3.3g}$$

$$(3.2\text{d}) - (3.2\text{g}), \tag{3.3h}$$

$$Y_{kjr} \in \{0,1\}, \quad \forall k \in K\cup\{0\}, j \in J\cup\{0\}, r = 1,2,\cdots,R+1, \tag{3.3i}$$

$$0 \le W_{kjr} \le 1, \quad \forall k \in K\cup\{0\}, j \in J\cup\{0\}, r = 1,2,\cdots,R+1. \tag{3.3j}$$

(RFL-SP$_i$) does not involve the relationships between customers, and is much smaller in size compared to the original (LRFL). Hence, it can be more efficiently handled by commercial solvers. However, since in the Lagrangian relaxation framework, each subproblem (RFL-SP$_i$) will be solved repeatedly across multiple iterations, we may still encounter heavy computational burden if relying on solvers. Thus, section 3.3.2 further develops an optional efficient approximate algorithm for the

subproblems.

The summation of the optimal objective values from the above two relaxed parts (which involve $\mathbf{X}$ and $\mathbf{Y}$ respectively) provides a lower bound to the original problem (LRFL). Further, based on the solutions to the relaxed subproblems, we use a simple heuristic to perturb them to obtain a feasible solution to the original problem, which provides an upper bound to (LRFL). Specifically, we fix the optimal facility location decisions from the first relaxed subproblem involving $\mathbf{X}$. Then for each customer i, we sort all built and connected access-facility pairs (i.e., $\{(k,j) : X_j = 1, l_{kj} = 1\}$) in ascending order of d_{ikj}. With the sorted sequence for each customer i, at every level r, we assign i to pair (k,j) with the smallest d_{ikj} as long as i has never been assigned to k at earlier levels $1, 2, \cdots, r-1$. The following proposition indicates that the feasible solution from this simple approach is likely to be near-optimum.

Proposition 1. *If $R = |K|$, then in any optimal solution $(\mathbf{X}, \mathbf{Y}, \mathbf{Z})$, a customer will be assigned to backup access-facility pairs based on the transportation costs; i.e., if $Y_{ik_1j_1r} = 1$ and $Y_{ik_2j_2(r+1)} = 1$ for some i, r, then $d_{ik_1j_1} \leq d_{ik_2j_2}$.*

Proof. Suppose, for a contradiction, that $(\mathbf{X}, \mathbf{Y}, \mathbf{Z})$ is optimal to (RFL) but there exist i, (k_1, j_1), (k_2, j_2) and r such that $Y_{ik_1j_1r} = 1$, $Y_{ik_2j_2(r+1)} = 1$ and $d_{ik_1j_1} > d_{ik_2j_2}$. We will show that by swapping (k_1, j_1) and (k_2, j_2) the objective of (RFL) will decrease. We construct a different solution $(\mathbf{X}', \mathbf{Y}', \mathbf{Z}')$ as follows:

(i) $X'_j = X_j$;

(ii) $$Y'_{hlms} = \begin{cases} 1, & \text{if } (h,l,m,s) = (i,k_1,j_1,r+1) \text{ or } (i,k_2,j_2,r); \\ 0, & \text{if } (h,l,m,s) = (i,k_1,j_1,r) \text{ or } (i,k_2,j_2,r+1); \\ Y_{hlms}, & \text{otherwise;} \end{cases}$$

(iii) $$Z'_{hlms} = \begin{cases} \frac{1-q_{k_2}}{1-q_{k_1}} Z_{ik_1j_1r}, & \text{if } (h,l,m,s) = (i,k_2,j_2,r); \\ q_{k_2} Z_{ik_1j_1r}, & \text{if } (h,l,m,s) = (i,k_1,j_1,r+1); \\ 0, & \text{if } (h,l,m,s) = (i,k_1,j_1,r) \text{ or } (i,k_2,j_2,r+1); \\ Z_{hlms}, & \text{otherwise.} \end{cases}$$

By construction, $(\mathbf{X}', \mathbf{Y}', \mathbf{Z}')$ is a feasible solution to (RFL). We use $\Phi(\mathbf{X}, \mathbf{Y}, \mathbf{Z})$ to denote the objective value of (RFL) associated with $(\mathbf{X}, \mathbf{Y}, \mathbf{Z})$, it follows that

$$\begin{aligned}
\Phi(\mathbf{X}, \mathbf{Y}, \mathbf{Z}) - \Phi(\mathbf{X}', \mathbf{Y}', \mathbf{Z}') &= \mu_i \left[d_{ik_1j_1} Z_{ik_1j_1r} + d_{ik_2j_2} Z_{ik_2j_2(r+1)} - \left(d_{ik_1j_1} Z'_{ik_1j_1(r+1)} + d_{ik_2j_2} Z'_{ik_2j_2r} \right) \right] \\
&= \mu_i \left[d_{ik_1j_1} Z_{ik_1j_1r} + d_{ik_2j_2} \frac{q_{k_1}\left(1 - q_{k_2}\right)}{1 - q_{k_1}} Z_{ik_1j_1r} \right. \\
&\qquad \left. - d_{ik_1j_1} q_{k_2} Z_{ik_1j_1r} - d_{ik_2j_2} \frac{1 - q_{k_2}}{1 - q_{k_1}} Z_{ik_1j_1r} \right] \\
&= \mu_i Z_{ik_1j_1r} \left(1 - q_{k_2}\right) \left(d_{ik_1j_1} - d_{ik_2j_2} \right)
\end{aligned}$$

Since $\mu_i Z_{ik_1j_1r}(1 - q_{k_2}) \geq 0$, $d_{ik_1j_1} > d_{ik_2j_2}$, we have $\Phi(\mathbf{X}, \mathbf{Y}, \mathbf{Z}) \geq \Phi(\mathbf{X}', \mathbf{Y}', \mathbf{Z}')$, which implies that $\Phi(\mathbf{X}, \mathbf{Y}, \mathbf{Z})$ is not optimal. This completes the proof. □

Hence, when $R = |K|$, given the facility location decisions, this heuristic yields an optimal customer assignment plan and a tight upper bound. In case $R < |K|$, it can only guarantee a feasible but not necessarily optimal assignment decisions. However, since the probabilities for large back-up levels to occur, i.e., the product of disruption probabilities of multiple access points, are often smaller by orders of magnitudes, the solution given by this sorting/greedy heuristic shall be quite close to the optimal solution.

In the remainder of the Lagrangian solution framework, we use standard subgradient techniques (Fisher, 2004) to update the multipliers λ; i.e.,

$$\lambda_{ikj}^{n+1} = \lambda_{ikj}^{n} + t_j^n \left(\sum_r Y_{ikjr}^n - X_j^n \right), \tag{3.4}$$

$$t_j^n = \frac{\xi_n \left(Z^* - Z_D(\lambda^n) \right)}{\| \sum_r Y_{ikjr}^n - X_j^n \|^2}, \tag{3.5}$$

where λ_{ikj}^n, t^n represent the lagrangian multiplier and step size in the nth iteration, respectively. ξ_n is a scalar, and Z^* and $Z_D(\lambda^n)$ are the best upper bound and the current lower bound, respectively.

Upon completing the lagrangian relaxation procedure, the above two bounds, especially the lower bound, may still not be close to optimum. If the Lagrangian relaxation algorithm fails to converge to a small enough gap in a certain number of iterations, we embed it into a branch-and-bound (B&B) framework to further reduce the gap. We construct a binary tree by branching on $\mathbf{X}$.

Specifically, among all unbranched variables, we select and branch on the one whose construction yields the least system cost. After building the branching tree, we run the Lagrangian relaxation algorithm at each node to determine the corresponding feasible solution and lower bound, and update them after finishing both child branches. While traversing the binary tree, depth-first search is found to perform slightly better than breadth-first or least-cost-first searches for small or moderate-sized instances (which are likely to be solved to optimality). However, if the instances are large, it is difficult to traverse the entire tree and completely close the gap. In such cases, least-cost-first search is preferable since it tends to yield a reasonably good lower bound before completely traversing the entire tree.

3.3.2 Approximate Solution to Subproblem

As we stated before, each (RFL-SP$_i$) is still a combinatorial problem with exponential complexity in the worst case. Furthermore, considering the large number of nodes we need to explore during the branch-and-bound process, even if we solve each subproblem (e.g., using commercial solvers) relatively quickly (e.g., 1-10s), it may take an excessively long time to complete the entire algorithm and find a good near-optimal solution. Therefore, in this section we develop an approximate algorithm which helps quickly find lower bounds to the relaxed subproblems.

Equations (3.3g) show the connections between Z_{kjr} and $Z_{kj(r-1)}, Y_{kj(r-1)}$, which brings difficulty in solving subproblem (RFL-SP$_i$). Therefore, instead of having Z_{kjr} directly in the formulation, we approximate them with fixed numbers. Let $k_1, k_2, \cdots, k_{|K|+1}$ be an ordering of the access points such that $q_{k_1} \leq q_{k_2} \leq \cdots \leq q_{k_{|K|+1}}$, note that $q_0 = 1$ and $k_{|K|+1} = 0$. We define two additional sets of numbers $\{\alpha_{kr}\}_{\forall k \in K, r \in \{1,2,\cdots,R+1\}}, \{\beta_r\}_{\forall r \in \{1,2,\cdots,R+1\}}$, such that

$$\alpha_{kr} = (1 - q_k) \prod_{l=1}^{r-1} q_{k_l}, \quad \beta_r = \prod_{l=1}^{r-1} q_{k_l}.$$

We next replace Z_{kjr} and Z_{00r} by their estimates α_{kr} and β_r, respectively. The relaxed subproblem (RFL-SP$_i$) is further relaxed into

$$\text{(RFL-RSP}_i\text{)} \quad \min \quad \sum_{k \in K} \sum_{j \in J} \sum_{r=1}^{R+1} \left(\mu_i d_{ikj} \alpha_{kr} + \lambda_{kjr} \right) Y_{kjr} + \sum_{r=1}^{R+1} \mu_i d_{i00} \beta_r Y_{00r} \tag{3.6a}$$

$$\text{s.t.} \quad Y_{kjr} \leq l_{kj}, \quad \forall k \in K \cup \{0\}, j \in J \cup \{0\}, r = 1, 2, \cdots, R+1, \tag{3.6b}$$

$$\sum_{j\in J}\sum_{r=1}^{R} Y_{kjr} \le 1, \quad \forall k \in K, \tag{3.6c}$$

$$\sum_{r=1}^{R+1} Y_{00r} = 1, \tag{3.6d}$$

$$\sum_{k\in K}\sum_{j\in J} Y_{kjr} + \sum_{s=1}^{r} Y_{00s} = 1, \quad \forall r = 1, 2, \cdots, R+1, \tag{3.6e}$$

$$Y_{kjr} \in \{0,1\}, \quad \forall k \in K \cup \{0\}, j \in J \cup \{0\}, r = 1, 2, \cdots, R+1. \tag{3.6f}$$

We observe that (RFL-RSP$_i$) is a combinatorial assignment problem, which can be solved by the Hungarian algorithm (as in Cui et al. (2010)). The following proposition states that the solution to (RFL-RSP$_i$) provides a lower bound to the relaxed subproblem (RFL-SP$_i$), and hence it can be embedded into the Lagrangian relaxation framework to facilitate computation speed, yet without affecting the validity of the resulted lower and upper bounds.

Proposition 2. *The solution to (RFL-RSP$_i$) provides a lower bound to the relaxed subproblem (RFL-SP$_i$).*

Proof. Let $\mathbf{Y}^*, \mathbf{Z}^*$ and $\mathbf{W}^*$ be the optimal solution to (RFL-SP$_i$). (RFL-RSP$_i$) can be built from (RFL-SP$_i$) by replacing Z^*_{kjr} and Z^*_{00r} with α^*_{kr} and β^*_r, respectively, and removing constraints (3.3f)–(3.3h) and (3.3j). Since we are relaxing contraints, the solution $\mathbf{Y}^*, \mathbf{Z}^*$ and $\mathbf{W}^*$ should still be feasible to (RFL-RSP$_i$), and based on the construction of α_{kr} and β_r, we know that $\alpha_{kr}Y_{kjr}$ and $\beta_r Y_{00r}$ are lower bounds of W_{kjr} and W_{00r}, respectively. Therefore, the optimal objective value of (RFL-RSP$_i$) is a lower bound to the optimal objective of (RFL-SP$_i$). This completes the proof. □

3.4 Numerical Examples

We apply the proposed model and solution algorithms to two examples so as to demonstrate their applicability and performance under different problem and parameter settings. The first example includes a series of hypothetical square grid networks with varying sizes. The second case focuses on planning railroad emergency response facility locations in the Chicago metropolitan area. The main purpose of this example is to illustrate the impacts of various system settings (e.g., heterogeneity) on the optimal design.[d]

[d] All input data for these case studies will be available at my webpage http://www.siyangxie.com.

The proposed solution algorithms are programmed in C++ and run on a 64-bit Intel i7-3770 computer with 3.40 GHz CPU and 8G RAM. The mixed-integer linear programs (LRFL) and (RFL-SP$_i$), if solved directly, are tackled by commercial solver CPLEX 12.4 using up to 4 threads. The reformulated problem (RFL-RSP$_i$) is solved by the Hungarian algorithm.

3.4.1 Hypothetical Grid Networks

For $n \in \{4, 5, 6, 7, 8, 10\}$, an $n \times n$ square grid network is generated to represent a hypothetical study region (e.g., a city like Venice) with n^2 cells (e.g., islands) and $2n(n-1)$ blockage segments (e.g., canal branches), as shown in Figure 3.2. We assume that the network corresponds to a coordinate system $(n_1, n_2) \in [1, 2, \cdots, n] \times [1, 2, \cdots, n]$, where n_1 increases from left to right, and n_2 increases from bottom to top; the bottom-left and top-right cells have coordinates $(1, 1)$ and (n, n), respectively. We further label the cell at location (n_1, n_2) with index $n_1 + n \times (n_2 - 1)$. The edge length between any two adjacent cells is set to 1, and each cell is considered to be both an individual customer and a candidate facility location. For cell $i = (n_1, n_2)$, the demand is $\bar{\mu}\left(1 + \tau_\mu \cos\left(\pi \frac{n_1 - 1}{n-1}\right)\right)$ and the fixed facility cost is $\bar{f}\left(1 + \tau_f \cos\left(\pi \frac{n_2 - 1}{n-1}\right)\right)$. The values of parameters τ_μ and τ_f determine the extent of heterogeneity of demand and facility cost distribution over the network, such that the customer demand density varies from $\bar{\mu}(1 + |\tau_\mu|)$ on the left side to $\bar{\mu}(1 - |\tau_\mu|)$ on the right side, and the facility cost varies from $\bar{f}(1 + |\tau_f|)$ on the bottom to $\bar{f}(1 - |\tau_f|)$ on the top. In these hypothetical test cases, we set their values to be $\bar{\mu} = 10.0, \bar{f} = 100.0, \tau_\mu = \tau_f = 0.25$. The middle point of each edge represents the access point (e.g., bridge) through which customers may travel to service facilities. The site-dependent failure probability of the edge between cells i and j are "randomly" generated as $0.015 + 0.005(\text{mod}(i + j, 5) + 1)$. The maximum assignment level is $R = 3$ for all cases.

To solve the reliable facility location problems in these networks, we use three solution approaches: (i) CPLEX directly applied to the linearized original problem (LRFL); (ii) Lagrangian relaxation based algorithm embedded in a branch-and-bound framework with each subproblem (RFL-SP$_i$) solved by CPLEX (LR+B&B+CPLEX); and (iii) Lagrangian relaxation based algorithm embedded in a branch-and-bound framework with each subproblem (RFL-RSP$_i$) solved by the approximate algorithm (LR+B&B+Approx.). The solution time limit is set to be 3600 seconds. Table 3.1 summarizes and compares the results obtained by the three approaches for a range of

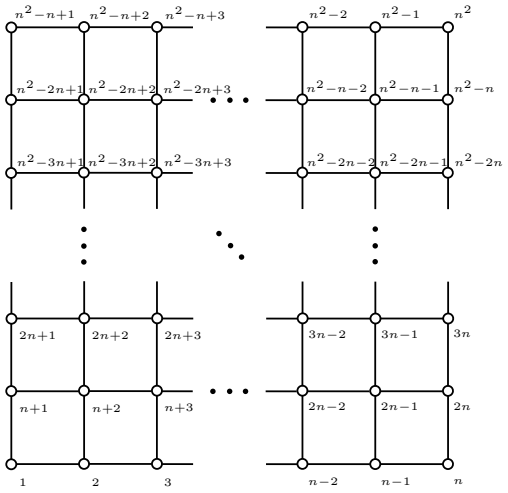


Figure 3.2: $n \times n$ hypothetical grid network.

test instances.

Table 3.1: Algorithm performance comparison.

	Network size	Number of facilities	Opt. facility locations	Final UB	Final LB	Final gap (%)	CPU time (s)
CPLEX	4×4	1	10	402.09	398.89	0.80	3600
	5×5	3	7, 20, 22	635.32	629.22	0.96	3600
	6×6	3	8, 23, 26	889.87	880.80	1.02	3600
	7×7	–	–	–	–	fail	3600
	8×8	–	–	–	–	fail	3600
	10×10	–	–	–	–	fail	3600
LR+B&B +CPLEX	4×4	1	10	402.09	402.09	0.0	2371
	5×5	2	7, 18	638.24	494.32	22.55	3600
	6×6	4	3, 23, 25, 34	934.74	598.01	36.02	3600
	7×7	5	2, 13, 32, 36, 42	1282.54	728.36	43.21	3600
	8×8	–	–	–	–	fail	3600
	10×10	–	–	–	–	fail	3600
LR+B&B +Approx.	4×4	1	10	402.09	402.09	0.0	1
	5×5	3	7, 20, 22	635.32	635.32	0.0	11
	6×6	3	8, 23, 26	889.87	889.87	0.0	65
	7×7	4	13, 16, 37, 40	1215.84	1215.84	0.0	2624
	8×8	6	4, 23, 26, 44, 50, 54	1603.98	1474.38	8.08	3600
	10×10	9	17,22,45,59,61,77,82,89,94	2524.32	2133.35	15.49	3600
	10×10	9	12,16,39,42,55,67,72,89,93	2506.72	2179.06	13.07	43200

Overall, it can be observed that the solution time and solution quality deteriorate with the network size, owing probably to the significant increase in the number of integer variables $\mathbf{Y}$. CPLEX cannot close the optimality gaps for the first three cases despite the relatively small network sizes. For the three larger networks, CPLEX ran out of memory and failed to provide even a

feasible solution. The LR+B&B+CPLEX approach can provide a feasible solution within one hour for the first 4 cases, however, the optimality gaps are relatively large except for the smallest 4×4 network. This is because when network size is large, it takes CPLEX a long time to solve even one instance of subproblem (RFL-SP$_i$), and thus the overall algorithm can only branch on a very limited number of nodes within the time limit. For the 10×10 network, the LR+B&B+CPLEX approach failed to give a feasible solution. In contrast, the LR+B&B+Approx. solution approach can obtain exact optimal solutions within 1 hour for the first 4 cases. For $n \in \{8, 10\}$, the optimality gaps after 1 hour of computation are 8.08%, and 15.49%, respectively. As such, the proposed LR+B&B+Approx. approach clearly outperforms the other two methods in terms of both solution quality and computation time.

To further check the quality of the solutions and explore the effectiveness of our solution approach, we run the case with network size 10×10 for another 11 hours. The final gap reduces from 15.49% (after 1 hour) to 13.07% (after 12 hours), and the objective value of the best known solution decreases slightly by 0.7%, from 2524.32 to 2506.72. We suspect that the best known feasible solution after 1 hour of computation is already of a reasonably good quality.

3.4.2 Railroad Emergency Response

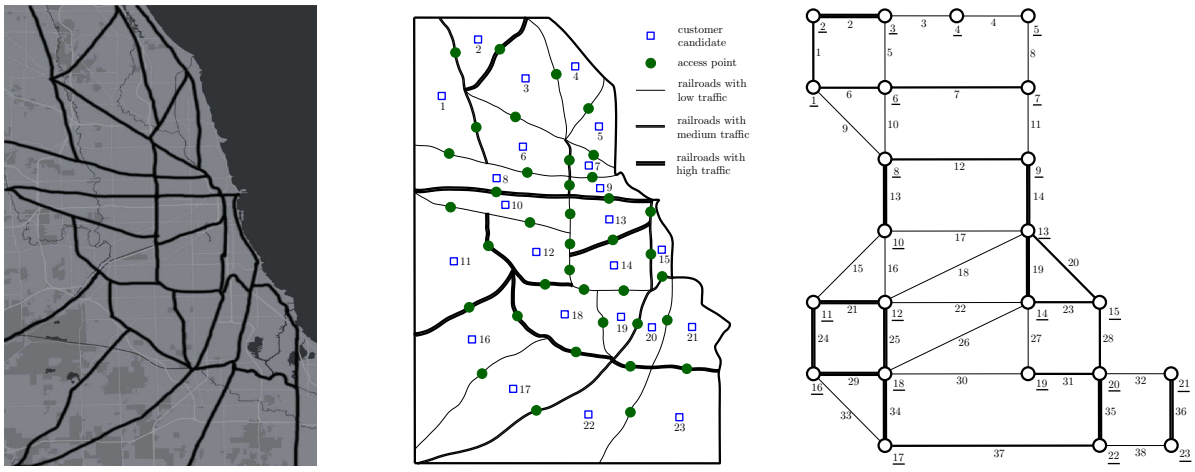
We now consider the Chicago area, a region with strong railroad network presence as shown in Figure 3.3(a). Target areas (e.g., towns and districts) are partitioned and surrounded by railroad segments. We assume there is one major access point (at-grade crossing)[e] on each railroad segment that allow first-responders to reach the regions. A limited number of emergency resource facilities are to be deployed among these regions in anticipation of random emergencies (e.g., fire, incidents). However, the rail crossings are subject to blockages, and hence the emergency resources of a facility might not be accessible if all of its surrounding rail crossings are blocked. The railroad network in Chicago contains $|J| = 23$ candidate facility locations, $|I| = 23$ customers (e.g., cities and towns), and $|K| = 38$ railway-highway crossings. In Figure 3.3(b), each line represents a railway segment, each dot represents a railway crossing (access point), and each square (surrounded by multiple railway

[e]Each segment may actually include multiple access points. Since a segment cannot be passed through if and only if all access points on it are disrupted, we can approximately consolidate these access points into one "representative" with the "composite" disruption probability $q = \prod_{k \in L} q_k$, where L is the set of all actual access points on this segment. Note that when a segment is long, the distances of passing through these different access points will likely be different. This issue is not addressed in this paper but deserves further study.

segments) represents a target demand area as well as a candidate facility location. The demand and fixed cost of each area are set to be proportional to its population and housing price, respectively. The railway crossings serve as access points to built facilities, and they are categorized into three groups, each having a high, median, or low risk of being blocked based on their annual train traffic volumes (denoted by $\mathcal{K}_h, \mathcal{K}_m, \mathcal{K}_l$, respectively). We further assume that blockage probability of each group can be specified as follows:

$$q_k = \begin{cases} \bar{q} + \hat{q} & \text{if } k \in \mathcal{K}_h \\ \bar{q} & \text{if } k \in \mathcal{K}_m \\ \bar{q} - \hat{q} & \text{if } k \in \mathcal{K}_l \end{cases}$$

where $\bar{q}$ is the average probability and $\hat{q}$ marks the level of spatial variation. A simplified graph of the network is shown in Figure 3.3(c), where each node is a candidate facility location and each link is a railway crossing. The distance d_{ikj} is calculated as the shortest path distance between node i and j through link k based on Figure 3.3(c). We assume that a customer receiving service from a facility elsewhere must pass through one of the railway crossings surrounding the facility.



(a) The Chicago map (b) The railroad network (c) The abstract representation

Figure 3.3: The railroads network setup in Chicago area.

We test our model with a range of $\bar{q}, \hat{q}$, and R, so as to examine their impacts on the optimal facility location design and algorithm performances. Case 10 (i.e., $\bar{q} = \hat{q} = 0$) represents the degenerated situation where crossings never get blocked, and hence backup assignments are not

necessary (i.e. $R = 1$). For other cases, we assume that $\bar{q} = 0.2$, and $\hat{q} \in \{0, 0.1, 0.2\}$ for identical probability, slight site-dependent probabilities, and high site-dependent probabilities, respectively. The value of R varies from 2 to 4.

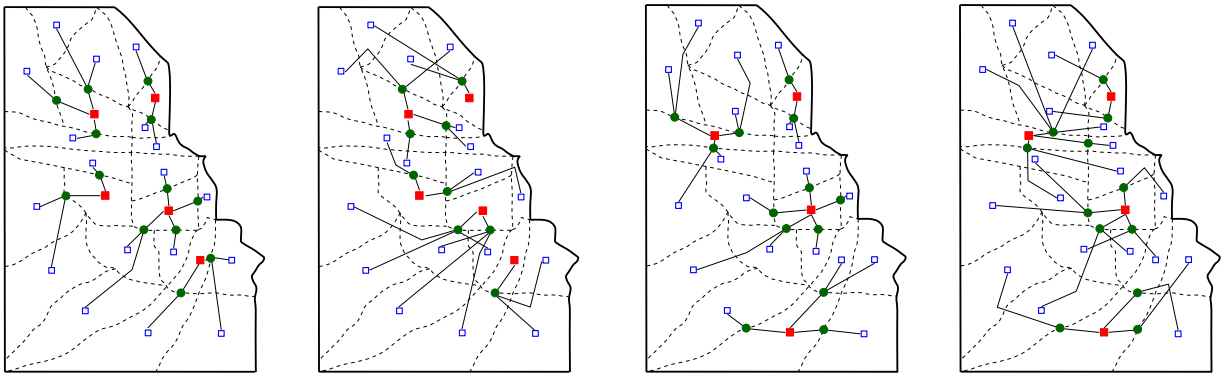
Solutions from our approximate algorithm (LR+B&B+Approx.) are presented in Table 3.2. The relatively large values of access point failure probabilities have led to longer computation times, i.e., not all cases can be solved to optimality within 2 hours. As R increases, the total cost decreases, due to a slightly lower likelihood for the customers to receive the penalty of losing service. In addition, the value of R does have observable impacts on the computation time and the optimal facility location design. These observations are consistent with those in earlier studies by Cui et al. (2010); Li and Ouyang (2012). Existence of access point failures generally has a noticeable impact on the optimal facility locations, total cost (including transportation cost), and the required computation time, if we use the no-failure counterpart (case 10) as the benchmark. Moreover, the spatial heterogeneity (as reflected by the value of $\hat{q}$) is possible to affect the optimal design, e.g., solutions to case 3 (with $\hat{q} = 0.2$), case 6 (with $\hat{q} = 0.1$), and case 9 (with $\hat{q} = 0$) are all quite different. It is worth noting that failing to consider site-dependent probabilities may lead to a cost increase, especially for transportation. For example, if we hold the failure probabilities of case 3 as the ground truth (where $\hat{q} = 0.2$), but solve the problem as if $\hat{q} = 0$. The corresponding solution (the one for case 9) will yield an actual total cost of 68596.4 and a transportation cost of 45636.4 under the assumed ground truth, which are 8.91% and 33.03% larger than the corresponding total cost of 62896.0 and transportation cost of 34566.0 obtained for case 3 (with $\hat{q} = 0.2$), respectively. It shall be also noted, however, that for many other cases, the cost "error" from ignoring access point failure heterogeneity is not as high as those observed in other studies (which directly consider facility failure heterogeneity). This result is somewhat intuitive because the presence of shared access points among the facilities tends to serve as another layer of "buffer" that averages out the spatial heterogeneity.

Figure 3.4 presents the location decisions and assignment path of each customer to access the facilities at each backup level (i.e., 1^{st} and 2^{nd}) for cases 3 and 9 (with $R = 2$). Generally, five facilities $\{5, 6, 12, 14, 20\}$ are built in case 3, as marked by the solid red squares, while another 4 facilities $\{5, 8, 14, 22\}$ are built in case 9. For both cases, the built facilities are located at regions with a higher concentration of demands. Specifically, in the southern half of the metropolitan area,

Table 3.2: Algorithm performance comparison.

Index	Prob. $\bar{q}$	Prob. $\hat{q}$	R	Opt. facility locations	Root UB	Root gap (%)	Final Obj.	Total Cost difference	Trans. cost difference	Final gap (%)	Time (s)
1	0.2	0.2	4	5, 6, 10, 14, 22	61376.0	46.0	60282.3	0.81%	3.57%	0	7054
2	0.2	0.2	3	5, 6, 10, 14, 22	61594.4	46.2	60881.2	1.33%	4.50%	2.91	7200
3	0.2	0.2	2	5, 6, 12, 14, 20	64719.2	48.0	62986.0	8.91%	33.03%	6.56	6936
4	0.2	0.1	4	5, 6, 10, 14, 22	61906.3	45.9	61377.6	0.16%	2.26%	0	7200
5	0.2	0.1	3	5, 6, 10, 14, 22	64663.3	48.2	62629.9	0.51%	2.83%	2.50	7200
6	0.2	0.1	2	5, 6, 12, 19, 20	70293.7	51.8	70115.3	3.54%	18.76%	13.07	7200
7	0.2	0	4	3, 7, 10, 14, 22	62690.3	46.5	62355.2	–	–	4.84	7200
8	0.2	0	3	3, 7, 10, 14, 22	65100.8	48.5	64454.7	–	–	8.40	7200
9	0.2	0	2	5, 8, 14, 22	76727.2	55.8	75635.2	–	–	22.35	7200
10	0	0	1	6, 14, 21	57120.0	36.2	57120.0	–	–	0	89

due to low demand, only one candidate location (e.g., 20 in case 3 and 22 in case 9) is selected. In contrast, the densely populated northern half always has two or even more built facilities. Moreover, regions/nodes with more access points (e.g., location 14) are more likely to be chosen since they can provide more backup access points/opportunities. As for the customer assignments, the 1$^{\text{st}}$-level assignments can be segregated into multiple groups, with each group clustered around one facility, while the 2$^{\text{nd}}$-level assignments are more intertwined with each other. In addition, a customer may visit the same facility or two different facilities through two different crossings at the two backup levels. For example, in the solution for case 3, customer 4 is first assigned to facility 5 through crossing 4 at level 1, then to facility 6 through crossing 5 at level 2; while customer 1 is first assigned to facility 6 through crossing 6 at level 1, then to facility 6 again through crossing 5 at level 2.



(a) 1$^{\text{st}}$ level assignment for case 3

(b) 2$^{\text{nd}}$ level assignment for case 3

(c) 1$^{\text{st}}$ level assignment for case 9

(d) 2$^{\text{nd}}$ level assignment for case 9

Figure 3.4: Facility locations and customer assignments at different backup levels of cases 3 and 9.

CHAPTER 4:

DECOMPOSITION OF FACILITY CORRELATIONS VIA AUGMENTATION OF VIRTUAL SUPPORTING STATIONS

In the previous chapter, the correlations of facility disruptions are caused by the failures of shared network access points, which can be intuitively represented as additional supporting stations (Section 3.2). However, facility disruptions could also be correlated when facilities are exposed to shared hazards. For example, facilities in a local geographical region are prone to simultaneous damage by a natural disaster (e.g., earthquake, hurricane, flooding). If one facility is known to have been disrupted, its neighboring facilities will bear a higher likelihood of being disrupted as well. To describe the correlations caused by shared hazards, since there exists no physical station, a straightforward modeling approach would need to enumerate or simulate an exponential number of scenarios, which is extremely computationally challenging.[a]

In this chapter, we develop a systematic methodological framework to describe and analyze correlated facility disruptions in succinct mathematical models. The framework proposes probabilistic representations of correlated facility disruptions, develops a decomposition scheme to augment the original customer-facility system into a customer-facility-station system with supporting stations (which are virtual) that experience only independent failures to capture the correlations among facilities. The added virtual supporting stations enable us to apply the optimization scheme developed in Chapter 3 to the new customer-facility-station system, which will be presented in Chapter 5. It is also proved that the decomposition scheme largely reduces the complexity associated with

[a]This chapter is based on a published paper, Xie et al. (2015).

system evaluation. We prove analytical properties of the decomposition scheme, and illustrate the proposed methodological framework using a set of numerical case studies and sensitivity analyses.

4.1 Introduction

Facility disruptions could be correlated when facilities are exposed to shared hazards or mutual interactions. For example, facilities in a local geographical region are prone to simultaneous damage by a natural disaster (e.g., earthquake, hurricane, flooding). If one facility is known to have been disrupted by an earthquake, its neighboring facilities will bear a higher likelihood of being disrupted as well – this shows a positive correlation pattern. In another context, suppose multiple facilities along a river are all threatened by flooding. If one facility is known to have been disrupted by flooding, then its downstream peers become less likely to be disrupted due to the release of water pressure. Similar correlations may also exist when facilities compete for scarce resources.

When facility disruptions exhibit correlations caused by the shared hazards, there exists no physical station, and thus a straightforward modeling approach would need to enumerate or simulate an exponential number of scenarios; this makes it computationally difficult to even just evaluate the system performance. To the best of our knowledge, in the supply chain context, only very limited efforts have been made to address certain special types of disruption correlations in facility location design (Li and Ouyang, 2010; Liberatore et al., 2012; Li et al., 2013). In other contexts such as computer networks (Bakkaloglu et al., 2002) and biometrics (Griffiths, 1973), researchers have attempted to use beta-binomial models to describe correlated disruptions. Such models however are only applicable to positive correlated disruptions that are spatially symmetric and homogeneous. Further, these models may not be capable of explicitly capturing the underlying disruption causalities or interdependencies among facilities. There remains a lack of a systematic methodology framework that can model general facility disruption correlations (i.e., including both positive and general correlations) in a computationally-tractable way.

Therefore, in this chapter, we develop a systematic station structure methodology to decompose facility disruption correlations, and make the following unique methodological contributions. First, we define three commonly-used representations of generally correlated facility disruption profiles (i.e., with scenario probabilities, marginal probabilities and conditional probabilities) and derive transformations between them. These transformations unify different representations that have

appeared in past studies (Liu et al., 2009; Lu et al., 2015; Li and Ouyang, 2010) and theoretically prove their equivalence. Second, we provide detailed formulas to transform these probabilistic correlation profiles into an adapted supporting station structure. This enables us to essentially decompose any correlated facility disruptions into a compact network structure that can be efficiently modeled with only independent failures, which in turn allows us to avoid enumerating an exponential number of disruption scenarios in evaluating system performance. Third, the number of needed supporting stations in this new framework is drastically reduced, and so is the formulation complexity, by allowing site-dependent station failure probabilities. It is shown in next chapter that the new station structure can be efficiently integrated into a design framework for optimal location decisions.

The remainder of this chapter is organized as follows. Section 4.2 introduces three probabilistic representations of facility disruption profile as well as the station structure representation, Section 4.3 presents the main decomposition scheme from a probabilistic disruption profile to the supporting station structure representation. Section 4.4 discusses some properties and miscellaneous issues associated with the decomposition framework. Section 4.5 presents several illustrative numerical experiments and sensitivity analyses, and draws managerial insights.

4.2 Facility Disruption Representations

This section proposes various succinct representations of correlated facility disruptions. Section 4.2.1 describes three commonly used probabilistic representations of a general facility disruption profile, and the pairwise transformations between them. Section 4.2.2 introduces a new supporting station structure representation, and its operational rules and properties.

4.2.1 Probabilistic Representations

For a given set of candidate facility locations $\mathcal{J}$, the state of each facility built at $j \in \mathcal{J}$ can be either functioning or disrupted, and the state of the entire facility set $\mathcal{J}$ is specified by the functioning states of all facilities. We call a specific underlying disruption pattern of all facilities $\mathcal{J}$ a disruption profile, which can be described in three equivalent ways as follows.

We define each unique realization of facilities' functioning states as a disruption scenario ω. A scenario ω occurs with probability p_ω, and a parameter $\delta_{j\omega}$ is used to indicate whether facility j

is functioning in scenario ω, $\delta_{j\omega} = 1$ if j is functioning, or 0 otherwise. Moreover, each scenario can be equivalently specified by the set of all disrupted facilities J. Let $p_J^S = p_\omega$ if ω satisfies $\delta_{j\omega} = 0, \forall j \in J, \delta_{j\omega} = 1, \forall j \in \mathcal{J}\backslash J$, which denotes the probability for all locations in J to be disrupted while all others are functioning. Any arbitrary disruption profile can be specified by the set $\mathcal{S} = \{p_J^S\}_{\forall J \subseteq \mathcal{J}}$, where $p_J^S \geq 0, \forall J \subseteq \mathcal{J}$ and $\sum_{J\subseteq\mathcal{J}} p_J^S = 1$. We call set $\{p_J^S\}_{\forall J\subseteq\mathcal{J}}$ a *scenario representation*, of the underlying disruption profile for convenience, which apparently includes a total of $2^{|\mathcal{J}|}$ elements.

A disruption profile can also be specified by marginal probabilities for facilities in set J to be disrupted, regardless of the states of all other facilities; i.e., $p_J^M = \sum_{\omega:\delta_{j\omega}=0,\forall j\in\mathcal{J}} p_\omega$. Then set $\mathcal{M} = \{p_J^M\}_{J\subseteq\mathcal{J}}$, where $p_\emptyset^M = 1$ and $p_{J_1}^M \geq p_{J_2}^M, \forall J_1 \subseteq J_2 \subseteq \mathcal{J}$, specifies an arbitrary disruption profile, which we call a *marginal representation*. Note that it also includes $2^{|\mathcal{J}|}$ elements.

Similarly, a disruption profile can be also represented by conditional disruption probabilities; i.e., $p_{j|J}^C = p_{\{j\}\cup J}^M / p_J^M$, which is the probability for facility $j \in \mathcal{J}$ to be disrupted given that all facilities in set $J \subseteq \mathcal{J}\backslash\{j\}$ have been disrupted. We call the collection of all conditional probabilities, i.e., $\mathcal{C} = \{p_{j|J}^C\}_{\forall J\subseteq\mathcal{J}, j\notin J}$, a *conditional representation*. Note that it includes $|\mathcal{J}| \cdot 2^{|\mathcal{J}|-1}$ elements.

It is well-known that these three disruption profile representations can be transformed equivalently from one another. Obviously, by definition,

$$p_J^M = \sum_{J_1: J\subseteq J_1} p_{J_1}^S, \ \forall J \subseteq \mathcal{J}, \tag{4.1}$$

$$p_{j|J}^C = \frac{p_{j\cup J}^M}{p_J^M} = \frac{\sum_{J_1: J\cup\{j\}\subseteq J_1} p_{J_1}^S}{\sum_{J_2: J\subseteq J_2} p_{J_2}^S}, \ \forall j \notin J \subseteq \mathcal{J}. \tag{4.2}$$

Next, from the Chain Rule of conditional probability (Russell and Norvig, 2009), we have,

$$p_J^M = \prod_{i=1}^{|J|} p_{j_i|\{j_1,\cdots,j_{i-1}\}}^C, \quad \forall J := \{j_1, \cdots, j_{|J|}\} \subseteq \mathcal{J}. \tag{4.3}$$

Finally, based on the Inclusion-Exclusion Principle Brualdi (2004), we conclude that,

$$p_J^S = \sum_{J_1: J\subseteq J_1} (-1)^{|J_1|-|J|} p_{J_1}^M, \quad \forall J \subseteq \mathcal{J}. \tag{4.4}$$

Proof. Let $\mathcal{P}(\mathcal{J})$ be the set of all subsets of $\mathcal{J}$, and $(\mathcal{P}(\mathcal{J}), \subseteq)$ be a partially ordered set of all

subsets of $\mathcal{J}$ that is partially ordered by containment with the smallest element $\emptyset$. The *Möbius function* μ of $(\mathcal{P}(\mathcal{J}, \subseteq))$ is defined as

$$\mu(J_1, J_2) = (-1)^{|J_2|-|J_1|}, \ \forall J_1 \subseteq J_2 \subseteq \mathcal{J}.$$

We define function $F : \mathcal{P}(\mathcal{J}) \to \mathbb{R}$ as $F(J) = p_J^S, \forall J \subseteq \mathcal{J}$, and $G : \mathcal{P}(\mathcal{J}) \to \mathbb{R}$ as $G(J) = p_J^M, \forall J \subseteq \mathcal{J}$. Based on the transformation equation (4.1), G can be expressed in terms of F as $G(J) = \sum_{J_1 : J_1 \subseteq J} F(J_1), \ \forall J \subseteq \mathcal{J}$. Theorem 6.6.1 in Brualdi (2004) directly implies that the above formula can be inverted to recover F from G as $F(J) = \sum_{J_1 : J_1 \subseteq J} \mu(J_1, J) G(J_1), \ \forall J \subseteq \mathcal{J}$ and further substituting $F(J) = p_J^S$ and $G(J) = p_J^M$ yields the following transformation

$$p_J^S = \sum_{J_1 : J \subseteq J_1} (-1)^{|J_1|-|J|} p_{J_1}^M, \ \forall J \subseteq \mathcal{J},$$

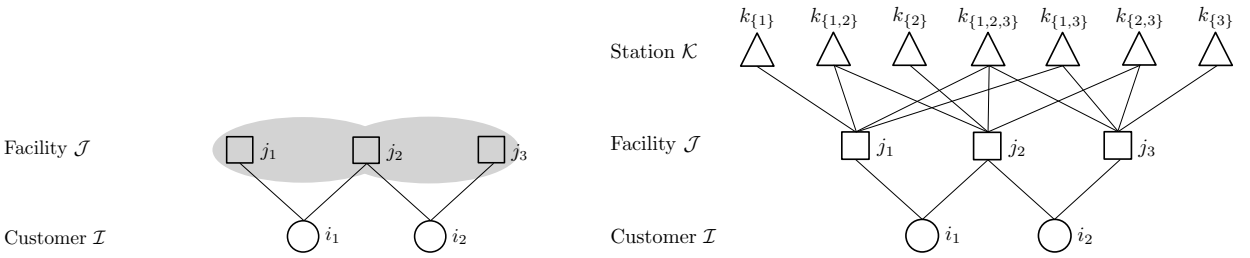
which completes the proof. □

4.2.2 Station Structure Representation

The idea of virtual supporting station is first introduced in Li et al. (2013). A set of virtual stations, $\mathcal{K}$, are added and connected to facilities in $\mathcal{J}$, as shown in Figure 4.1(b). A station $k \in \mathcal{K}$ could be connected to multiple facilities, and we use a binary parameter $l_{kj} = 1$ to indicate that facility j is connected to station k, 0 $l_{kj} = 0$ otherwise. We further assume that each station $k \in \mathcal{K}$ is associated with a non-negative site-dependent disruption quasi-probability $q_k \in [0, \infty)$[b]. The larger this quasi-probability value, the more damaging that station is to its connected facilities. Each sta- tion itself is also in a binary state: functioning or disrupted. By augmenting the original facility system with the additional virtual supporting stations, a station can be specified by the set of facilities connected to it; i.e., let k_J denote the station connected to all facilities in J but no other facilities, i.e., $l_{k_J j} = 1, \forall j \in J$ and $l_{k_J j} = 0, \forall j \notin J$.

With the additional layer of virtual supporting stations, correlations among facilities could be equivalently captured by the stations $\mathcal{K}$ with purely independent failures. The basic mechanism

[b]In theoretical physics and especially quantum mechanics, a similar concept is sometimes called "quasi-probability" (e.g., Dirac (1942); Feynman (1987)), where "conditional probabilities and probabilities of imagined intermediary states may be negative in a calculation of probabilities of physical events or states. ... The other possibility is that the situation for which the probability appears to be negative is not one that can be verified directly"(Feynman, 1987).



(a) Correlated facility disruption (b) Augmented customer-facility-station structure

Figure 4.1: Conceptual illustration of the station structure.

of the augmented facility-station system is defined as follows: *a facility remains operational if and only if at least one of its connected stations is functioning.* Hence the operating state of the facility system is determined collectively by the states of all stations. For example, in Figure 4.1(b), j_1 and j_2 are disrupted and j_3 is functioning if and only if stations $k_{\{1\}}, k_{\{2\}}, k_{\{1,2\}}, k_{\{1,3\}}, k_{\{2,3\}}, k_{\{1,2,3\}}$ are all disrupted and $k_{\{3\}}$ is functioning. Following this mechanism, we can ensure that the probability of a disruption scenario is equal to the product of q_k for each disrupted station k, and $(1-q_k)$ for each functioning station k. For example, in Figure 4.1(b), $p^S_{\{1,2\}} = q_{k_{\{1\}}} q_{k_{\{2\}}} q_{k_{\{1,2\}}} q_{k_{\{1,3\}}} q_{k_{\{2,3\}}} q_{k_{\{1,2,3\}}} (1 - q_{k_{\{3\}}})$.

For the set of all possible supporting stations $\mathcal{K}$, the corresponding probability formulation is $\{q_k\}_{\forall k \in \mathcal{K}}$, which we now call a *station representation*. In the worst case, the maximum number of supporting stations can be up to $|\{q_k\}_{\forall k \in \mathcal{K}}| = |\mathcal{K}| = 2^{|\mathcal{J}|}$. However, as we will show in Section 4.4.1, in most real-world cases, the number of necessary supporting stations in practice is likely to be quite small.

4.3 Decomposition of Correlations

4.3.1 Independent and Correlated Disruptions

We say that a disruption profile is independent if any subset of facilities $J \subseteq \mathcal{J}$ are independent; i.e., $p^M_J = p^M_{J_1} \cdot p^M_{J \setminus J_1}, \forall J_1 \subset J$. In this case, the three probabilistic representations can be simply expressed in terms of individual facility disruption probabilities $\{p_j\}_{\forall j \in \mathcal{J}}$, i.e.,

$$p^S_J = \prod_{j \in J} p_j \prod_{j \notin \bar{J}} (1 - p_j), \quad p^M_J = \prod_{j \in J} p_j, \quad p^C_{j|J} = \frac{\prod_{j \in j \cup J} p_j}{\prod_{j \in J} p_j}, \quad \forall J \subseteq \mathcal{J}, j \notin J. \tag{4.5}$$

More generally, a disruption profile may be correlated. Next, we define positive correlation for a disruption profile, $\{p_J^M\}_{\forall J \subseteq \mathcal{J}}$. We first define a set of cue fractions $\frac{p_J^M}{p_{J\setminus\{j\}}^M} \leq 1, \forall j \in J \subseteq \mathcal{J}$, and then consider a series of iterative operations. In the first iteration, we construct a new fraction with each of the cue fraction $\frac{p_J^M}{p_{J\setminus\{j\}}^M}$ as the denominator and $\frac{p_{J\setminus\{j'\}}^M}{p_{J\setminus\{j,j'\}}^M}$ as the numerator. The numerator is simply obtained by removing an arbitrary common subscript element $j' \in J, j' \neq j$ from every item in the original cue fraction. In each of the following iterations, we just take the resulting fractions from the previous iteration as cue fractions and repeat the same operation until the numerator of every resulting fraction contains an item of $p_\emptyset^M$. If every resulting fraction throughout the iterations is no greater than one we call the disruption profile positively correlated. Obviously, the disruptions are independent only if all of the resulting fractions at all iterations are equal to one. This definition is formally stated below.

Definition 1. *A set of facilities $J \subseteq \mathcal{J}$ are positively correlated if*

$$Q(J, J_1, j) := \frac{\prod_{J_1\setminus\{j\} \subseteq L \subseteq J\setminus\{j\}} (p_L^M)^{(-1)^{|L|-|J_1|}}}{\prod_{J_1 \subseteq L \subseteq J} (p_L^M)^{(-1)^{|L|-|J_1|+1}}} \leq 1, \quad \forall j \in J_1 \subseteq J \subseteq \mathcal{J}, \tag{4.6}$$

and $\exists j \in J_1 \subseteq J \subseteq \mathcal{J}$ such that $Q(J, J_1, j) < 1$. If facilities are correlated but not positively correlated, we say they are generally correlated. If all facilities in $\mathcal{J}$ are positively (or generally) correlated, we say the disruption profile is positively (or generally) correlated.

As a specific example, when $\mathcal{J}$ only has two facilities, say $\mathcal{J} := \{j_1, j_2\}$, then the above conditions are equivalent to $\frac{p_{\{j_1\}}^M / p_\emptyset^M}{p_{\mathcal{J}}^M / p_{\{j_2\}}^M} \leq 1$. This is obviously consistent with the classic definition of positively correlated disruptions $\frac{p_{j_1|\emptyset}^C}{p_{j_1|j_2}^C} = \frac{p_{\{j_1\}}^M p_{\{j_2\}}^M}{p_{\mathcal{J}}^M} < 1$, since $p_{\mathcal{J}}^M / p_{\{j_2\}}^M \leq 1$ by definition.

When facility disruptions are correlated, specifying any of the three probabilistic representations would typically require enumerating an exponential number of the representation elements. To circumvent this complexity, the following section describes how an arbitrary probabilistic disruption representation can be transformed into an equivalent station representation with only independent station failures.

4.3.2 Decomposition Scheme

Station structure with only independent station failures are much easier for analysis and design (Snyder, 2006; Chen et al., 2011). This section presents recipes for decomposing a facility system with an arbitrary correlated disruption profile (probabilistic representation) into an equivalent network with additional supporting stations (i.e., station representation).

Proposition 3. *For a given station representation* $\{q_{k_J}\}_{\forall J \subseteq \mathcal{J}}$*, the equivalent probabilistic disruption profile representations are formulated as:*

$$p_J^S = \sum_{J_1 : J \subseteq J_1} (-1)^{|J_1|-|J|} \left[\prod_{J_2 : J_2 \cap J_1 \neq \emptyset} q_{k_{J_2}} \right], \ \forall J \subseteq \mathcal{J}, \tag{4.7}$$

$$p_J^M = \prod_{J_1 : J_1 \cap J \neq \emptyset} q_{k_{J_1}}, \ \forall J \subseteq \mathcal{J}, \tag{4.8}$$

$$p_{j|J}^C = \prod_{J_1 \ni j, J_1 \subseteq \bar{J}} q_{k_{J_1}}, \ \forall J \subseteq \mathcal{J}, j \notin J. \tag{4.9}$$

Conversely, given a probabilistic disruption representations, we can construct an equivalent station representation by solving equations (4.7), (4.8) or (4.9). We prove in the following proposition that the station representation satisfying (4.7), (4.8) or (4.9) exists and is unique.

Proposition 4. *For any probabilistic disruption profile representation, there exists one and only one station representation* $\{q_{k_J}\}_{\forall J \subseteq \mathcal{J}}$ *that satisfies* (4.7)*,* (4.8) *or* (4.9)*.*

Proof. See section 4.6.1. □

Next, the following two propositions show how to construct a supporting station structure (station representation) from a probabilistic disruption profile representation $\{p_{j|J}^C\}_{\forall J \subseteq \mathcal{J}, j \notin J}$, $\{p_J^M\}_{\forall J \subseteq \mathcal{J}}$, or $\{p_J^S\}_{\forall J \subseteq \mathcal{J}}$.

Proposition 5. *An arbitrary conditional representation* $\{p_{j|J}^C\}_{\forall J \subseteq \mathcal{J}, j \notin J}$ *of a correlated disruption profile can be represented by a station representation* $\{q_{k_J}\}_{\forall J \subseteq \mathcal{J}}$*, where:*

$$q_{k_J} = \prod_{i=0}^{|J|-1} \left[\prod_{\substack{L : L \subseteq J \setminus \{j\} \\ |L| = i}} p_{j|(\bar{J} \cup L)}^C \right]^{(-1)^i} = \prod_{L : L \subseteq J \setminus \{j\}} \left[p_{j|(\bar{J} \cup L)}^C \right]^{(-1)^{|L|}}, \ j \in J, \ \forall J \subseteq \mathcal{J}. \tag{4.10}$$

Proof. See section 4.6.2. □

Proposition 6. *An arbitrary marginal representation* $\{M_J\}_{\forall J \subseteq \mathcal{J}}$ *or scenario representation* $\{S_J\}_{\forall J \subseteq \mathcal{J}}$ *of a correlated disruption profile can be transformed into a station representation* $\{P(K_J)\}_{\forall J \subseteq \mathcal{J}}$*, where:*

$$q_{k_J} = \prod_{i=0}^{|J|} \left[\prod_{\substack{L:\bar{J} \subseteq L \\ |L|=|\bar{J}|+i}} p_L^M \right]^{(-1)^{i-1}} = \prod_{L:\bar{J} \subseteq L \subseteq \mathcal{J}} \left[p_L^M \right]^{(-1)^{|L|-|\bar{J}|+1}} \tag{4.11}$$

$$= \prod_{L:\bar{J} \subseteq L \subseteq \mathcal{J}} \left[\sum_{J_1:L \subseteq J_1} p_{J_1}^S \right]^{(-1)^{|L|-|\bar{J}|+1}}, \ \forall J \subseteq \mathcal{J}. \tag{4.12}$$

Proof. See section 4.6.3. □

This decomposition scheme could be illustrated using the simple three-facility system in Figure 4.1. Suppose the scenario-based facility disruption probabilities are given as input shown in Table 4.1. Following (4.1) and (4.11), the marginal disruption probability for facility 1, $p_{\{1\}}^M$, and disruption quasi-probability for a station solely connected to this facility, $q_{k_{\{1\}}}$, are computed respectively as

$$p_{\{1\}}^M = p_{\{1\}}^S + p_{\{1,2\}}^S + p_{\{1,3\}}^S + p_{\{1,2,3\}}^S = 0.60,$$
$$q_{k_{\{1\}}} = \frac{p_{\{1,2,3\}}^M}{p_{\{2,3\}}^M} = \frac{0.30}{0.35} = 0.86.$$

In so doing, the entire marginal representation and associated station structure can be computed, and the results are summarized in Table 4.1.

Table 4.1: Different representations of the correlated disruption example.

Scenario representation	$p_{\{1\}}^S$ 0.05	$p_{\{2\}}^S$ 0.05	$p_{\{3\}}^S$ 0.05	$p_{\{1,2\}}^S$ 0.15	$p_{\{1,3\}}^S$ 0.10	$p_{\{2,3\}}^S$ 0.05	$p_{\{1,2,3\}}^S$ 0.30
Marginal representation	$p_{\{1\}}^M$ 0.60	$p_{\{2\}}^M$ 0.55	$p_{\{3\}}^M$ 0.50	$p_{\{1,2\}}^M$ 0.45	$p_{\{1,3\}}^M$ 0.40	$p_{\{2,3\}}^M$ 0.35	$p_{\{1,2,3\}}^M$ 0.30
Station structure	$q_{k_{\{1\}}}$ 0.86	$q_{k_{\{2\}}}$ 0.75	$q_{k_{\{3\}}}$ 0.67	$q_{k_{\{1,2\}}}$ 0.93	$q_{k_{\{1,3\}}}$ 0.95	$q_{k_{\{2,3\}}}$ 1.00	$q_{k_{\{1,2,3\}}}$ 0.79

Finally, we claim in the next proposition a property that validates the values of $\{q_k\}$ as proba-

bilities under positive correlation.

Proposition 7. *The values of q_k in a station structure representation satisfies $q_k \in [0,1], \forall k \in \mathcal{K}$ if and only if the disruption profile is positively correlated.*

Proof. See section 4.6.4. □

4.4 Discussions

This section illustrates some important properties and miscellaneous issues assoicated with the proposed supporting station structure.

4.4.1 Compactness of the Station Structure

One may wonder how many stations will be needed to represent a complex correlation profile. Recall that $q_{k_J} = 1$ indicates that the station corresponding to k_J fails surely and is unnecessary. Hence, we only focus on the subset of necessary stations that must be included in the station structure. Correspondingly, in the scenario representation $\mathcal{S} = \{p_J^S\}_{\forall J \subseteq \mathcal{J}}$, useful information is contained only in the subset of scenarios with positive probabilities, i.e., $\widetilde{\mathcal{J}} = \{J : p_J^S > 0, \forall J \subseteq \mathcal{J}\}$. This section analyzes the size of $\mathcal{K}$ and $\mathcal{S}$ and shows that $|\mathcal{K}|$ is very likely to be comparable to $|\mathcal{S}|$ ($|\mathcal{S}| = |\widetilde{\mathcal{J}}|$) in realistic settings.

First, we present a property of the supporting station structure in the following lemma.

Lemma 1. *For any facility subset $J \subseteq \mathcal{J}$, if $\exists j \in J$ such that $p_{\bar{J}}^M = p_{\bar{J} \cup \{j\}}^M$, then $q_{k_J} = 1$.*

Proof. See section 4.6.5. □

Lemma 1 leads to the following lemma that connects $|\mathcal{K}|$ to $|\mathcal{S}|$.

Lemma 2. *In a supporting station structure, $q_{k_J} \neq 1$ only if $\exists \tilde{J} \subseteq \widetilde{\mathcal{J}}$ such that $\bar{J} = \bigcap_{J_1 \in \tilde{J}} J_1$.*

Proof. See section 4.6.6. □

Lemma 2 indicates that if the station connected to J is necessary (i.e., $q_{k_J} \neq 1$), then there must be some effective scenarios in $\widetilde{\mathcal{J}}$ whose intersection is exactly the complement of J. Lemmas 1 and 2 show the condition under which a station set is necessary – this helps determine the total number of necessary stations $|\mathcal{K}|$. We denote $\left\{\bigcap_{J \in \tilde{J}} J\right\}_{\forall \tilde{J} \subseteq \widetilde{\mathcal{J}}}$ as the set of distinct intersections across any arbitrary subset of $\widetilde{\mathcal{J}}$, and develop the following theorem to further relate $|\mathcal{S}|$ to $|\mathcal{K}|$.

Theorem 1. *In a facility system with facility set* $\mathcal{J}$ *and effective disruption scenario set* $\widetilde{\mathcal{J}}$*, the maximum number of necessary supporting stations* $|\mathcal{K}|$ *is* $\min\left\{2^{|\mathcal{J}|}, \left|\left\{\bigcap_{J\in\tilde{J}} J\right\}_{\forall \tilde{J}\subseteq\widetilde{\mathcal{J}}}\right|\right\}$*. Specially,* $|\mathcal{K}| \leq |\mathcal{S}| + 1$ *if* $\left\{\bigcap_{J\in\tilde{J}} J\right\}_{\forall \tilde{J}\subseteq\widetilde{\mathcal{J}}} = \widetilde{\mathcal{J}} \bigcup \emptyset$*.*

Proof. Given facility set $\mathcal{J}$, on one hand, since a station can either be connected to a facility or not, the maximum possible number of distinct stations equals the number of subsets of $\mathcal{J}$ which is $2^{|\mathcal{J}|}$. On the other hand, Lemma 2 tells us that a station is effective (i.e., its failure propensity is unequal to 1) only if the complement of its connected facility set corresponds to an intersection of some effective scenarios in $\widetilde{\mathcal{J}}$, so the maximum number of stations is no more than $\left|\left\{\bigcap_{J\in\tilde{J}} J\right\}_{\forall \tilde{J}\subseteq\widetilde{\mathcal{J}}}\right|$. Consequently, the number of necessary supporting stations is bounded as $\min\left\{2^{|\mathcal{J}|}, \left|\left\{\bigcap_{J\in\tilde{J}} J\right\}_{\forall \tilde{J}\subseteq\widetilde{\mathcal{J}}}\right|\right\}$. □

Remark: Theorem 1 implies that when the facility system is globally correlated, the number of necessary stations is usually comparable to the number of scenario/marginal/conditional probabilities that are needed to describe the correlation and thus the station structure is compact from a modeling point of view. In real world cases, most intersections of effective scenarios in $\widetilde{\mathcal{J}}$ is likely to be contained within $\widetilde{\mathcal{J}} \bigcup \emptyset$ as well. Then the maximum number of distinct intersections of any effective scenarios is around $|\widetilde{\mathcal{J}}| + 1$. So according to theorem 1, the size of the resulting station structure, $|\mathcal{K}|$, is very likely to be compact.

When the facility disruptions are "locally" correlated, the number of stations $|\mathcal{K}|$ (generally larger than the number of facilities $|\mathcal{J}|$) shall be far smaller than the number of scenarios $|\mathcal{S}|$, and the difference between $|\mathcal{K}|$ and $|\mathcal{S}|$ grows sharply with the level of localness of the correlations (i.e., correlations being confined within a local area). In particular, if the facility system $\mathcal{J}$ could be partitioned into N mutually exclusive subsets $\{J_n\}_{n=1,2,\cdots,N}$, such that the facilities within each subset J_i are correlated with one other, while facilities in different subsets are independent, the maximum number of needed stations is $|\mathcal{K}| \leq \sum_{n=1}^{N} 2^{|J_n|}$, which is typically much smaller compared to the maximum number of scenarios that are used/needed to describe the correlation, $\mathcal{S} = 2^{|\mathcal{J}|}$. We state this obvious result in the following proposition without proof.

Proposition 8. *If* $\mathcal{J} = \cup_{n=1,2,\cdots,N} J_n$ *for some* $N > 1$*, such that for all* $i = 1, 2, \cdots, N$*, the disruptions of all facilities in* J_i *are independent of those in* $\mathcal{J}\backslash J_i$*, then the maximum number of needed stations*

$|\mathcal{K}|$ and the number of scenarios $|\mathcal{S}|$ satisfy $|\mathcal{K}| \leq \sum_{n=1}^{N} 2^{|J_n|}$ and $|\mathcal{S}| = 2^{\sum_{n=1}^{N}|J_n|}$, respectively, which further yields

$$\frac{|\mathcal{K}|}{|\mathcal{S}|} \leq \frac{\sum_{n=1}^{N} 2^{|J_n|}}{2^{\sum_{n=1}^{N}|J_n|}} \leq \begin{cases} 1, & \text{if } N = 1 \text{ (globally correlated);} \\ \frac{\sum_{n=1}^{N} 2^{|J_n|}}{2^{|\mathcal{J}|}}, & \text{if } 2 \leq N \leq |\mathcal{J}|/2 \text{ (locally correlated).} \end{cases}$$

As an example, we consider a facility system $\mathcal{J} = \cup_{n=1,2,\cdots,N} J_n$ where $J_i = \{3i-2, 3i-1, 3i\}$, and the disruptions of J_i and $\mathcal{J} \backslash J_i$ are independent. Each J_i has a system structure as shown in Figure 4.1(a), and is subject to the scenario disruption profile in Table 4.1. For this particular system $\mathcal{J}$, the total number of scenarios is $|\mathcal{S}| = \left(2^3\right)^N = 8^N$, while the number of stations is only $|\mathcal{K}| = 6N$, which is much smaller than $|\mathcal{S}|$. As such, the formulations we will present in the next chapter indicate that when $N = 4$, a scenario-based formulation would require at least $3N + 3N(3N+1) \cdot 8^N = 638988$ binary variables to describe the scenarios, while our proposed formulation will only need at most $3N + 3N(3N+1)(6N+1)6N = 93612$ binary variables.

Although Proposition 8 addresses the very special case where correlation among facilities can be divided into disjoint groups, similar relationships between $|\mathcal{S}|$ and $|\mathcal{K}|$ can be obtained for other cases. The following proposition, for example, shows that when the disruptions are positively correlated, no station is needed to connect any two independent facilities.

Proposition 9. *If the facility disruptions are positively correlated per Definition 1; i.e.,*

$$\prod_{L: J_1 \backslash \{j\} \subseteq L \subseteq J \backslash \{j\}} (p_L^M)^{(-1)^{|L|-|J_1|}} \leq \prod_{L: J_1 \subseteq L \subseteq J} (p_L^M)^{(-1)^{|L|-|J_1|+1}}, \forall j \in J_1 \subseteq J \subseteq \mathcal{J},$$

then for any two facilities j_1 and j_2 that are independent of each other; i.e., $p^M_{\{j_1\}} \cdot p^M_{\{j_2\}} = p^M_{\{j_1, j_2\}}$, we have $q_{k_J} = 1$ for any facility set J that contains both j_1 and j_2. That is, no station is connected to both j_1 and j_2.

Proof. See section 4.6.7. □

Proposition 9 is quite revealing; it implies that the number of stations shall be limited if the correlations are local and positive (which often occurs in the real world). For example, consider a chain of N facilities located sequentially in a line, such that any two adjacent facilities are

positively correlated while any non-adjacent facilities are mutually independent. In such a system, the maximum number of scenarios is $|\mathcal{S}| = 2^N$, while the maximum number of stations needed to express the correlation is only $|\mathcal{K}| = 2N - 1$, which is far smaller than $|\mathcal{S}|$ for all $N > 2$.

4.4.2 Identical Station Failure Quasi-Probability

Our framework allows supporting stations to have site-dependent failure quasi-probabilities so as to reduce the size of the station structure. This is very appealing. However, for location design models, it is sometimes convenient to have identical quasi-probabilities across stations (Snyder and Daskin, 2005). A station structure with identical quasi-probabilities can be constructed based on the following lemma.

Lemma 3. *Li et al. (2013) Given a station structure with site-dependent failure quasi-probabilities, all elements can be equivalently represented by powers of an identical constant $p > 0$, i.e.,*

$$q_{k_J} = p^{I_{k_J}},\ I_{k_J} \in \mathbb{R}. \tag{4.13}$$

Where I_{k_J} is the corresponding exponent of q_{k_J}. If we further limit the powers to be integers, these quasi-probabilities can be approximated arbitrarily accurately within error $\epsilon > 0$, i.e.,

$$q_{k_J} \in \left[p^{N_{k_J}} - \epsilon, p^{N_{k_J}} + \epsilon\right],\ N_{k_J} \in \mathbb{Z}. \tag{4.14}$$

Where N_{k_J} is the corresponding integer exponent of q_{k_J} under error ϵ.

The basic idea behind Lemma 3 is to split each original station into one or multiple new ones with identical failure quasi-probabilities. Hence, this transformation will undesirably increase the number of the necessary stations and introduce certain approximation errors.

The following example illustrates the effects of enforcing identical station failure probabilities. We consider a facility system $\mathcal{J} = \{1, 2, 3\}$, and arbitrarily generate facility disruption scenario representation as $p^S_{\{1\}} = 0.16, p^S_{\{2\}} = p^S_{\{3\}} = 0.12, p^S_{\{1,2\}} = p^S_{\{1,3\}} = 0.08, p^S_{\{2,3\}} = 0.07, p^S_{\{1,2,3\}} = 0.05$. After computing the station failure quasi-probabilities, we expand each station by introducing some identical stations with equal failure quasi-probabilities which are grouped together as a substitute of the orignal station. We further ensure that the difference between the approximated failure quasi-probability(multiplication of the quasi-probabilities of all grouped stations) and the accurate failure

probability of each station is no larger than a pre-set approximation error. The station structure with site-dependent station failure quasi-probabilities and no approximation (corresponds to the last row), and the ones with identical failure quasi-probabilities under different approximation error tolerances are presented in Table 4.2. The values 0.1, 0.01, and 0.001 in this table are different required approximation error tolerances (i.e., ϵ in (4.14)), and the associated rows list the results of the corresponding cases.

Table 4.2: Size and accuracy of station structures with identical failure quasi-probability.

	ϵ	$k_{\{1\}}$	$k_{\{2\}}$	$k_{\{3\}}$	$k_{\{1,2\}}$	$k_{\{1,3\}}$	$k_{\{2,3\}}$	$k_{\{1,2,3\}}$
Number of stations	0.1	25	25	25	3	3	5	4
	0.01	30	33	33	1	1	3	2
	0.001	653	712	712	19	19	68	51
Approx failure prob.	0.1	0.4167	0.4167	0.4167	0.9003	0.9003	0.8394	0.8693
	0.01	0.4167	0.3817	0.3817	0.9712	0.9712	0.9162	0.9433
	0.001	0.4167	0.3849	0.3849	0.9748	0.9748	0.9129	0.9339
Accurate failure prob.		0.4167	0.3846	0.3846	0.9750	0.9750	0.9135	0.9341

It can be seen that as the approximation becomes more accurate, the number of necessary stations increases dramatically. Therefore, cautions shall be taken when one decides which supporting station structure to use.

4.4.3 Computational Treatment

Consider a facility system $\mathcal{J}$ where all facilities are possible to be disrupted (otherwise we can simply ignore those which are always functioning). In a positively correlated disruption profile regarding this system, $\forall j \notin J \subseteq \mathcal{J}$, we shall have

$$p^C_{j|J} = \frac{p^M_{\{j\}\cup J}}{p^M_J} \geq p^M_{\{j\}},\ \forall j \notin J \subseteq \mathcal{J}. \tag{4.15}$$

In such a case, $p^S_{\mathcal{J}} > 0$ must hold; otherwise if $p^S_{\mathcal{J}} = 0$, there must exist $j \notin J$ such that $p^M_{\{j\}} > 0, p^M_J > 0, p^M_{\{j\}\cup J} = 0$, which violates (4.15). Hence, according to the transformations from scenario reprentation to marginal and conditional representations, any element in both the marginal representation $\{p^M_J\}_{\forall J\subseteq\mathcal{J}}$ and the conditional representation $\{p^C_{j|J}\}_{\forall j\in\mathcal{J}, J\subseteq\mathcal{J}\setminus\{j\}}$ must have a positive value, and it is always feasible to use (4.10)-(4.11) to construct a supporting station representation.

However, when facility disruptions are not positively correlated, or if observed data is incom-

plete, the probability for all facilities to simultaneously disrupt may be 0 ($p^S_{\mathcal{J}} = 0$), and our decomposition equations (4.10)-(4.11) for station construction would divide a positive value by 0. To avoid such a mathematical artifact, we introduce a sufficient small value $\epsilon > 0$ to replace $p^S_{\mathcal{J}}$ if $p^S_{\mathcal{J}} = 0$. In this way, the original $\frac{C}{0}$ and $\frac{0}{C}$ expressions in equations (4.10)-(4.11) become $\frac{C}{\epsilon}$ and $\frac{\epsilon}{C}$, respectively, and consequently our decomposition approach can continue to be applied. From the proof of Proposition 6, this simple treatment will preserve the equivalence of the disruption profile representations and the station structure, except that only the original probability $p^S_{\mathcal{J}}$ will bear a small approximation error ϵ. By setting ϵ sufficiently small, we can limit the approximation error within an acceptable tolerance.

4.5 Numerical Examples

4.5.1 Hypothetical Examples

This section illustrates application of the proposed methodological framework to two examples of disaster patterns as shown in Figure 4.2(a) and 4.3(a). We also conduct sensitivity analysis to study how the station structure is dependent on various parameter settings in the next section.

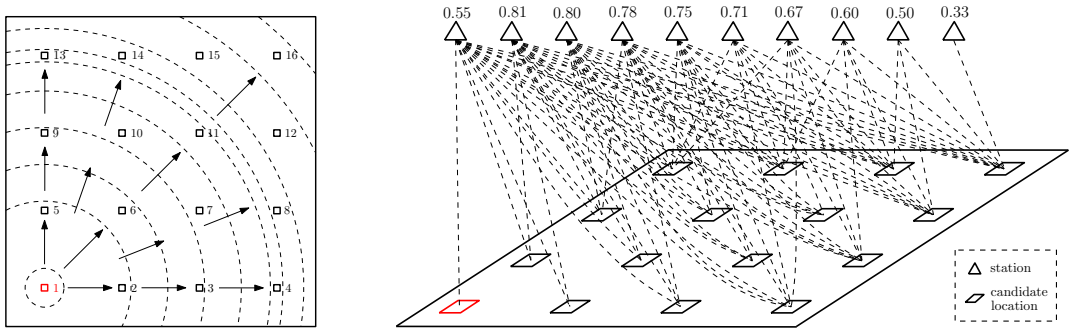
Example 1: Earthquake

Figure 4.2(a) illustrates sixteen evenly distributed facility locations in an 4×4 square area. We assume that the epicenter of a potential earthquake hazard is at location 1, since earthquake intensity drops with distance, facilities closer to the epicenter are more likely to be disrupted. We thus divide the city region into 10 rings centered at location 1. The facilities in each ring will fail together, and if that happens, all other facilities closer to the epicenter are already disrupted. Such a correlated disruption pattern can be described by a scenario representation as listed in Table 4.3.

Table 4.3: Scenario facility disruption probabilities under the earthquake hazard.

J	p^S_J	J	p^S_J
1	0.10	1,2,3,5,6,7,9,10,11	0.05
1,2,5	0.09	1,2,3,4,5,6,7,9,10,11,13	0.04
1,2,5,6	0.08	1,2,3,4,5,6,7,8,9,10,11,13,14	0.03
1,2,3,5,6,9	0.07	1,2,3,4,5,6,7,8,9,10,11,12,13,14,15	0.02
1,2,3,5,6,7,9,10	0.06	1,2,3,4,5,6,7,8,9,10,11,12,13,14,15,16	0.01

Marginal and conditional representations are first computed from (4.1) and (4.2), respectively, and part of the results are illustrated in Tables 4.4 and 4.5. It is observed that only positive



(a) Input network. (b) Station structure.

Figure 4.2: Illustrations of the network and station structure for the earthquake hazard.

correlation exists among the facility disruptions (and we can verify from Definition 1 that the disruption profile is positively correlated).

Table 4.4: Marginal facility disruption probabilities under the earthquake hazard.

J	p_J^M	J	p_J^M	J	p_J^M	J	p_J^M	J	p_J^M	J	p_J^M
1	0.55	5	0.45	9	0.28	13	0.10	1,2	0.45	1,6	0.36
2	0.45	6	0.36	10	0.21	14	0.06	2,6	0.36	1,2,3	0.28
3	0.28	7	0.21	11	0.15	15	0.03	1,2,6	0.36	2,3,6	0.28
4	0.10	8	0.06	12	0.03	16	0.01	2,5,6	0.36	1,2,5,6	0.36

Table 4.5: Conditional facility disruption probabilities under the earthquake hazard.

j	J	$p_{j\|J}^C$	j	J	$p_{j\|J}^C$	j	J	$p_{j\|J}^C$	j	J	$p_{j\|J}^C$
1	2	1.0000	1	2,6	1.0000	3	2	0.6222	3	1,2	0.6222
2	1	0.8182	2	1,3	1.0000	3	6	0.7778	3	2,6	0.7778
2	5	1.0000	2	1,5	1.0000	6	1	0.6545	6	1,2,3	1.0000
2	6	1.0000	2	1,5,6	1.0000	6	2	0.8000	6	1,2,5	0.8000

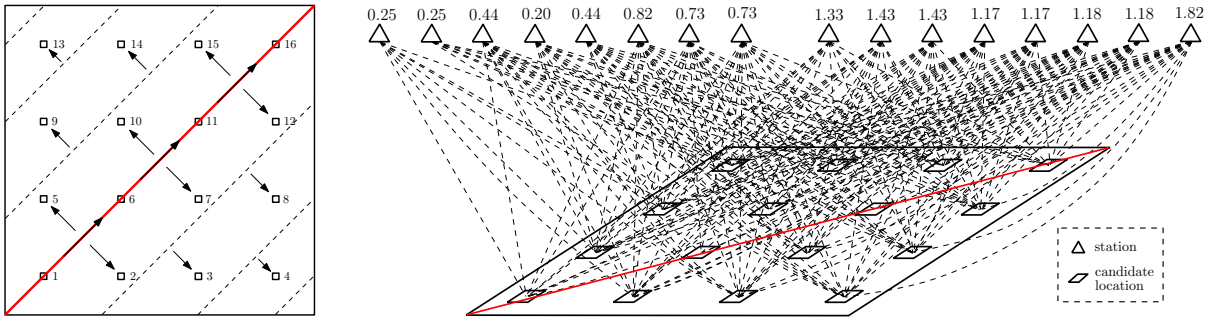
The supporting station structure from equation (4.11) includes 10 supporting stations, each with a site-dependent failure quasi-probability. Note that the number of stations is equal to the total number of input scenarios, indicating the compactness of the station structure. The detailed connections between the constructed stations and the facilities and the failure quasi-probabilities of each station are presented in Table 4.6 (J_k is denoted as the set of facilities connected to station k) and shown in Figure 4.2(b). All stations have a failure quasi-probability between 0 and 1, which verifies that the disruption profile is positively correlated.

Table 4.6: Station failure quasi-probabilities under the earthquake hazard.

k	J_k	q_k	k	J_k	q_k
1	16	0.3333	6	4,7,8,10,11,12,13,14,15,16	0.7500
2	12,15,16	0.5000	7	3,4,7,8,9,10,11,12,13,14,15,16	0.7778
3	8,12,14,15,16	0.6000	8	3,4,6,7,8,9,10,11,12,13,14,15,16	0.8000
4	4,8,12,13,14,15,16	0.6667	9	2,3,4,5,6,7,8,9,10,11,12,13,14,15,16	0.8182
5	4,8,11,12,13,14,15,16	0.7143	10	1,2,3,4,5,6,7,8,9,10,11,12,13,14,15,16	0.5500

Example 2: Flooding

Now we consider a different type of hazard, flooding, for the same square area. Figure 4.3(a) illustrates a river passing diagonally through locations 1, 6, 11, 16, as illustrated by the red line. Flooding may start randomly at any point along the river, i.e. locations 1, 6, 11, 16, and once it happens, water will spread in all directions and may disrupt nearby facilities. On the other hand, releasing flood water at one point could release the pressure and reduce the risk of flooding at other points. Hence, the facility disruptions exhibit both positive (e.g., along the lateral direction) and negative correlations (e.g., along the longitudinal direction). For this example, a total of 16 input scenarios are listed in Table 4.7.



(a) Input network.

(b) Station structure.

Figure 4.3: Illustrations of the network and station structure for the flooding hazard.

Table 4.7: Scenario facility disruption probabilities under the flooding hazard.

J	p_J^S	J	p_J^S	J	p_J^S
1	0.10	6,11	0.08	1,2,5,6	0.05
6	0.10	11,16	0.08	6,7,10,11	0.05
11	0.10	1,6,11	0.06	11,12,15,16	0.05
16	0.10	6,11,16	0.06	1,2,3,5,6,7,9,10,11	0.03
1,6	0.08	1,6,11,16	0.04	6,7,8,10,11,12,14,15,16	0.03
				1,2,3,4,5,6,7,8,9,10,11,12,13,14,15,16	0.01

Similar to Example 1, we first transform the scenario representation into marginal and conditional representations (the tables showing these transformation results are omitted to save space). Then the supporting station structure is constructed from (4.11). Again, the station structure is compact with only 16 effective supporting stations (i.e., identical to the number of input scenarios). The station-facility connections and the station quasi-probabilities are presented in Table 4.8 and Figure 4.3(b). The first subset of stations (#1 - 8) have disruption quasi-probabilities less than 1 (i.e., providing positive support), while the other stations (#9 - 16) have disruption quasi-probabilities larger than 1 (i.e., providing negative support). It can be also seen from the disruption profile representations that this disruption profile is generally correlated. For instance, the results of conditional and marginal representations show that $p^C_{6|\{11,16\}} = 0.5185 < p^M_{\{6\}} = 0.59$, which implies that disruption of facility 6 is generally correlated with disruptions of facilities 11 and 16. Meanwhile, $p^C_{1|\{6\}} = 0.4576 > p^M_{\{1\}} = 0.37$ shows that disruption of facility 1 is positively correlated with disruption of facility 6.

Table 4.8: Station failure quasi-probabilities under the flooding hazard.

k	J_k	q_k	k	J_k	q_k
1	1,2,3,4,5,9,13	0.2500	9	1,2,3,4,5,8,9,12,13,14,15,16	1.3333
2	4,8,12,13,14,15,16	0.2500	10	1,2,3,4,5,7,8,9,10,12,13,14,15	1.4286
3	1,2,3,4,5,6,7,8,9,10,13,14	0.4444	11	2,3,4,5,7,8,9,10,12,13,14,15,16	1.4286
4	2,3,4,5,7,8,9,10,12,13,14,15	0.2000	12	1,2,3,4,5,6,7,8,9,10,12,13,14,15	1.1667
5	3,4,7,8,9,10,11,12,13,14,15,16	0.4444	13	2,3,4,5,7,8,9,10,11,12,13,14,15,16	1.1667
6	1,2,3,4,5,7,8,9,10,12,13,14,15,16	0.8167	14	1,2,3,4,5,6,7,8,9,10,12,13,14,15,16	1.1768
7	1,2,3,4,5,6,7,8,9,10,11,12,13,14,15	0.7297	15	1,2,3,4,5,7,8,9,10,11,12,13,14,15,16	1.1768
8	2,3,4,5,6,7,8,9,10,11,12,13,14,15,16	0.7297	16	1,2,3,4,5,6,7,8,9,10,11,12,13,14,15,16	1.8158

4.5.2 Sensitivity Analysis

In this section, we introduce more variations to the earthquake and flooding cases to study the impacts of various system parameters on the size of the station structure. We now focus on a square grid network in a 2-dimensional Euclidean space, $\{1,2,\cdots,n\} \times \{1,2,\cdots,n\}$, in which each of the n^2 nodes is indexed by its coordinates $(i,j), 1 \leq i,j \leq n$. We consider four disaster cases and set up the relevant parameters as follows.

(1) Earthquake I: Again, we assume the epicenter is at the most bottom left location (1,1), and the distance from the epicenter to every node (i,j) is $d_{i,j}$. We first set the individual failure probability of location (i,j), $p_{i,j}$, to be approximately inversely proportional to distance $d_{i,j}$;

i.e., $p_{i,j} = (\alpha d_{i,j} + 1)^{-1}$, where $\alpha = 9/d_{n,n}$, so that $p_{1,1} = 1$ and $p_{n,n} = 0.1$. Then we generate all 2^{n^2} possible scenarios, each corresponding to a combination of independent disruptions at these n^2 locations. The probability of a scenario is calculated as the product of individual location's disruption or survival probability (i.e., either $p_{i,j}$ or $1 - p_{i,j}$) across all locations. After generating the complete set of scenarios, we conduct a random draw (based on each scenario's probability) to select a subset of these scenarios to form the scenario representation of the disruption profile. The probabilities for unselected scenarios are all set to zero. Then, based on the above scenario profile, we apply our model to calculate the number of stations needed in the station structure.

(2) Earthquake II: Now we assume a more realistic case where the affected areas form contour layers around the epicenter. Each layer consists of a subset of nodes that approximately have similar distance to the epicenter; i.e., $L_k = \{(i,j) : \max\{i,j\} = k\}$, and $|L_k| = 2k - 1$, for all $k = 1, 2, \cdots, n$. The probability for layer k to be affected by the earthquake is approximately inversely proportional to the distance of (k,k) to the epicenter, and if layer k is affected, then all $2^{|L_k|}$ possible combinations of location failures inside this layer are equally likely to occur. Therefore, if $(i,j) \in L_k$, then $p_{i,j} = 2^{1-2k}(\alpha d_{k,k} + 1)^{-1}$, where $\alpha = 9/d_{n,n}$. Furthermore, while generating scenarios, we enforce the additional rule that if any location in layer $k+1$ is disrupted, then all facilities in layers $1, \cdots, k$ must have been disrupted. The remainder of the process is similar to that for Earthquake I.

(3) Flooding I: Again, we assume that a river passes the area diagonally from $(1,1)$ to (n,n). Distance $d_{i,j}$ now measures the minimum distance from location (i,j) to the river, and the individual failure probability $p_{i,j} = (\alpha d_{i,j} + 2)^{-1}$, where $\alpha = 8/d_{n,1}$ so that $p_{1,1} = 0.5$ and $p_{n,1} = 0.1$. Then, we generate all possible scenarios, each corresponds to a combination of at least one but at most two failures along the river and arbitrary disruptions at the remaining $n^2 - n$ locations. By forcing the number of failures along the river to be no more than 2, we have introduced negative correlation into the disruption profile. The scenario probability is then computed as the product of the corresponding location probabilities. The remainder of the process is similar to that for Earthquake I.

(4) Flooding II: Similar to Earthquake II, we now introduce contour layers, each of which consists

of a subset of nodes that approximately have similar distance to the river; i.e., $L_k = \{(i,j) : |i-j| = k\}$, and $L_0 = n$, $|L_k| = 2(n-k)$, for $k \leq n-1$. The probability for layer k to be affected by flooding is now approximately inversely proportional to the distance of this layer to the river (i.e., all locations in a layer have equal distance to the river), and if layer k is affected, then all $2^{|L_k|}$ possible combinations of location failures inside this layer are equally likely to occur. Therefore, if $(i,j) \in L_k, k > 0$, then $p_{i,j} = 2^{2k-2n}(\alpha d_{k,k} + 2)^{-1}$, where α takes the same value as that in Flooding I. Furthermore, while generating scenarios, we enforce an additional rule: if a location (i,j) in layer k is disrupted, then its two neighboring locations in layers $k-1$ must have both been disrupted. The rest of the process is the same as that for Flooding I.

The setup of the above cases introduces two types of correlations: those due to omission of scenarios in the representation, and those due to generation of each scenario. For Earthquake I and Flooding I, although the disruptions of individual locations in each scenario are independent, only a subset of these scenarios are selected into the profile representation for station computation. The omission of unselected scenarios reflects on the possibility that while studying real-world data, only a subset of disaster scenarios may have ever been observed or documented in a limited period of time. For Earthquake II and Flooding II, we further enforce for each scenario that the impacts of the disaster spread out spatially from its source. This is similar to what happens in the real world, and also introduces correlation inside each of the selected scenarios. In summary, the failures in Earthquake II and Flooding II tend to be more positively correlated than those in Earthquake I and Flooding I, respectively. This is because of the distance-based spatial spreading pattern we enforce for each scenario. The failures in Earthquake I/II is more positively correlated than those in Flooding I/II, respectively, because of the way we generate the scenarios (e.g., for flooding we introduce negative correlation along the river direction).

In our computational experiments, we vary the number of locations n^2, and the number of scenarios selected for the profile representation. For each of the above four cases, we randomly generate scenario representations and compute the station structure based on our proposed model. In particular, we are interested in the number of supporting stations needed to equivalently represent the correlation profile. This simulation-computation process is repeated 50 times for each case, and the statistics of results are summarized in Table 4.9.

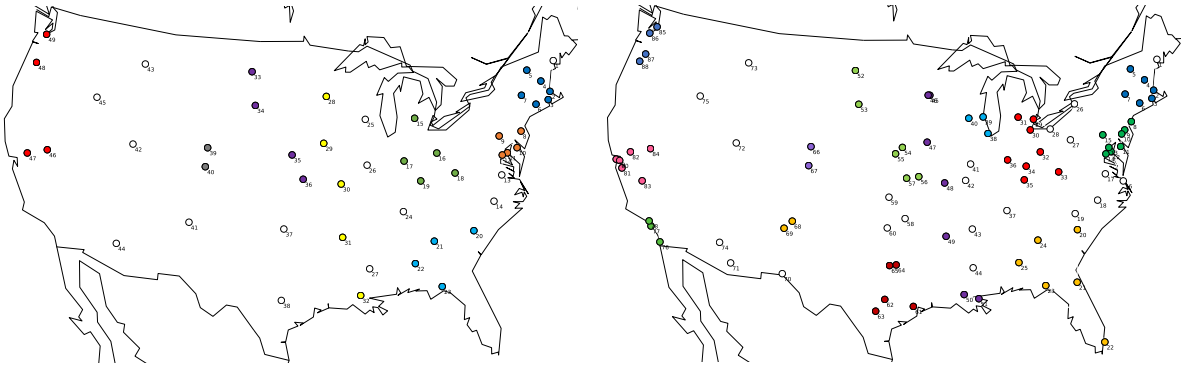
Table 4.9: Statistics on # of stations with various parameters and disaster patterns.

	Case setup parameters		Statistics on # of stations			
	# of locations	# of scenarios	mean	st.d.	max	min
Earthquake I	9	10	12.5	1.43	16	10
	9	15	17.9	1.57	22	15
	9	20	23.3	1.69	27	19
	16	10	16.8	2.83	24	11
	16	15	27.4	4.03	37	18
	16	20	35.2	4.54	46	25
Earthquake II	9	10	11.6	1.23	14	8
	9	15	16.8	4.25	24	9
	9	20	22.3	8.79	33	9
	16	10	13.3	1.53	16	10
	16	15	20.9	2.68	28	15
	16	20	29.5	3.98	36	18
Flooding I	9	10	16.1	2.48	23	12
	9	15	24.3	3.83	34	18
	9	20	31.7	2.45	38	27
	16	10	23.1	4.08	34	17
	16	15	37.4	5.53	49	24
	16	20	51.8	7.54	69	35
Flooding II	9	10	13.7	1.93	19	10
	9	15	20.7	3.50	26	16
	9	20	27.0	2.92	32	22
	16	10	18.8	2.96	25	12
	16	15	30.6	6.34	45	20
	16	20	44.3	7.21	59	29

We observe that for each of the four cases, the number of needed stations is comparable to, sometimes even smaller than, the number of input scenarios in the profile representation. This supports our earlier analysis that the number of needed stations does not grow exponentially with the number of input scenarios. Furthermore, by comparing across these four cases, we find that the number of stations needed in Earthquake I (or Flooding I) is generally larger than that in Earthquake II (or Flooding II). Also, the number of needed stations in Flooding I/II is generally larger than its counterpart in Earthquake I/II. These observations imply that the station structure is usually more compact when the disruptions involve more positive correlations. Those involving more negative correlations, in contrast, would require more stations. Although these results obviously depend on the choices of system parameters (e.g., disruption probabilities, scenario generation rules), we find consistent patterns and trends across these different cases. This suggests that our results are quite representative.

4.5.3 U.S. Network

We further test our methodology on the U.S. map, with data derived from 1990 census data: (i) a 49-node network with locations as the state capitals of the continental United States plus Washington, D.C.; and (ii) a 88-node network with the 49-node locations and other 39 largest cities in the United States. The data set is available from Professor L. Snyder's website http://www.lehigh.edu/lvs2. The two networks are shown in Figures 4.4(a) and 4.4(b), with 49 and 88 nodes, respectively.



(a) 49-node network. (b) 88-node network.

Figure 4.4: Input networks of the U.S. map with locations from 1990 census data.

49-node

As shown in Figure 4.4(a), Local disruption correlations are observed among the locations in each of the 8 local areas in the 49-node network. Facility disruptions across these local areas, however, are assumed to be independent. The scenario-based correlation profile (as if obtained from historical observations) for each local area is presented in Table 4.10, in which the column with header "J" lists the facility locations disrupted in each scenario, and the column with header "p_J^S" presents the corresponding scenario probabilities.

The resulting station structure, including the set of facilities J_k connected to each station k as well as its disruption quasi-probability q_k, is listed in Table 4.11. The total number of stations is 75, which is much smaller than the total number of scenarios $12{\times}7{\times}10{\times}7{\times}13{\times}11{\times}4{\times}8{\times}2^{14} = 4.41{\times}10^{11}$.

88-node

Similarly, the 88-node network as shown in Figure 4.4(b) includes 13 local areas. The scenario-based disruption correlation profile for each local area is presented in Table 4.12, and the resulting

Table 4.10: Scenario representation of disruption profile for the 49-node network.

J	p_J^S	J	p_J^S	J	p_J^S	J	p_J^S
Local I		**Local III**		**Local V**		**Local VI**	
2	0.008	15	0.010	28	0.010	33	0.010
3	0.008	15,16	0.008	29	0.008	34	0.008
2,3	0.006	15,17	0.008	30	0.006	35	0.006
2,3,4	0.005	15,16,17	0.008	28,29	0.008	36	0.005
2,3,4,5	0.004	15,16,18	0.006	29,30	0.006	33,34	0.008
2,3,6	0.005	15,17,19	0.006	30,31	0.005	34,35	0.006
2,3,6,7	0.004	15,16,17,18	0.005	28,29,30	0.006	35,36	0.005
2,3,4,6	0.005	15,16,17,19	0.005	29,30,31	0.005	33,34,35	0.006
2,3,4,5,6	0.003	15,16,17,18,19	0.005	30,31,32	0.004	34,35,36	0.005
2,3,4,6,7	0.003	**Local IV**		28,29,30,31	0.005	33,34,35,36	0.004
2,3,4,5,6,7	0.002	20	0.010	29,30,31,32	0.003	**Local VIII**	
Local II		23	0.012	28,29,30,31,32	0.005	47	0.010
12	0.010	20,23	0.008	**Local VII**		48	0.005
11,12	0.010	20,21,23	0.006	39	0.010	49	0.005
10,11,12	0.008	20,22,23	0.005	40	0.010	46,47	0.010
9,11,12	0.008	20,21,22,23	0.004	39,40	0.010	48,49	0.010
9,10,11,12	0.008					47,48,49	0.005
8,9,10,11,12	0.006					46,47,48,49	0.005

Table 4.11: Station failure quasi-probabilities for the 49-node network.

k	J_k	q_k	k	J_k	q_k	k	J_k	q_k	k	J_k	q_k
1	7	0.4000	20	18,19	0.8696	39	28,31,32	0.9600	58	48,49	0.3333
2	6,7	0.5556	21	17,19	0.6250	40	28,30,31,32	0.9783	59	46,47,48	0.8000
3	5	0.4000	22	17,18,19	0.9946	41	28,29	0.6667	60	46,47,49	0.8000
4	5,7	0.9615	23	16,18	0.6350	42	28,29,30,31,32	0.0978	61	46,47,48,49	0.0938
5	5,6,7	1.0636	24	16,18,19	0.9946	43	36	0.4000	62	1	0.0200
6	4,5	0.5556	25	16,17,18,19	0.9758	44	35,36	0.5556	63	13	0.0200
7	4,5,7	1.0636	26	15,16,17,18,19	0.0610	45	34,35,36	0.6429	64	14	0.0200
8	4,5,6,7	1.0062	27	22	0.4000	46	33	0.4444	65	24	0.0200
9	3,4,5,6,7	0.8222	28	21	0.4444	47	33,36	1.0714	66	25	0.0200
10	2,4,5,6,7	0.8222	29	21,22	0.9783	48	33,35,36	1.0216	67	26	0.0200
11	2,3,4,5,6,7	0.0547	30	21,22,23	0.6970	49	33,34	0.6429	68	27	0.0200
12	8	0.4286	31	20,21,22	0.6571	50	33,34,36	1.0208	69	37	0.0200
13	8,10	0.6364	32	20,21,22,23	0.0502	51	33,34,35	0.7368	70	38	0.0200
14	8,9	0.6364	33	32	0.5000	52	33,34,35,36	0.1190	71	41	0.0200
15	8,9,10	0.8643	34	31,32	0.6250	53	39	0.5000	72	42	0.0200
16	8,9,10,11	0.8000	35	30,31,32	0.6667	54	40	0.5000	73	43	0.0200
17	8,9,10,11	0.0500	36	29,30,31,32	0.7059	55	39,40	0.0400	74	44	0.0200
18	18	0.5000	37	28	0.6250	56	46	0.5000	75	45	0.0200
19	19	0.5000	38	28,32	0.8889	57	46,47	0.5000			

station structure is listed in Table 4.13. Again, the total number of stations is only 129, while the total number of scenarios is 4.4×10^{18}.

Table 4.12: Scenario representation of disruption profile for the 88-node network.

J	p_J^S	J	p_J^S	J	p_J^S	J	p_J^S
Local I		**Local III**		**Local IV**		**Local VII**	
2	0.008	22	0.012	29	0.010	52	0.010
3	0.008	21,22	0.010	31	0.010	53	0.008
2,3	0.006	22,23	0.010	29,31	0.008	54,55	0.006
2,3,4	0.005	21,22,23	0.008	29,30,31	0.007	56,57	0.006
2,3,4,5	0.004	20,21,22,23	0.006	29,30,31,32	0.006	52,53	0.008
2,3,6	0.005	21,22,23,24	0.006	29,30,31,32,33	0.004	53,54,55	0.006
2,3,6,7	0.004	21,22,23,25	0.006	29,30,31,32,34	0.005	54,55,56,57	0.005
2,3,4,6	0.005	21,22,23,24,25	0.005	29,30,31,32,36	0.005	52,53,54,55	0.005
2,3,4,5,6	0.003	20,21,22,23,24,25	0.004	29,30,31,32,34,36	0.004	53,54,55,56,57	0.004
2,3,4,6,7	0.003	**Local VI**		29,30,31,32,33,34	0.003	52,53,54,55,56,57	0.003
2,3,4,5,6,7	0.002	45,46	0.010	29,30,31,32,33,34,36	0.003	**Local VIII**	
Local II		47	0.008	29,30,31,32,34,35,36	0.003	61	0.010
8	0.008	48	0.006	29,30,31,32,33,34,35,36	0.002	62	0.008
9,10	0.008	49	0.005	**Local X**		63	0.008
11	0.008	50,51	0.004	68	0.010	64,65	0.008
8,9,10	0.006	45,46,47	0.008	69	0.010	62,63	0.006
9,10,11	0.006	47,48	0.006	68,69	0.020	61,62,63	0.005
8,9,10,11	0.005	48,49	0.005	**Local XI**		61,64,65	0.005
8,9,10,11,12	0.004	49,50,51	0.004	76	0.010	62,63,64,65	0.005
8,9,10,11,15	0.004	45,46,47,48	0.006	77,78	0.010	61,62,63,64,65	0.003
8,9,10,11,12,13,14	0.004	47,48,49	0.005	76,77,78	0.005	**Local XII**	
8,9,10,11,12,13,14,15	0.003	48,49,50,51	0.004	**Local XIII**		79,80	0.008
Local V		45,46,47,48,49	0.005	85	0.005	81	0.008
38	0.010	47,48,49,50,51	0.004	86	0.005	79,80,81	0.010
39	0.010	45,46,47,48,49,50,51	0.003	87	0.005	79,80,81,82	0.006
40	0.010	**Local IX**		88	0.005	79,80,81,83	0.005
38,39	0.008	66	0.010	85,86	0.010	79,80,81,82,83	0.004
39.40	0.008	67	0.010	87,88	0.010	79,80,81,82,84	0.004
38,39,40	0.005	66,67	0.010	85,86,87,88	0.005	79,80,81,82,83,84	0.003

4.6 Proof of Propositions

4.6.1 Proof of Proposition 3

Proof. First, equations (4.8) can be converted into

$$\log p_J^M = \sum_{J_1: J_1 \cap J \neq \emptyset} \log q_{k_{J_1}}, \ \forall J \subseteq \mathcal{J}, \tag{4.16}$$

which turn to be a system of linear equations with $(2^{|\mathcal{J}|} - 1)$ equations and variables. Rewrite is as

$$\mathbf{M} = \mathbf{A} \cdot \mathbf{P}, \tag{4.17}$$

where $\mathbf{M} := [\log p_J^M]_{J \subseteq \mathcal{J}}^{\mathrm{T}}$, $\mathbf{P} := [\log q_{k_J}]_{J \subseteq \mathcal{J}}^{\mathrm{T}}$, and $\mathbf{A}$ is the $(2^{|\mathcal{J}|} - 1) \times (2^{|\mathcal{J}|} - 1)$ coefficient matrix. To show that the station structure constructed from (4.16) is unique, we just need to prove that the solution to (4.17) is unique, i.e., $\mathbf{A}$ is not singular (or is invertible).

Table 4.13: Station failure quasi-probabilities for the 88-node network.

k	J_k	q_k	k	J_k	q_k	k	J_k	q_k
1	7	0.4000	45	40	0.3846	89	68,69	0.0450
2	6,7	0.5556	46	38,39	0.5652	90	76	0.3333
3	5	0.4000	47	38,40	1.0903	91	77,78	0.3333
4	5,7	0.9615	48	39,40	0.5652	92	76,77,78	0.0450
5	5,6,7	1.0636	49	38,39,40	0.0970	93	84	0.4286
6	4,5	0.5556	50	50,51	0.3750	94	83	0.4286
7	4,5,7	1.0636	51	49,50,51	0.5714	95	83,84	0.9608
8	4,5,6,7	1.0062	52	48,49,50,51	0.6364	96	82,84	0.5833
9	3,4,5,6,7	0.8222	53	47,48,49,50,51	0.6875	97	82,83,84	0.9107
10	2,4,5,6,7	0.8222	54	45,46	0.4286	98	81,82,83,84	0.8000
11	2,3,4,5,6,7	0.0547	55	45,46,50,51	1.0980	99	79,80,82,83,84	0.8000
12	15	0.4286	56	45,46,49,50,51	1.0259	100	79,80,81,82,83,84	0.0500
13	13,14,15	0.6364	57	45,46,48,49,50,51	1.0127	101	87,88	0.3333
14	12,13,14	0.4286	58	45,46,47	0.6364	102	86,87,88	0.7500
15	12,13,14,15	1.2833	59	45,46,47,50,51	1.0275	103	85,87,88	0.7500
16	11,12,13,14,15	0.7692	60	45,46,47,49,50,51	1.0080	104	85,86	0.3333
17	9,10,11,12,13,14,15	0.7647	61	45,46,47,48	0.7333	105	85,86,88	0.7500
18	8,12,13,14,15	0.7692	62	45,46,47,48,50,51	1.0130	106	85,86,87	0.7500
19	8,11,12,13,14,15	0.8450	63	45,46,47,48,49	0.7895	107	85,86,87,88	0.1422
20	8,9,10,12,13,14,15	0.7647	64	45,46,47,48,49,50,51	0.1693	108	1	0.0200
21	8,9,10,11,12,13,14,15	0.0684	65	56,57	0.3750	109	16	0.0200
22	24,25	0.4000	66	54,55,56,57	0.5000	110	17	0.0200
23	20	0.4444	67	53,54,55,56,57	0.6154	111	18	0.0200
24	20,25	0.6000	68	52	0.4286	112	19	0.0200
25	20,24	0.6000	69	52,56,57	1.0370	113	26	0.0200
26	20,24,25	1.7857	70	52,54,55,56,57	1.0588	114	27	0.0200
27	20,23,24,25	0.7778	71	52,53	0.5844	115	28	0.0200
28	20,21,24,25	0.7778	72	52,53,56,57	1.0640	116	37	0.0200
29	20,21,23,24,25	0.8635	73	52,53,54,55	0.6667	117	41	0.0200
30	20,21,22,23,24,25	0.0670	74	52,53,54,55,56,57	0.1335	118	42	0.0200
31	35	0.4000	75	64,65	0.3750	119	43	0.0200
32	35,36	0.6250	76	62,63	0.3750	120	44	0.0200
33	34,35,36	0.6667	77	62,63,64,65	0.9275	121	58	0.0200
34	33	0.4000	78	61	0.3750	122	59	0.0200
35	33,35	1.0417	79	61,64,65	1.1228	123	60	0.0200
36	33,35,36	0.9600	80	61,63,64,65	0.7037	124	70	0.0200
37	33,34,35	0.7059	81	61,62,64,65	0.7037	125	71	0.0200
38	33,34,35,36	1.2143	82	61,62,63	1.0159	126	72	0.0200
39	32,33,34,35,36	0.8333	83	61,62,63,64,65	0.1086	127	73	0.0200
40	30,32,33,34,35,36	0.8400	84	66	0.5000	128	74	0.0200
41	30,31,32,33,34,35,36	0.8333	85	67	0.5000	129	75	0.0200
42	29,30,32,33,34,35,36	0.8333	86	66,67	0.0400			
43	29,30,31,32,33,34,35,36	0.0720	87	68	0.6667			
44	38	0.3846	88	69	0.6667			

We first order all non-empty subsets in $\mathcal{J}$ into a sequence $\mathbf{S1} := [J^1, J^2, \cdots, J^{2^{|\mathcal{J}|}-1}]$ according to the partial order such that $J^j \not\subseteq J^i, \forall 1 \leq i < j \leq 2^{|\mathcal{J}|} - 1$. Apparently, in sequence $\mathbf{S1}$, $J^{2^{|\mathcal{J}|}-1} = \mathcal{J}$ and $J^i \subset \mathcal{J},, \forall 1 \leq i < 2^{|\mathcal{J}|} - 1$. Base on sequence $\mathbf{S1}$, we construct another sequence $\mathbf{S2} := [\mathcal{J}, \bar{J^1}, \bar{J^2}, \cdots, J^{2^{|\bar{\mathcal{J}}|}-2}]$. Apparently, $\mathbf{S1}$ contains every non-empty subset in $\mathcal{J}$ exactly once. Then we order equations (4.16) such that column indices $\{J_1\}$ follows sequence $\mathbf{S1}$ and row indices $\{J\}$ follow

sequence **S2**. Then the corresponding coefficient matrix $\mathbf{A}$ should be in the following form

$$\mathbf{A} = \begin{bmatrix} 1 & 1 & 1 & \dots & 1 & 1 & 1 \\ 0 & 1 & 1 & \dots & 1 & 1 & 1 \\ * & 0 & 1 & \dots & 1 & 1 & 1 \\ \vdots & \ddots & \ddots & \ddots & \vdots & \vdots & \vdots \\ \vdots & & \ddots & \ddots & 1 & 1 & 1 \\ \vdots & & & \ddots & 0 & 1 & 1 \\ * & \dots & \dots & \dots & * & 0 & 1 \end{bmatrix} = \mathbf{B} + \mathbf{u}\mathbf{v}^{\mathrm{T}},$$

where $*$ denotes an entry of either 0 or 1, $\mathbf{u} = [1, 1, 1, \cdots, 1, 1]^{\mathrm{T}}$, $\mathbf{v} = [1, 1, 1, \cdots, 1, 2]^{\mathrm{T}}$, and

$$\mathbf{B} = \begin{bmatrix} 0 & 0 & 0 & \dots & 0 & 0 & -1 \\ -1 & 0 & 0 & \dots & 0 & 0 & -1 \\ * & -1 & 0 & \dots & 0 & 0 & -1 \\ \vdots & \ddots & \ddots & \ddots & \vdots & \vdots & \vdots \\ \vdots & & \ddots & \ddots & 0 & 0 & -1 \\ \vdots & & & \ddots & -1 & 0 & -1 \\ * & \dots & \dots & \dots & * & -1 & -1 \end{bmatrix}.$$

It is obvious that $|\mathbf{B}| = (-1)^{2^{|\mathcal{J}|}} \neq 0$. We denote the (m, n) element of $\mathbf{B}^{-1}$ as $b'_{m,n}$. Since the inner product of the last column of $\mathbf{B}$ and the last row of $\mathbf{B}^{-1}$ should be 1, we obtain and have the following observations

$$\sum_{n=1}^{2^{|\mathcal{J}|}-1} b'_{2^{|\mathcal{J}|}-1,n} = -1. \tag{4.18}$$

Further, since the inner product of the last column of $\mathbf{B}$ and any other row (but the last row) of $\mathbf{B}^{-1}$ should be zero, we obtain

$$\sum_{m=1}^{2^{|\mathcal{J}|}-2} \sum_{n=1}^{2^{|\mathcal{J}|}-1} b'_{m,n} = 0. \tag{4.19}$$

Based on (4.18) and (4.19), we calculate the value of $\mathbf{v}^{\mathrm{T}}\mathbf{B}^{-1}\mathbf{u}$ as

$$\mathbf{v}^{\mathrm{T}}\mathbf{B}^{-1}\mathbf{u} = [1,1,\cdots,1,2]\ \mathbf{B}^{-1}\begin{bmatrix}1\\1\\\vdots\\1\\1\end{bmatrix} = 2\sum_{n=1}^{2^{|\mathcal{J}|}-1} b'_{2^{|\mathcal{J}|}-1,n} + \sum_{m=1}^{2^{|\mathcal{J}|}-2}\sum_{n=1}^{2^{|\mathcal{J}|}-1} b'_{m,n} = -2.$$

Since $\mathbf{B}$ is an invertible square matrix, and $1 + \mathbf{v}^{\mathrm{T}}\mathbf{B}^{-1}\mathbf{u} \neq 0$, the Sherman-Morrison formula (Sherman and Morrison, 1950) tells us that $\mathbf{A} = \mathbf{B} + \mathbf{u}\mathbf{v}^{\mathrm{T}}$ is also invertible, implying that the solution to (4.8) is unique.

Since there is a one-to-one correspondence between marginal representation and conditional representation (or scenario representation), we can infer that the solution to (4.7) (or (4.9)) should also exist and be unique, which completes the proof. □

4.6.2 Proof of Proposition 5

Proof. First, the second equality is straightforward. Since (4.9) expresses conditional representation $\{p^C_{j|J}\}_{\forall J\subseteq\mathcal{J}, j\notin J}$ in terms of station representation $\{q_{k_J}\}_{\forall J\subseteq\mathcal{J}}$, the right-hand side of (4.10) can be equivalently written as

$$\prod_{L:L\subseteq J\setminus\{j\}}\left[p^C_{j|(\bar{J}\cup L)}\right]^{(-1)^{|L|}} = \prod_{L:L\subseteq J\setminus\{j\}}\left[\prod_{\substack{J_1: j\in J_1\\ J_1\subseteq J\setminus L}} q_{k_{J_1}}\right]^{(-1)^{|L|}}, \quad j\in J,\ \forall J\subseteq\mathcal{J}. \tag{4.20}$$

Consider all L in (4.20) with cardinality $|L| = i \in [0, |J|-1]$, there are $\binom{|J|-|J_1|}{i}$ such L that satisfies $j\in J_1$ and $J_1\subseteq J\setminus L$. So the exponent of $q_{k_{J_1}}$ is

$$\sum_{i=0}^{|J|-|J_1|}(-1)^i\binom{|J|-|J_1|}{i} = \begin{cases}1, & \text{if } j\in J_1 = J,\\ 0, & \text{if } j\in J_1\subset J.\end{cases}$$

Hence we conclude that the right-hand side of (4.20) is q_{k_J}, which is equal to the left-hand side of (4.10). This completes the proof. □

4.6.3 Proof of Proposition 6

Proof. First, the second equality in (4.11) is straightforward, and equation (4.1) directly implies (4.12). Next, we express p_L^M in terms of the station representation $\{q_{k_J}\}_{\forall J \in \mathcal{J}}$ as

$$p_L^M = \prod_{J: J \cap L \neq \emptyset} q_{k_J} = \frac{\prod_{J \subseteq \mathcal{J}} q_{k_J}}{\prod_{J \subseteq \bar{L}} q_{k_J}}. \tag{4.21}$$

Substituting (4.21) and (4.1) into the right-hand side of (4.11) yields

$$\prod_{L: \bar{J} \subseteq L \subseteq \mathcal{J}} \left[p_L^M\right]^{(-1)^{|L|-|\bar{J}|+1}} = \frac{\prod_{L: \bar{J} \subseteq L \subseteq \mathcal{J}} \left[\prod_{J_1 \subseteq \mathcal{J}} q_{k_{J_1}}\right]^{(-1)^{|L|-|\bar{J}|+1}}}{\prod_{L: \bar{J} \subseteq L \subseteq \mathcal{J}} \left[\prod_{J_1 \subseteq \bar{L}} q_{k_{J_1}}\right]^{(-1)^{|L|-|\bar{J}|+1}}}, \ \forall J \subseteq \mathcal{J}. \tag{4.22}$$

The numerator in (4.22) is the product of $\left[\prod_{J_1 \subseteq \mathcal{J}} q_{k_{J_1}}\right]^{(-1)^{|L|-|\bar{J}|+1}}$ over all elements in $\{L : \bar{J} \subseteq L \subseteq \mathcal{J}\}$, which equals 1. To compute the denominator in (4.22), for any facility set $J_1 \subseteq J$, the number of L that satisfies $|L| - |\bar{J}| = i \in [0, |J|]$ and $J_1 \subseteq \bar{L} \subseteq J$ is $\binom{|J|-|J_1|}{i}$. Therefore the exponent of $q_{k_{J_1}}$ is

$$\sum_{i=0}^{|J|-|J_1|} (-1)^{i+1} \binom{|J| - |J_1|}{i} = \begin{cases} -1, & \text{if } J_1 = J, \\ 0, & \text{if } J_1 \subset J. \end{cases} \tag{4.23}$$

Finally, we substitute (4.23) into the right-hand side of (4.22) and obtain

$$\prod_{L: \bar{J} \subseteq L \subseteq \mathcal{J}} \left[p_L^M\right]^{(-1)^{|L|-|\bar{J}|+1}} = \frac{1}{q_{k_J}^{-1}} = q_{k_J}.$$

This completes the proof. □

4.6.4 Proof of Proposition 7

Proof. We prove this proposition using the marginal representation. The proofs with other representations can be similarly deduced with their transformation equations (4.1) - (4.4).

We first prove sufficiency. Recall the expression of q_{k_J} in terms of marginal representation

$\{p_J^M\}_{\forall J \subseteq \mathcal{J}}$ as (4.11), we pick an arbitrary j from J, and rewrite it as

$$q_{k_J} = \left[\prod_{L:\bar{J} \subseteq L \subseteq \mathcal{J} \setminus \{j\}} (p_L^M)^{(-1)^{|L|-|\bar{J}|+1}} \right] \cdot \left[\prod_{L:\bar{J} \cup \{j\} \subseteq L \subseteq \mathcal{J}} (p_L^M)^{(-1)^{|L|-|\bar{J}|+1}} \right], \tag{4.24}$$

which is exactly $Q(\mathcal{J}, \bar{J} \cup \{j\}, j)$ (see equation (4.6)). Since the disruption profile is positively correlated, we can conclude that $q_{k_J} = Q(\mathcal{J}, \bar{J} \cup \{j\}, j) \leq 1, \forall j \in J \subseteq \mathcal{J}$ according to Definition 1.

Then we prove necessity. In this case, we are given that $Q(\mathcal{J}, \bar{J} \cup \{j\}, j) = q_{k_J} \leq 1, \forall j \in J \subseteq \mathcal{J}$. We further can obtain

$$\begin{aligned} Q(J, J_1, j) &= \frac{\prod_{J \setminus \{j\} \supseteq L \supseteq J_1 \setminus \{j\}} (p_L^M)^{(-1)^{|L|-|J_1|}}}{\prod_{J \supseteq L \supseteq J_1} (p_L^M)^{(-1)^{|L|-|J_1|+1}}} \\ &= \frac{\prod_{J \cup \{j'\} \setminus \{j\} \supseteq L \supseteq J_1 \setminus \{j\}} (p_L^M)^{(-1)^{|L|-|J_1|}} \cdot \prod_{J \cup \{j'\} \setminus \{j\} \supseteq L \supseteq J_1 \cup \{j'\} \setminus \{j\}} (p_L^M)^{(-1)^{|L|-|J_1|+1}}}{\prod_{J \cup \{j'\} \supseteq L \supseteq J_1} (p_L^M)^{(-1)^{|L|-|J_1|+1}} \cdot \prod_{J \cup \{j'\} \supseteq L \supseteq J_1 \cup \{j'\}} (p_L^M)^{(-1)^{|L|-|J_1|}}} \\ &= Q(J \cup \{j'\}, J_1, j) \cdot Q(J \cup \{j'\}, J_1 \cup \{j'\}, j), \ \forall j \in J_1 \supseteq J \subseteq \mathcal{J}, j' \notin J. \end{aligned}$$

The above equation can be further expanded so we can obtain that

$$Q(J, J_1, j) = \prod_{J_1 \subseteq L \subseteq J_1 \cup \mathcal{J} \setminus \bar{J}} Q(\mathcal{J}, L, j) = \prod_{J_1 \subseteq L \subseteq J_1 \cup \mathcal{J} \setminus \bar{J}} q_{k_{\overline{L \setminus \{j\}}}} \leq 1, \ \forall j \in J_1 \supseteq J \subseteq \mathcal{J}.$$

This completes the proof. □

4.6.5 Proof of Lemma 1

Proof. Assume that $J = J_1 \cup \{j\}$ such that $p_{\bar{J}}^M = p_{\bar{J}_1}^M$, we have the following equation

$$p_{\bar{J}}^M = p_{\bar{J}_1}^M + \sum_{L:L \subseteq J_1} p_{\bar{J} \cup L}^S. \tag{4.25}$$

Then $p_{\bar{J}}^M = p_{\bar{J}_1}^M$ implies that

$$p_{\bar{J} \cup L}^S = 0, \ \forall L \subseteq J_1.$$

So for any $J_2 \subseteq J_1$, similar to (4.25) we have the following equation

$$p^M_{\bar{J}\cup J_2} = p^M_{\bar{J}_1\cup J_2} + \sum_{J_3: J_3 \subseteq \bar{J}_1 \setminus J_2} p^S_{\bar{J}_1 \cup J_2 \cup J_3} = p^M_{\bar{J}_1 \cup J_2} = p^M_{\bar{J}\cup J_2 \cup \{j\}}, \quad \forall J_2 \subseteq J_1. \tag{4.26}$$

Given (4.24), for any $L_1 \in \{L : \bar{J} \subseteq L \subseteq \mathcal{J}\backslash\{j\}\}$, (4.26) indicates that there exists $L_2 = L_1 \cup \{j\}$ such that $p^M_{L_1} = p^M_{L_2}$ and $|L_1| = |L_2| - 1$, which implies that

$$q_{k_J} = \frac{\prod_{L:\bar{J}\subseteq L\subseteq \mathcal{J}\backslash\{j\}} (p^M_L)^{(-1)^{|L|-|\bar{J}|+1}}}{\prod_{\bar{J}\cup\{j\}\subseteq L\subseteq \mathcal{J}} (p^M_L)^{(-1)^{|L|-|\bar{J}|}}} = 1.$$

This completes the proof. □

4.6.6 Proof of Lemma 2

Proof. According to Lemma 1, if $q_{k_J} \neq 1$, then $p^M_{\bar{J}} \neq p^M_{\bar{J}\cup\{j\}}, \forall j \in J$, based on which we conclude that

$$\sum_{J_1 \subseteq J\backslash\{j\}} p^S_{\bar{J}\cup J_1} \neq 0, \quad \forall j \in J. \tag{4.27}$$

That is, for any $j \in J$, at least one of the scenario probabilities $\{p^S_{\bar{J}\cup J_1}\}_{\forall J_1 \subseteq J\backslash\{j\}}$ is larger than 0. Let $J = \{j_1, j_2, \cdots, j_{|J|}\}$, we assume that

$$p^S_{\bar{J}\cup J_k} \neq 0,\ J_k \subseteq J\backslash\{j_k\}, \quad \forall k = 1, 2, \cdots, |J|.$$

If we denote $J^{'}_k = \bar{J} \cup J_k, \forall k = 1, 2, \cdots, |J|$, then each $J^{'}_k$ is an input scenario with probability $p^S_{J^{'}_k} \neq 0$, implying the following

$$\bigcap_{k=1}^{|J|} J^{'}_k = \bigcap_{k=1}^{|J|} \bar{J} \cup J_k = \bar{J}.$$

Denote the set formed by all distinct $J_k^{'}$ as $\tilde{J}$, then there must exist a scenario subset $\tilde{J} \subseteq \widetilde{\mathcal{J}}$ such that

$$\bar{J} = \bigcap_{J_1 \in \tilde{J}} J_1, \quad \tilde{J} \subseteq \widetilde{\mathcal{J}}.$$

This completes the proof. □

4.6.7 Proof of Proposition 9

Proof. For any two facilities j_1 and j_2 that are independent of each other, (4.8) implies

$$p^M_{\{j_1\}} = \prod_{J: j_1 \in J} q_{k_J},\ p^M_{\{j_2\}} = \prod_{J: j_2 \in J} q_{k_J},\ p^M_{\{j_1, j_2\}} = \prod_{J: J \cap \{j_1, j_2\} \neq \emptyset} q_{k_J}.$$

Substituting these equations into $p^M_{\{j_1\}} \cdot p^M_{\{j_2\}} = p^M_{\{j_1, j_2\}}$ yields

$$\prod_{J: j_1 \in J, j_2 \in J} q_{k_J} = 1.$$

Proposition 7 shows that, when facility disruptions are positively correlated, the disruption quasi-probability for any station $k \in \mathcal{K}$ satisfies $q_k \in [0, 1]$. Then, $\prod_{J: j_1 \in J, j_2 \in J} q_{k_J} = 1$ implies that $q_{k_J} = 1$ for any facility set J containing both j_1 and j_2. This completes the proof. □

CHAPTER 5:

RELIABLE FACILITY LOCATION UNDER CORRELATED FACILITY DISRUPTIONS WITH VIRTUAL SUPPORTING STATIONS

The previous chapter develops a recipe for transforming facility disruption correlations (e.g., those caused by shared hazards) into an augmented network with additional virtual supporting stations. It was proven that the augmented facility-station system with independent station failures can equivalently represent the facility system with correlated facility disruptions. How to optimally design the reliable locations of service facilities, however, remains an open nontrivial question. Therefore, a more complete systematic methodology framework including optimization module is needed to design reliable facility locations under correlated facility disruptions.[a]

In this chapter, we extend and combine the optimization and decomposition frameworks in the previous two chapters to develop a compact mixed-integer mathematical model. The facility location and customer assignment decisions are optimized to strike a balance between system reliability and cost efficiency. Additional virtual stations are derived using methods in Chapter 4 and can capture the effect of shared hazards. Extenstions are added to the optimization model proposed in Chapter 3. To hedge against the complexity associated with the new optimization model, new customized algorithms based on Lagrangian relaxation and approximation subroutines are further developed. Multiple numerical case studies with various types of correlated facility disruptions are carried out to demonstrate the performance and applicability of our models and algorithms.

[a]This chapter is based on a submitted paper, Xie et al. (2018a).

5.1 Motivation

Facility disruption correlations tend to have a strong impact on the performance of a reliable facility location design. Cui et al. (2010) proves that site-dependent failure probabilities do have impacts. Here we illustrate how correlated facility disruptions could further affect the facility location design. Based on the statement in Cui et al. (2010), when facilities located in the Gulf coast area (TX, LA, MS, AL, and FL) have a higher 10% chance of disruption, while other potential sites have a much lower failure probability of 5%, it is preferrable to locate facilities in the capitals of CA, PA, IL, GA, and OK in the 49-node U.S. network. However, if we further assume that GA and IL are perfectly positively correlated (i.e., they are disrupted simultaneously), it is even more cost efficient to hedge against the positive correlation by locating facilities in the capitals of PA, MI, GA, IA, OK, and CA. Specifically, we choose not to build facilities in both GA and IL. The expected system cost of this solution is \$952214, compared to \$955506 in Cui's solution, as shown in Table 5.1.

Table 5.1: System costs of independent and correlated disruptions.

Pattern	Item	Disrupted Facility	Disruption cost	Probability	Expected cost
Independent	Disruption cases	Harrisburg PA	908672	0.0429	38954
		Oklahoma City OK	660985	0.0429	28336
		Sacramento CA	1058226	0.0429	45365
		Atlanta GA Springfield IL	861292	0.0429	36923
	Normal case		549509	0.8145	446764
	Fixed cost		339500	1.0000	339500
	Total cost				**955506**
Correlated	Disruption cases	Harrisburg PA	748126	0.0387	28946
		Lansing MI	580746	0.0387	22469
		Atlanta GA	679479	0.0387	26288
		Des Moines IA	566169	0.0387	21905
		Oklahoma City OK	621288	0.0387	24037
		Sacramento CA	1023910	0.0387	39614
	Normal case		517239	0.7351	380218
	Fixed cost		378200	1.0000	378200
	Total cost				**952214**

From this example, we can see that the presence of correlation significantly affects the expected service cost. Such factors should be carefully considered and incorporated. When facility disruptions exhibit correlations, a straightforward modeling approach would involve some type of enumeration (or simulation and sampling), which may incur an exponential number of random scenarios; this makes it computationally difficult to even just evaluate the performance of a given

design. To the best of our knowledge, only a few efforts have been made to address correlated facility disruptions, either exactly or approximately (e.g., Liberatore et al. (2011); Li and Ouyang (2010); Lu et al. (2015)).

In light of these challenges, we build upon the idea of supporting station structure in the previous chapter, to address the reliable facility location problem with any patterns of facility disruption correlations. The additional layer of independent yet heterogeneous supporting stations are incorporated into the optimization framework to capture the effect of correlated disruptions.[b] As a result, the optimization model developed in this chapter, which transfers correlated facility disruptions to independent disruptions of such stations, is capable of addressing the facility location problem equivalently. A compact mixed-integer mathematical model is proposed to determine the optimal facility location and customer assignment plans. Several customized solution approaches based on Lagrangian relaxation are also developed. Case studies involving multiple patterns of correlations are conducted to demonstrate the performance and applicability of our methodology.

The remainder of this chapter is organized as follows. Section 5.2 introduces the mixed-integer mathematical models for the reliable facility location problem under facility disruption correlations. Section 5.3 presents the customized solution approaches to efficiently solve the optimization model. In Section 5.4, a range of case studies involving multiple patterns of correlations are shown.

5.2 Model Formulation

This section first presents the traditional scenario-based formulation of the reliable facility location problem under correlated facility disruptions. Then, the scenario-based reliable facility location model is transformed into an equivalent station-based model using the decomposition scheme presented in the previous chapter.

5.2.1 Scenario-based Formulation

We denote $\mathcal{I}$ as the set of discrete customers, and each customer $i \in \mathcal{I}$ has a demand μ_i. We define $\mathcal{J}$ to be the set of discrete candidate facility locations, and associate each location $j \in \mathcal{J}$ with a fixed facility cost f_j. The cost for a facility at location j to satisfy one unit of demand from customer i is denoted by d_{ij}.

Customers can go to candidate location $j \in \mathcal{J}$ for service if a facility is built and no disruption

[b] These stations are representations of real supporting infrastructures (instead of being virtual) in Chapter 3.

has occurred there. Under any realization of the facility states, each customer i seeks service by visiting the available and functioning facility that has the smallest transportation cost. Moreover, a penalty cost π_i per unit demand will be imposed if customer i does not receive any service. This situation occurs if no facility is reachable, or if the cost of serving customer i by the nearest available facility already exceeds π_i. We model this possibility by adding an "emergency" facility index by $j = 0$ with fixed cost $f_0 = 0$ and transportation costs $d_{i0} = \pi_i,\ \forall i \in \mathcal{I}$.

Let $\Omega = \{0,1\}^{|\mathcal{J}|}$ be the set of all possible disruption scenarios/realizations if facilities were built at all candidate locations (including the "emergency" facility). For each $\omega \in \Omega$, which occurs with probability p_ω, we use parameter $\delta_{j\omega} = 1$ to indicate that the facility at j (if built) is functioning in scenario ω, or 0 otherwise. The emergency facility is assumed to be always functioning, i.e., $\delta_{0\omega} = 1, \forall \omega \in \Omega$.

We denote X_j and $Y_{ij\omega}$ as binary variables indicating whether a facility is built at location j, and whether customer i visits facility j in scenario ω, respectively. Specifically,

$$X_j = \begin{cases} 1 & \text{if a facility is built at location } j; \\ 0 & \text{otherwise.} \end{cases}$$

$$Y_{ij\omega} = \begin{cases} 1 & \text{if customer } i \text{ visits facility } j \text{ in scenario } \omega; \\ 0 & \text{otherwise.} \end{cases}$$

Then it is straightforward to see that the reliable facility location problem could be formulated as the following scenario-based formulation (RFL-SCE):

$$\text{(RFL-SCE)} \quad \min \quad \sum_{j\in\mathcal{J}} f_j X_j + \sum_{i\in\mathcal{I}} \sum_{j\in\mathcal{J}\cup\{0\}} \sum_{\omega\in\Omega} \mu_i d_{ij} Y_{ij\omega} p_\omega \tag{5.1a}$$

$$\text{s.t.} \quad \sum_{j\in\mathcal{J}\cup\{0\}} Y_{ij\omega} = 1,\ \forall i \in \mathcal{I}, \omega \in \Omega, \tag{5.1b}$$

$$Y_{ij\omega} \le \delta_{j\omega} X_j,\ \forall i \in \mathcal{I}, j \in \mathcal{J}, \omega \in \Omega, \tag{5.1c}$$

$$X_j, Y_{ij\omega} \in \{0,1\},\ \forall i \in \mathcal{I}, j \in \mathcal{J}\cup\{0\}, \omega \in \Omega. \tag{5.1d}$$

The objective (5.1a) is the summation of the fixed facility costs and the expected transportation costs (including the penalty costs) across all possible facility disruption scenarios. Constraints

(5.1b) enforce that in any disruption scenario $\omega \in \Omega$, each customer i is either assigned to a regular facility or assigned to the emergency facility. Constraints (5.1c) require each customer to be assigned to only a functioning open facility. Given a problem with correlated disruptions at $|\mathcal{J}|$ candidate locations, the total number of possible scenarios that need to be enumerated is $2^{|\mathcal{J}|}$. This implies that formulation (RFL-SCE), which is an integer program, requires an exponential number of variables and constraints; thus it is extremely difficult to solve, if not impossible. So in the next sections, we introduce a supporting station structure as well as an alternative station-based formulation that is more compact in size and can be solved more efficiently.

5.2.2 Station-based Formulation

In this section, we take advantage of the station representation developed in Section 4.2.2 and the optimization model (RFL-SIF) proposed in Section 3.2 to propose a new formulation for the reliable facility location problem under any correlated facility disruptions. Each customer $i \in \mathcal{I}$ is assigned to one most preferred station-facility pair (k, j) plus a set of $(R-1)$ backup pairs as part of the service plan. Again, we enforce that a station will appear at no more than one of the backup pairs assigned to a customer. The transportation cost for station-facility pair (k, j) to satisfy one unit of demand from customer i is now denoted by $d_{ikj} = d_{ij}$ since no physical station exists. For all $i \in \mathcal{I}, j_1, j_2 \in \mathcal{J}$, we let $c_{ij_1j_2} = 1$ if $d_{ij_1} \leq d_{ij_2}$, or 0 otherwise.

We inherit the sets of decision variables from Section 3.2: variables $\{X_j\}_{j\in\mathcal{J}}$ denote the location decisions, $\{Y_{ikjr}\}_{i\in\mathcal{I},k\in\mathcal{K}\cup\{0\},j\in\mathcal{J}\cup\{0\},r\in\{1,2,\cdots,R+1\}}$ specify the assignment of customers to station-facility pairs at different backup levels, and $\{Z_{ikjr}\}_{i\in\mathcal{I},k\in\mathcal{K}\cup\{0\},j\in\mathcal{J}\cup\{0\},r\in\{1,2,\cdots,R+1\}}$ where $Z_{ikjr} \in \mathbb{R}$ denote the quasi-probabilities for customer i to be assigned to station-facility pair (k, j) at level r. The station-based reliable facility location problem (RFL-STA) under correlated facility disruptions is formulated as the following mixed-integer programming model:

$$\text{(RFL-STA)} \quad \min \quad \sum_{j\in\mathcal{J}} f_j X_j + \sum_{i\in\mathcal{I}} \sum_{k\in\mathcal{K}\cup\{0\}} \sum_{j\in\mathcal{J}\cup\{0\}} \sum_{r=1}^{R+1} \mu_i d_{ij} Z_{ikjr} Y_{ikjr} \tag{5.2a}$$

$$\text{s.t.} \quad \sum_{r=1}^{R} Y_{ikjr} \leq X_j, \ \forall i \in \mathcal{I}, j \in \mathcal{J}, k \in \mathcal{K}, \tag{5.2b}$$

$$Y_{ikjr} \leq l_{kj}, \ \forall i \in \mathcal{I}, j \in \mathcal{J} \cup \{0\}, k \in \mathcal{K} \cup \{0\}, r = 1, 2, \cdots, R+1, \tag{5.2c}$$

$$\sum_{j\in\mathcal{J}}\sum_{r=1}^{R} Y_{ikjr} \le 1,\ \forall i\in\mathcal{I}, k\in\mathcal{K}, \tag{5.2d}$$

$$\sum_{r=1}^{R+1} Y_{i00r} = 1,\ \forall i\in\mathcal{I}, \tag{5.2e}$$

$$\sum_{k\in\mathcal{K}}\sum_{j\in\mathcal{J}} Y_{ikjr} + \sum_{s=1}^{r} Y_{i00s} = 1,\ \forall i\in\mathcal{I}, r = 1,2,\cdots,R+1, \tag{5.2f}$$

$$Y_{ik_1j_1r} \le \sum_{s=1}^{r-1} Y_{ik_2j_2s} + c_{ij_1j_2} + 2 - l_{k_2j_2} - \left[\frac{\sum_{h\in\mathcal{I}}\sum_{k\in\mathcal{K}}\sum_{s=1}^{R} Y_{hkj_2s}}{|\mathcal{I}||\mathcal{K}|R}\right], \quad \forall i\in\mathcal{I}, j_1,j_2\in\mathcal{J}, k_1,k_2\in\mathcal{K}, 2\le r\le R, \tag{5.2g}$$

$$Z_{ikj1} = l_{kj}\left(1-q_k\right),\ \forall i\in\mathcal{I}, j\in\mathcal{J}\cup\{0\}, k\in\mathcal{K}\cup\{0\}, \tag{5.2h}$$

$$Z_{ikjr} = l_{kj}\left(1-q_k\right)\cdot\sum_{k'\in\mathcal{K}}\sum_{j'\in\mathcal{J}}\frac{q_{k'}}{1-q_{k'}} Z_{ik'j'(r-1)} Y_{ik'j'(r-1)}, \quad \forall i\in\mathcal{I}, j\in\mathcal{J}\cup\{0\}, k\in\mathcal{K}\cup\{0\}, r = 2,3,\cdots,R+1, \tag{5.2i}$$

$$X_j, Y_{ikjr}\in\{0,1\},\ \forall i\in\mathcal{I}, j\in\mathcal{J}\cup\{0\}, k\in\mathcal{K}\cup\{0\}, r = 1,2,\cdots,R+1. \tag{5.2j}$$

The objective function (5.2a) and constraints (5.2b)-(5.2f), (5.2h)–(5.2j) are exactly the same as those in (RFL). Constraints (5.2g) enforce that a customer is always assigned to the closest functioning station-facility pair for service; i.e., for any $1\le r\le R$ and two arbitrary station-facility pairs $(k_1,j_1),(k_2,j_2)$ with $d_{ij_2} < d_{ij_1}$, if facility j_2 is built, and a customer i is assigned to (k_1,j_1) at level r, then i should have been assigned to (k_2,j_2) at some level $s < r$. Constraints (5.2g) ensure equivalence between (RFL-STA) and (RFL-SCE), as we will prove in 10.

We then show in Proposition 10 below that the above formulation (RFL-STA) correctly captures the effect of correlated facility disruptions: with sufficiently large R, the station-based formulation (RFL-STA) is equivalent to the scenario-based formulation (RFL-SCE).

Proposition 10. *When $R = |\mathcal{K}|$, the station-based formulation (RFL-STA) with station structure is guaranteed to yield exactly the same optimal objective value and optimal solutions as the scenario-based formulation (RFL-SCE).*

Proof. See section 5.5.1. □

If $R < |\mathcal{K}|$, the two formulations are not necessarily equivalent. However, when only a limited number of facilities are built in the optimal solution, if R is as large as the total number of all

stations connected to the open facilities, the two formulations are equivalent. Furthermore, as the value of R only influences very high order terms in the formulation, even choosing an R value smaller than $|\mathcal{K}|$ would have only a small impact on the optimal location decisions.[c] More discussion on this choice can be found in Cui et al. (2010) and Section 5.4.2.

Formulation (RFL-STA) is nonlinear because the objective and constraints (5.2i) contain nonlinear terms $Z_{ikjr}Y_{ikjr}$. However, since each $Z_{ikjr}Y_{ikjr}$ is a product of a bounded continuous variable and a binary variable, we can linearize it by applying a variant of the technique introduced by Sherali and Alameddine (1992). First, since quasi-probability q_k can take any nonnegative value in $[0, \infty)$, Z_{ikjr} can take any real value in $[\overline{M}_k, \widehat{M}_k]$, where

$$\overline{M}_k = \min_{\forall L \subseteq \mathcal{K}\setminus\{k\}} \left[(1-q_k) \prod_{l \in L} q_l \right], \ \forall k \in \mathcal{K},$$
$$\widehat{M}_k = \max_{\forall L \subseteq \mathcal{K}\setminus\{k\}} \left[(1-q_k) \prod_{l \in L} q_l \right], \ \forall k \in \mathcal{K}.$$

We then can replace each $Z_{ikjr}Y_{ikjr}$ by a new continuous variable W_{ikjr} and enforce their equivalence by adding the following four sets of constraints

$$W_{ikjr} \leq Z_{ikjr} + \overline{M}_k(Y_{ikjr} - 1), \tag{5.3a}$$
$$W_{ikjr} \geq Z_{ikjr} + \widehat{M}_k(Y_{ikjr} - 1), \tag{5.3b}$$
$$W_{ikjr} \leq \widehat{M}_k Y_{ikjr}, \tag{5.3c}$$
$$W_{ikjr} \geq \overline{M}_k Y_{ikjr}. \tag{5.3d}$$

The model formulation is now transformed into the following:

$$\text{(LRFL-STA)} \quad \min \quad \sum_{j \in \mathcal{J}} f_j X_j + \sum_{i \in \mathcal{I}} \sum_{k \in \mathcal{K} \cup \{0\}} \sum_{j \in \mathcal{J} \cup \{0\}} \sum_{r=1}^{R+1} \mu_i d_{ij} W_{ikjr} \tag{5.4a}$$
$$\text{s.t.} \quad (5.2\text{b}) - (5.2\text{h}), \tag{5.4b}$$
$$Z_{ikjr} = (1-q_k) \sum_{k' \in \mathcal{K}} \sum_{j' \in \mathcal{J}} \frac{q_{k'}}{1-q_{k'}} W_{ik'j'(r-1)},$$

[c] One of the side-effects of the approximate model formulation (with a small R) is that the customers do not necessarily go to the nearest operational facility, but rather they may, at least theoretically, go to a more reliable yet farther one so as to lower the risk of losing service completely.

$$\forall i \in \mathcal{I}, j \in \mathcal{J} \cup \{0\}, k \in \mathcal{K} \cup \{0\}, r = 2, 3, \cdots, R+1, \quad (5.4c)$$

$$(5.3a) - (5.3d), \ \forall i \in \mathcal{I}, j \in \mathcal{J} \cup \{0\}, k \in \mathcal{K} \cup \{0\}, r = 1, 2, \cdots, R+1, \quad (5.4d)$$

$$X_j, Y_{ikjr} \in \{0, 1\}, \ \forall i \in \mathcal{I}, j \in \mathcal{J} \cup \{0\}, k \in \mathcal{K} \cup \{0\}, r = 1, 2, \cdots, R+1. \quad (5.4e)$$

This mixed-integer linear program (LRFL-STA) could in theory be solved by commercial solvers such as CPLEX and Gurobi. However, the existence of station-facility pairs as well as their associated site-dependent disruption quasi-probability exacerbates the model complexity. In light of this, we develop customized solution approaches in the next section.

5.3 Solution Approach

In this section, we extend the solution algorithm designed in Section 3.3 to provide a more accurate and efficient approach for (LRFL-STA) with virtual stations.

5.3.1 Lagrangian Relaxation

We choose to relax constraints (5.2b) in (LRFL-STA) with Lagrangian multipliers $\{\lambda_{ikj}\}_{\forall i \in \mathcal{I}, \forall k \in \mathcal{K}, \forall j \in \mathcal{J}}$ and move them as penalty terms to the objective function. The objective function becomes

$$\min \quad \sum_{j \in \mathcal{J}} \left(f_j - \sum_{i \in \mathcal{I}} \sum_{k \in \mathcal{K}} \lambda_{ikj} \right) X_j + \sum_{i \in \mathcal{I}} \sum_{k \in \mathcal{K} \cup \{0\}} \sum_{j \in \mathcal{J} \cup \{0\}} \sum_{r=1}^{R+1} \mu_i d_{ij} W_{ikjr} + \sum_{i \in \mathcal{I}} \sum_{k \in \mathcal{K}} \sum_{j \in \mathcal{J}} \lambda_{ikj} \sum_{r=1}^{R} Y_{ikjr}.$$

The above relaxation of the set of constraints (5.2b) essentially decouples the location and assignment variables $\mathbf{X}$ and $\mathbf{Y}$. The remaining model can be decomposed into multiple disjoint parts. The part involving $\mathbf{X}$,

$$\min_{X_j \in \{0,1\}, \forall j} \quad \sum_{j \in \mathcal{J}} \left(f_j - \sum_{i \in \mathcal{I}} \sum_{k \in \mathcal{K}} \lambda_{ikj} \right) X_j,$$

can be solved by simple inspection; i.e., given any $\{\lambda_{ikj}\}$, we can easily find the optimal $\mathbf{X}$ as follows:

$$X_j = \begin{cases} 1 & \text{if } f_j - \sum_{i \in \mathcal{I}} \sum_{k \in \mathcal{K}} \lambda_{ikj} < 0; \\ 0 & \text{otherwise.} \end{cases}$$

We further notice that the remaining problem can be further separated into individual subproblems, one for each customer, as long as we relax the term $\sum_{h\in\mathcal{I}}\sum_{k\in\mathcal{K}}\sum_{s=1}^{R} Y_{hkj_2s}$ in (5.2g) by $\sum_{k\in\mathcal{K}}\sum_{s=1}^{R} Y_{ikj_2s}$. The relaxed subproblem (RFL-STA-SP$_i$) with respect to customer i is

$$\text{(RFL-STA-SP}_i\text{)} \quad \min \quad \sum_{k\in\mathcal{K}\cup\{0\}}\sum_{j\in\mathcal{J}\cup\{0\}}\sum_{r=1}^{R+1}\mu_i d_{ij} W_{kjr} + \sum_{k\in\mathcal{K}}\sum_{j\in\mathcal{J}}\lambda_{kj}\sum_{r=1}^{R} Y_{kjr} \tag{5.5a}$$

$$\text{s.t.} \quad Y_{kjr} \le l_{kj}, \ \forall j \in \mathcal{J}\cup\{0\}, k \in \mathcal{K}\cup\{0\}, r = 1,2,\cdots,R+1, \tag{5.5b}$$

$$\sum_{j\in\mathcal{J}}\sum_{r=1}^{R} Y_{kjr} \le 1, \ \forall k \in \mathcal{K}, \tag{5.5c}$$

$$\sum_{r=1}^{R+1} Y_{00r} = 1, \tag{5.5d}$$

$$\sum_{k\in\mathcal{K}}\sum_{j\in\mathcal{J}} Y_{kjr} + \sum_{s=1}^{r} Y_{00s} = 1, \ \forall r = 1,2,\cdots,R+1, \tag{5.5e}$$

$$Y_{k_1j_1r} \le \sum_{s=1}^{r-1} Y_{k_2j_2s} + c_{ij_1j_2} + 2 - l_{k_2j_2} - \frac{\sum_{k\in\mathcal{K}}\sum_{s=1}^{R} Y_{kj_2s}}{|\mathcal{I}||\mathcal{K}|R},$$
$$\forall i \in \mathcal{I}, j_1, j_2 \in \mathcal{J}, k_1, k_2 \in \mathcal{K}, 2 \le r \le R, \tag{5.5f}$$

$$Z_{kj1} = 1 - q_k, \ \forall j \in \mathcal{J}, k \in \mathcal{K}, \tag{5.5g}$$

$$Z_{kjr} = (1-q_k)\sum_{k'\in\mathcal{K}}\sum_{j'\in\mathcal{J}}\frac{q_{k'}}{1-q_{k'}} W_{j'k'(r-1)},$$
$$\forall j \in \mathcal{J}, k \in \mathcal{K}, r = 2,3,\cdots,R+1, \tag{5.5h}$$

$$(5.3a) - (5.3d), \tag{5.5i}$$

$$Y_{kjr} \in \{0,1\}, \ \forall j \in \mathcal{J}, k \in \mathcal{K}, r = 1,2,\cdots,R+1. \tag{5.5j}$$

Note that (RFL-STA-SP$_i$), although still a mixed-integer linear program, is much smaller in size than the original (LRFL-STA), and hence it can often be efficiently handled by commercial solvers like CPLEX. However, solving this subproblem repeatedly (for each customer, and across Lagrangian relaxation iterations) could pose as a computational burden. Thus, section 5.3.3 further proposes an optional customized algorithm to solve (RFL-STA-SP$_i$).

The optimal objective values from the relaxed subproblems provide a lower bound to the original problem. Section 5.3.2 describes a heuristic to perturb the subproblem solutions in order to obtain a feasible solution to the original problem (which provides an upper bound). With the upper bound

and lower bound, we use standard subgradient techniques (Fisher, 2004) to update the multipliers λ in the Lagrangian procedure; i.e.,

$$\lambda_{ikj}^{n+1} = \lambda_{ikj}^{n} + t_j^n \left(\sum_r Y_{ikjr}^n - X_j^n \right), \tag{5.6}$$

$$t_j^n = \frac{\xi^n \left(Z^* - Z_D(\lambda^n) \right)}{\| \sum_r Y_{ikjr}^n - X_j^n \|^2}, \tag{5.7}$$

where λ_{ikj}^n represents a generic multiplier in the nth iteration, t^n is the step size, ξ^n is a scalar, and Z^* and $Z_D(\lambda^n)$ are the best upper bound and the current lower bound, respectively.

The above bounds, especially the lower bound, may be far from optimum (e.g., due to duality gaps from the relaxed constraints). If the Lagrangian relaxation algorithm fails to obtain a small enough gap in a certain number of iterations, we embed it into a branch-and-bound (B&B) framework to further reduce the gap. We construct a binary tree by branching on $\mathbf{X}$. Specifically, among all unbranched variables, we select and branch on the one whose construction yields the least system cost. After building the branching tree, we run the Lagrangian relaxation algorithm at each node to determine the corresponding feasible solution and lower bound, and update them after finishing both child branches. While traversing the binary tree, depth-first search is found to perform slightly better than breadth-first or least-cost-first searches for small or moderate-sized instances (which are likely to be solved to optimality). However, if the instances are large, it is difficult to traverse the entire tree and completely close the gap. In such cases, least-cost-first search is preferable since it tends to yield a reasonably good lower bound before completely traversing the entire tree.

5.3.2 Upper Bound

To obtain a good upper bound to the original model (RFL-STA), we first fix the facility location decisions from the relaxed subproblem. Then for each customer i, we sort all station-facility pairs associated with open facilities (i.e., pair (k,j) is considered if $X_j = 1, l_{kj} = 1$) in ascending order of (d_{ij}, p_k); (k_1, j_1) comes before (k_2, j_2) if $d_{ij_1} < d_{ij_2}$ or $d_{ij_1} = d_{ij_2}, q_{k_1} < q_{k_2}$. Then at every level r, we assign customer i to pair (k,j) with the smallest (d_{ij}, q_k) as long as i has never been assigned to any pair $(k,j'), \forall j'$ at levels $1, 2, \cdots, r-1$ before. The following two propositions state two properties of the optimal solution to (RFL-STA) and indicate that the feasible solution constructed from this

heuristic approach is likely to be near optimum.

First, constraints (5.2g) ensure the following property, which we state without proof:

Proposition 11. *(**Property I**) In any optimal solution* $(\mathbf{X}, \mathbf{Y}, \mathbf{Z})$ *to (RFL-STA), a customer will be assigned to backup station-facility pairs based on the corresponding distances; i.e., if* $Y_{ikjr} = 1$ *for some* i, k, j, r*, then* $X_{j'} = 0$ *or* $l_{k'j'} = 0$ *or* $\exists r' < r$ *s.t.* $Y_{ik'j'r'} = 1, \forall k', j'$ *with* $d_{ij'} < d_{ij}$*.*

Next, the following proposition reveals the relationship between assignment decisions and station disruption quasi-probabilities:

Proposition 12. *(**Property II**) In any optimal solution* $(\mathbf{X}, \mathbf{Y}, \mathbf{Z})$ *to (RFL-STA), a customer will be assigned to backup station-facility pairs that involve the same facility based on the corresponding disruption quasi-probabilities of the associated stations; i.e., if* $Y_{ikjr} = 1$ *for some* i, k, j, r*, then* $l_{k'j} = 0$ *or* $\exists j', r' < r$ *s.t.* $Y_{ik'j'r'} = 1$, $\forall k'$ *with* $q_{k'} \leq q_k$*.*

Proof. See section 5.5.2. □

Based on these two properties, given location decisions from the relaxed subproblem solutions, if R is sufficiently large, this heuristic yields the optimal customer assignments; otherwise, it can only guarantee feasible but not necessarily optimal assignments. Nevertheless, since the quasi-probabilities for using high-level back-ups (i.e., the product of multiple station disruption quasi-probabilities, which is equivalently the product of multiple facility disruption scenario probabilities) are often smaller by orders of magnitude, the solution given by this sorting/greedy heuristic shall be quite close to the optimal one.

5.3.3 Lower Bound

As mentioned before, although the relaxed problem is separable by customer i, each subproblem is still combinatorial and the worst-case complexity is exponential. Therefore, in this section, we develop an algorithm which helps quickly find lower bounds to the relaxed subproblems (RFL-STA-SP$_i$).

Equations (5.5h) show that Z_{kjr} depends on $Z_{kj(r-1)}$ and $Y_{kj(r-1)}$, which builds connections across the decision variables and brings difficulty in solving subproblem (RFL-STA-SP$_i$). Instead of having Z_{kjr} directly in the formulation, we approximate them with fixed numbers.

Similar properties as those stated in Propositions 11 and 12 apply to (RFL-STA-SP$_i$), which suggest that certain customer-station-facility assignments would never appear in the optimal solution to (RFL-STA-SP$_i$). We summarize them into the following rules:

Rule 1 If customer i is assigned to two different facilities at some levels, then it will always be assigned to the closer facility at a lower level; i.e., $Y_{ik_1j_1r_1} = Y_{ik_2j_2r_2} = 1$, and $d_{ij_1} < d_{ij_2} \Rightarrow r_1 < r_2$;

Rule 2 If customer i is assigned to a facility j, then it will be assigned to all station-facility pairs associated with j, $\{(k,j)\}_{\forall k: l_{kj}=1}$ at consecutive backup levels as long as k has never been used at a lower level; i.e., let $K_j = \{k : l_{kj} = 1, Y_{ikls} = 0, \forall l \in \mathcal{J}, l \neq j, s < r\}$, then $Y_{ikjr} = 1$ for some $k, r \Rightarrow Y_{ik_1js} = Y_{ik_2j(s+1)} = \cdots = Y_{ik_nj(s+n-1)} = 1$ for some permutation $k_1, k_2, \cdots, k_n$ of K_j and some s;

Rule 3 If customer i is assigned to multiple station-facility pairs $(k_1,j),(k_2,j),\cdots,(k_n,j)$ that involve the same facility j, then these pairs should be used in ascending order of the involved station disruption quasi-probabilities; i.e., $Y_{ik_1jr} = Y_{ik_2j(r+1)} = \cdots = Y_{ik_nj(r+n-1)} = 1 \Rightarrow q_{k_1} \leq q_{k_2} \leq \cdots \leq q_{k_n}$;

Based on these rules, we can set $Y_{kjr} = Z_{kjr} = 0$ for some (k,j,r) without affecting the optimal solution. For example, in the system shown in Figure 4.1(b), if we assume that $d_{i_1j_1} < d_{i_1j_2}, q_{k_{\{1\}}} < q_{k_{\{1,2\}}} < q_{k_{\{2\}}} < q_{k_{\{1,2,3\}}} < q_{k_{\{1,3\}}} < q_{k_{\{2,3\}}}$, then we have: (i) $Y_{i_1k_{\{1,2\}}j_11} = Y_{i_1k_{\{1,2\}}j_13} = 0$ because j_1 should be used first if it is built, and $k_{\{1,2\}}$ should be used after $k_{\{1\}}$ and before $k_{\{1,3\}}, k_{\{1,2,3\}}$; (ii) $Y_{i_1k_{\{2\}}j_21} = Y_{i_1k_{\{2\}}j_23} = Y_{i_1k_{\{2\}}j_24} = Y_{i_2k_{\{2\}}j_26} = 0$ because $k_{\{2\}}$ should always be used after $k_{\{1,2\}}$, and also after $k_{\{1\}}, k_{\{1,2,3\}}, k_{\{1,3\}}$ if j_1 is used. After setting these Y_{kjr} and Z_{kjr} to be zero, we can use algorithm **LowerBound**(i), for each customer i, to construct lower bounds to Z_{kjr} and Z_{00r} as α_{kjr} and β_r, respectively, and relax (RFL-STA-SP$_i$) into the following (RFL-STA-RSP$_i$):

We replace Z_{kjr} and Z_{00r} respectively by their estimates α_{kjr} and β_r, and relax (RFL-STA-SP$_i$) into the following (RFL-STA-RSP$_i$):

$$\text{(RFL-STA-RSP}_i\text{)} \quad \min \quad \sum_{k\in\mathcal{K}}\sum_{j\in\mathcal{J}}\sum_{r=1}^{R+1}\left(\mu_i d_{ij}\alpha_{kjr} + \lambda_{kjr}\right)Y_{kjr} + \sum_{r=1}^{R+1}\mu_i d_{i0}\beta_r Y_{00r} \tag{5.8a}$$

$$\text{s.t.} \quad Y_{kjr} \leq l_{kj}, \quad \forall j \in \mathcal{J}\cup\{0\}, k \in \mathcal{K}\cup\{0\}, r = 1,2,\cdots,R+1, \tag{5.8b}$$

Algorithm 1 Construct α_{kjr} and β_r as lower bounds to Z_{kjr} and Z_{00r}, respectively, $\forall i \in \mathcal{I}$

LowerBound(i)

for r **do**
  for (k,j) **do**
    $\alpha_{kjr} = 0$
    if $l_{kj} = 1$ **then**
      $K_k^r = 0$, prob $= 1.0$
      for k' **do**
        if $k' \neq k, l_{k'j} = 1, q_{k'} < q_k$ **then**
          $K_k^r = K_k^r + 1$, prob $=$ prob $\cdot\, q_{k'}$
        end if
      end for
      if $K_k^r < r$ **then**
        minProduct $= 1.0$
        if $K_k^r < r - 1$ **then**
          for (k', j') **do**
            if $j' \neq j, k' \neq k, d_{ij'} < d_{ij}, \alpha_{k'j',r-1-K_k^r} \in (0, \text{minProduct})$ **then**
              minProduct $= \alpha_{k'j',r-1-K_k^r}$
            end if
          end for
        end if
        $\alpha_{kjr} = q_k \cdot$ prob $\cdot$ minProduct
      end if
    end if
  end for
end for
for r **do**
  $\beta_r = 1.0$, minProduct $= 1.0$
  for (k,j) **do**
    if $l_{kj} = 1, \alpha_{kjr} \in (0, \text{minProduct})$ **then**
      minProduct$= \alpha_{kjr}$
    end if
  end for
  $\beta_r =$ minProduct
  for (k,j) **do**
    $\alpha_{kjr} = \alpha_{kjr} \cdot (1 - q_k)/q_k$
    if $\alpha_{kjr} = 0$ **then**
      $\alpha_{kjr} = \infty$
    end if
  end for
end for

$$\sum_{j\in\mathcal{J}}\sum_{r=1}^{R} Y_{kjr} \le 1, \quad \forall k \in \mathcal{K}, \tag{5.8c}$$

$$\sum_{r=1}^{R+1} Y_{00r} = 1, \tag{5.8d}$$

$$\sum_{k\in\mathcal{K}}\sum_{j\in\mathcal{J}} Y_{kjr} + \sum_{s=1}^{r} Y_{00s} = 1, \quad \forall r = 1,2,\cdots,R+1, \tag{5.8e}$$

$$Y_{kjr} \in \{0,1\}, \quad \forall j \in \mathcal{J}, k \in \mathcal{K}, r = 1,2,\cdots,R+1. \tag{5.8f}$$

Proposition 13. *The optimal objective of (RFL-STA-RSP$_i$) provides a lower bound to the optimal objective of (RFL-STA-SP$_i$).*

Proof. Let $\mathbf{Y}^*, \mathbf{Z}^*$ and $\mathbf{W}^*$ be the optimal solution to (RFL-STA-SP$_i$). We can construct (RFL-STA-RSP$_i$) from (RFL-STA-SP$_i$) in three sequential steps: (i) replace Z_{kjr} and Z_{00r} by α_{kjr} and β_r, respectively, and add constraints to set $Y_{kjr} = 0$ for some (k,j,r) pairs; (ii) remove constraints (5.5f)–(5.5i); and (iii) remove those constraints $Y_{kjr} = 0$ that were added in step (i), and instead, set the corresponding α_{kjr} to be sufficiently large. In step (i), we know that adding those $Y_{ikj} = 0$ does not change the optimal solution to (RFL-STA-SP$_i$), and based on the construction of α_{kjr} and β_r, we know $\alpha_{kjr}Y^*_{kjr}$ and $\beta_r Y^*_{00r}$ are lower bounds to Z^*_{kjr} and Z^*_{00r}, respectively. In step (ii), removing constraints obviously never increases the objective value of a minimization problem. Step (iii) just uses an alternative way to enforce the $Y_{kjr} = 0$ constraints; i.e., when the coefficients of those Y_{kjr} are set to be infinity, these variables cannot equal 1 at optimality (because a finite feasible solution is known to exist). Therefore, each of the three steps provides a lower bound to the model built in the previous step, hence the optimal objective value of (RFL-STA-RSP$_i$) is a lower bound to the optimal objective value of (RFL-STA-SP$_i$). This completes the proof. □

We observe that (RFL-STA-RSP$_i$) is a combinatorial generalized assignment problem, which can be solved by an adapted Hungarian algorithm as in Cui et al. (2010). (RFL-STA-RSP$_i$) aims at assigning one station-facility pair to each level (up to $R+1$) based on the updated coefficients associated with each Y_{kjr}, so as to minimize the total expected system cost. However, the actual maximum assignment level $R_{\max}$ (i.e., the largest r such that $Y_{kjr} = 1$ for some (k,j) pair) may be smaller than R due to lower cost associated with the emergency station-facility pair than all other remaining pairs at some level $r < R$. The main challenge is to identify the level that the

emergency station-facility pair should be assigned to. As such, we enumerate $R_{\max}$ from 0 to R and for each $R_{\max}$, we fix $Y_{00,R_{\max}+1} = 1$ and $Y_{kjr} = 0, r > R_{\max}+1$. In this way, the (RFL-STA-RSP$_i$) is simplified into a standard assignment problem that can be solved by conventional Hungarian algorithm. We solve (RFL-STA-RSP$_i$) and calculate the associated total cost for each enumeration of $R_{\max}$. By comparison, the value of $R_{\max}$ corresponding to the lowest total cost is the actual maximum assignment level $R_{\max}$. After fixing $R_{\max}$, It is worth noting that in the enumeration process, the assignment solutions to model with $R_{\max} = r$ can be used as a warm start to the model with $R_{\max} = r + 1$, which helps expedite the computation. Specifically, if the penalty cost d_{i0} (or say π_i) is sufficiently large, we only need to solve (RFL-STA-RSP$_i$) for one iteration, i.e., $R_{\max} = R$.

5.4 Numerical Examples

We apply the proposed model and solution algorithms to three sets of examples under different correlation patterns and parameter settings. The first set investigates the optimal facility location design for the hypothetical examples presented in section 4.5.1. The second set of examples demonstrate reliable facility system planning for the U.S. networks with 49 and 88 nodes under given correlation patterns; see Figure 4.4.[d]. The third set of examples are used to compare the results and computational performance of the scenario-based and station-based formulations.

The proposed solution algorithms are programmed in C++ and run on a 64-bit Intel i7-3770 computer with 3.40 GHz CPU and 8G RAM. The reformulated problem (RFL-STA-RSP$_i$) is solved by the Hungarian algorithm.

5.4.1 Hypothetical Examples

We first illustrate our optimization methodology on the two hypothetical cases with Earthquake and Flooding disaster patterns from Section 4.5.1.

Example 1: Earthquake

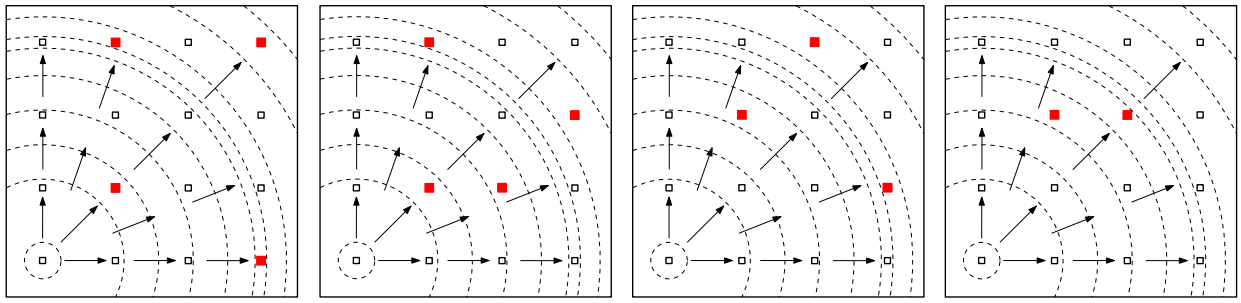
We consider the square urban area in Figure 4.2(a), which is subject to earthquake risks. The area is evenly divided into 16 indexed square cells, and the centroid of each represents a candidate facility location as well as a population center, i.e. $|\mathcal{J}| = |\mathcal{I}| = 16$. Each point generates an

[d]The location data set is from Snyder and Daskin (2005) and can be accessed at http://www.lehigh.edu/~lvs2, all input data, including the correlation profile, will be available for download at my website http://www.siyangxie.com.

equal demand of $\mu_i = 1.25, \forall i \in \mathcal{I}$, and this adds up to a city-wide total demand of 20. The construction cost of each facility is $f_j = 30, \forall j \in \mathcal{J}$. Euclidean distance between customer i and facility j (regardless of virtual supporting station k) is used to measure d_{ij}. Penalty for loss of unit demand is $\pi_i = 60, \forall i$. The maximum assignment level is $R = 10$, which is equal to the number of added virtual stations. For comparison, we also study other cases when there is no correlation, and when customer demand displays heterogeneity. When there is no correlation, we assume that facility disruptions are independent but maintain the same marginal probabilities; the optimization model degrades to the one in Cui et al. (2010). For the heterogeneous cases, we assume that $\mu_i = 3.0 - 0.5d_i$ (such that $\sum_i \mu_i = 20$), where d_i is the Euclidean distance from location i to the city center. The computational time for every case is within 1.5 minutes. Figures 5.1(a)-5.1(d) depict the optimal facility location design for each of the four cases (correlated vs. uncorrelated disruption, homogeneous vs. heterogeneous demand).

It can be seen that disruption correlations and demand heterogeneity both have significant impacts on the optimal facility locations. In the first case (with correlated disruption and homogeneous demand), the four built facilities are quite spatially dispersed so as to minimize the positive correlations among these facilities. For example, at most one facility is selected in each ring, because those in the same ring will suffer simultaneous disruptions and will not be able to back up each other. In Figure 5.1(b), facilities are more clustered toward the city center when there is no disruption correlation. Facilities at locations 7 and 12 (which are closest to the city center) in this case are obviously geographically more advantageous over those at corner locations 4 and 16 in Figure 5.1(a). Meanwhile, we also observe that the existence of disruption correlation pushes the facilities away from the epicenter: the facility at location 4 is two rings farther than that at location 7, and the one at 16 is one ring farther than the one at 12. The same impact of disruption correlation can be found under heterogeneous demands. To study the impact of demand heterogeneity, we now compare Figure 5.1(a) with 5.1(c), and also Figure 5.1(b) with 5.1(d). Intuitively, candidate locations with high customer demand are more favorable, and hence the facilities in Figure 5.1(b) and 5.1(d) are somewhat more clustered toward the city center than their respective counterparts. The facility number is also reduced by 1 due to the economic benefits of demand concentration.

Table 5.2 summarizes the computational results. For each case, we list the optimal objective value from the respective model (e.g., ignoring correlation), as well as the evaluated system-wide



(a) correlation & homogeneous demand (b) no correlation & homogeneous demand (c) correlation & heterogeneous demand (d) no correlation & heterogeneous demand

Figure 5.1: Optimal facility locations for the four earthquake cases.

cost assuming the ground-truth that disruption correlation does exist. It can be seen that the model without considering correlation results in a sub-optimal facility design (i.e., with higher system costs) for both homogeneous and heterogeneous demand. Heterogeneity in demand further exacerbates the cost difference to up to 28.6%.

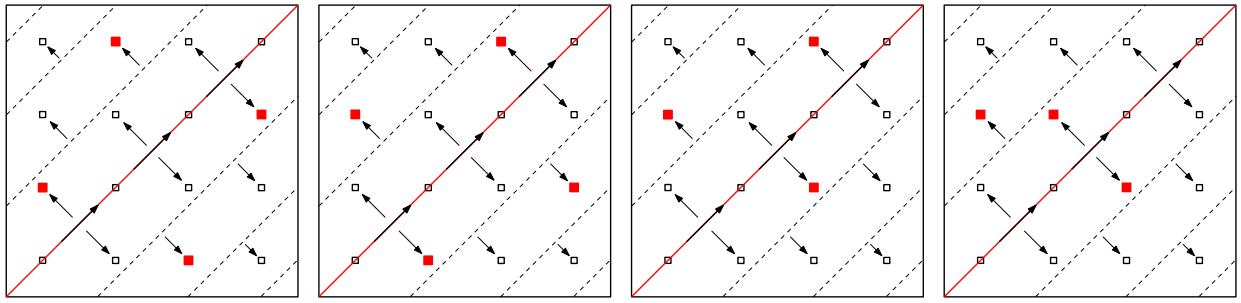
Table 5.2: Solution statistics for the earthquake cases.

Case		Facility	Objective	Evaluated cost	Cost	Comp.
Corr	Homo	location		under correlation	difference (%)	time (s)
Yes	Yes	4,6,14,16	339	339	–	83
No	Yes	6,7,12,14	315	349	2.9	70
Yes	No	8,10,15	325	325	–	69
No	No	10,11	302	418	28.6	2

Example 2: Flooding

We then consider the flooding case in Figure 4.3(a), in which the same city area is threatened by flooding from a river passing diagonally. A station structure with 16 additional virtual stations are built from Section 4.5.1. Other system parameters are similar to those in the earthquake case, except that the maximum customer assignment level is now $R = 16$, and we also examine four cases (correlated vs. uncorrelated disruption, homogeneous vs. heterogeneous demand). The optimal facility location designs are shown in Figure 5.2.

Again, disruption correlation and demand heterogeneity are observed to influence the optimal facility locations to some extent. Under homogeneous demand, our model determines that at optimality four facilities should be built somewhat evenly along the river, each at a different distance to the river. It shall be noted that the facility layouts in Figures 5.2(a) and 5.2(b) are actually iden-



(a) correlation & homogeneous demand (b) no correlation & homogeneous demand (c) correlation & heterogeneous demand (d) no correlation & heterogeneous demand

Figure 5.2: Optimal facility locations for the four flooding cases.

tical since the failure disruption scenarios are set to be symmetric along the river. It indicates that correlation does not affect the facility locations under homogeneous demand, probably because (i) facilities at locations 3, 5, 12, 14 or 2, 8, 9, 15 have relative low individual failure probabilities, and (ii) there are only very weak correlation among them (due to co-existence of positive and negative correlations). Under heterogeneous demand, three facilities are clustered toward the concentrated demand near the city center. Particularly, facilities in Figure 5.2(d) (with no correlation) are more clustered than those in Figure 5.2(c). The expected system cost and the cost difference for the four cases are summarized in Table 5.3. Similar to those for the earthquake cases, heterogeneous demand could reduce the system cost, and ignoring negative correlation leads to sub-optimal solutions. The computation time is a little larger as a result of having more virtual stations, but our proposed algorithm still solves the problem quite effectively.

Table 5.3: Solution statistics for the flooding cases.

Case		Facility	Objective	Evaluated cost	Cost	Comp.
Corr	Homo	location		under correlation	difference (%)	time (s)
Yes	Yes	2,8,9,15	294	294	–	189
No	Yes	3,5,12,14	284	294	0.0	4
Yes	No	7,9,15	291	291	–	155
No	No	7,9,10	278	320	10.0	12

5.4.2 U.S. Network

We next test our methodology on the U.S. map: (i) a 49-node network with locations as the state capitals of the continental United States plus Washington, D.C.; and (ii) a 88-node network with the 49-node locations and 39 other largest cities in the United States. The correlated facility

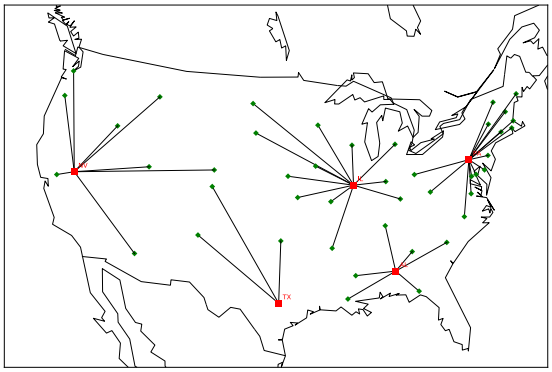
disruption profiles are as those specified in Section 4.5.3.

We first test our model and algorithm for $R = 10, 15, 20, 75 = |\mathcal{K}|$ using the 49-node dataset with the original system parameter values from Snyder and Daskin (2005)(case 49-I), and $R = 10, 20, 30, 129 = |\mathcal{K}|$ using the 88-node dataset. In addition, we run another set of instances (case 49-II) using the 49-node network but reducing the fixed facility costs to 1/3 of their original values in Snyder and Daskin (2005). The Lagrangian relaxation/B&B procedure is executed to a tolerance of 0.5%, or up to 3600 seconds in CPU time. The algorithm performance is summarized in Table 5.4, and the optimal facility locations and initial customer assignments are shown in Figure 5.3.

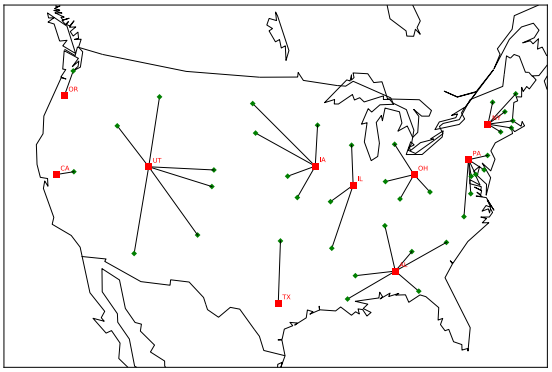
Table 5.4: Algorithm performance for the U.S. networks with 49 and 88 nodes.

Nodes	Pattern	R^e	Facility	Root UB	Root LB	Root gap (%)	Overall UB	Overall LB	Overall gap (%)	CPU time
49-I	Indp	5	9,22,26,38,46	891150	860404	3.450	887868	883507	0.491	96
	Corr	10	9,22,26,38,46	893049	834126	6.598	887881	883451	0.499	1557
		15	9,22,26,38,46	888855	835584	5.993	887868	883493	0.493	2373
		20	9,22,26,38,46	891150	834113	6.400	887868	883486	0.494	2568
		75	9,22,26,38,46	891150	834741	6.330	887868	879537	0.938	3600
49-II	Indp	5	7,9,16,22,26,36,38,47,48	586886	572538	2.445	586194	583291	0.495	654
	Corr	10	9,13,15,23,26,29,38,47,48	597735	554304	7.266	595102	570308	4.166	3600
		15	7,9,16,22,26,29,38,42,47,48	589864	554304	6.029	587996	569597	3.129	3600
		20	7,9,16,22,26,29,38,42,47,48	589657	554304	5.996	587794	568639	3.259	3600
		75	7,9,16,22,26,29,38,42,47,48	589122	554229	5.923	587793			3600
88	Indp	5	10,25,29,39,57,61,71,83,87	1242190	1200420	3.363	1242200	1221060	1.702	3600
	Corr	10	10,25,29,39,57,61,71,83,87	1254600	1112030	11.364	1254600	1150340	8.310	3600
		20	10,25,29,39,57,61,71,83,87	1242200	1115950	10.163	1242200	1150620	7.372	3600
		30	10,25,29,39,57,61,71,83,87	1244970	1107200	11.066	1242200	1150830	7.355	3600
		129	10,25,29,39,57,61,71,83,87	1244970	1102360	11.455	1242200	1140060	8.223	3600

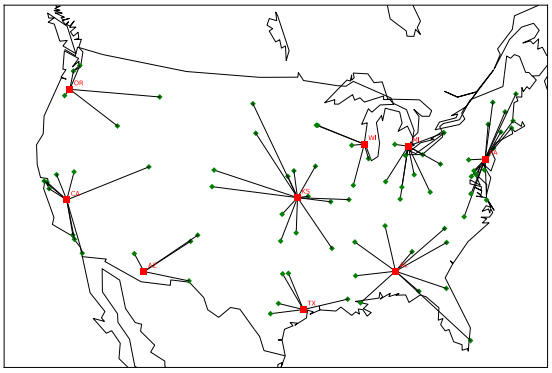
From Table 5.4 and Figure 5.3, we observe the following: First, our model and algorithm perform very well, especially on the 49-node cases, solving the first set (i.e., case 49 -I) to less than 1% gap and the second set (i.e., 49-II) to less than 5% gap within 1 hour. For the larger 88-node cases, we can still obtain 7%-8% gap within 1 hour computation time. Second, the maximum back-up level R does affect the location decision and optimal system cost when R is small; however, when R becomes large, the optimal facility locations are insensitive to it. This implies that we can set an arbitrary yet reasonably large value of R for many applications. Third, even with a sufficiently large R, i.e., $R = |\mathcal{K}|$, such that the station-based formulation is exactly equivalent to the scenario-based formulation, our algorithm still works quite effectively, as it is capable of providing solutions with a small optimality gap within a short amount of time. Finally, by comparing the second 49-node case set (i.e., 49-II) with independent and correlated facility disruptions, we can see that ignoring correlations in facility location models may lead to very sub-optimal system design.



(a) 49 nodes with input 1 (49-I)



(b) 49 nodes with input 2 (49-II)



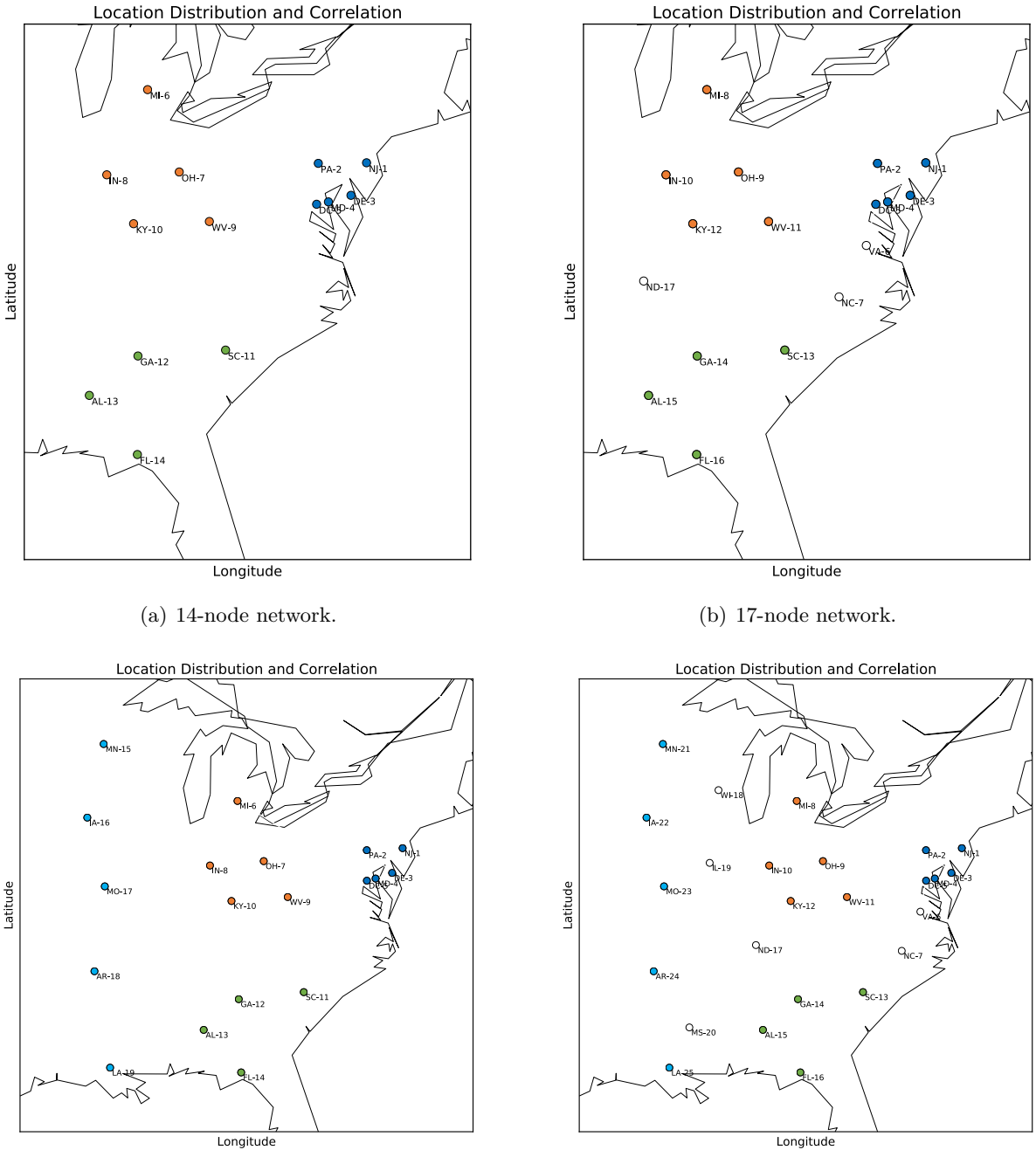
(c) 88 nodes

Figure 5.3: Facility location solutions for U.S. networks with 49 and 88 nodes.

5.4.3 Model Comparison

To better demonstrate the advantage of the proposed modeling framework, we further test both the scenario-based formulation (RFL-SCE) and station-based formulation (RFL-STA) on four networks, with 14, 17, 19, and 25 nodes, respectively. As shown in Figure 5.4, each of the four networks is part of the 49-node network in Figure 4.4, with the corresponding local correlation profiles presented in Table 4.10. For example, the three local areas in Figure 5.4(a) correspond to Local 2, Local 3, and Local 4 in Table 4.10, respectively, with the location indices adjusted accordingly. Independent stations (those not appearing in Table 4.10) each fail independently with probability 0.02.

Table 5.5 compares the results from (RFL-SCE) and (RFL-STA). It can be observed that for each case, the solutions (i.e., the final UB and location decisions) from the two formulations are exactly the same. However, our station-based formulation can be solved to optimality much



(a) 14-node network.

(b) 17-node network.

(c) 19-node network.

(d) 25-node network.

Figure 5.4: Network setup and correlation pattern for the four cases.

Table 5.5: Performance comparison between (RFL-SCE) and (RFL-STA).

Pattern	# of nodes	Scenario-based (RFL-SCE)			Station-based (RFL-STA)		
		UB	Facilities	Time (s)	UB	Facilities	Time (s)
No disruption	14	257581	PA-2,MI-6,FL-14	0.1	257581	PA-2,MI-6,FL-14	0.1
	17	304372	PA-2,MI-8,AL-15	0.2	304372	PA-2,MI-8,AL-15	0.3
	19	338555	PA-2,MI-6,AL-13	0.4	338555	PA-2,MI-6,AL-13	0.5
	25	426942	PA-2,IN-10,AL-15	0.5	426942	PA-2,IN-10,AL-15	0.9
Correlation	14	266412	PA-2,MI-6,AL-13	0.5	266412	PA-2,MI-6,AL-13	0.8
	17	313426	PA-2,MI-8,AL-15	35.8	313426	PA-2,MI-8,AL-15	1.6
	19	348426	PA-2,IN-8,AL-13	252.2	348426	PA-2,IN-8,AL-13	3.2
	25	–	–	–	436833	PA-2,AL-15,IL-19	6.7

more quickly than the scenario-based formulation, especially when correlations are present. In particular, the scenario-based formulation cannot even provide a feasible solution to the 25-node network due to exhaustion of computer memory (by the excessive number of scenarios). This comparison verifies the correctness and effectiveness of our station-based formulation (RFL-STA), as well as its clear superiority over the traditional scenario-based formulation (RFL-SCE), even for moderate-sized problem instances. We can easily project that the advantage will be even bigger for larger applications.

5.5 Proof of Propositions

5.5.1 Proof of Proposition 10

Proof. We first map an optimal solution to (RFL-STA) to a feasible solution to (RFL-SCE). Let $(\mathbf{X}, \mathbf{Y}, \mathbf{Z})$ be an optimal solution to (RFL-STA). We let $j(i,r) = j : Y_{ikjr} = 1, k(i,r) = k : Y_{ikjr} = 1, J(i,r) = \{j \in \mathcal{J} \cup \{0\} : j \neq j(i,r), \exists k,l \leq r-1, Y_{ikjl} = 1\}, \mathcal{R}_i = \{1\} \cup \{r > 1 : \exists j \neq j', k \neq k',$ s.t. $Y_{ikjr} = Y_{ik'j'r-1} = 1\}$, and for each $r \in \mathcal{R}$, we let $r_i(r) \in \{r' : r' \in \mathcal{R}_i, r' > r, r' \leq r'', \forall r'' > r\}$. We construct a solution $(\mathbf{X}', \mathbf{Y}')$ as follows

(i) $X'_j = X_j$;

(ii) $Y'_{ij\omega} = \begin{cases} 1, & \text{if } j = j(i,r) \text{ for some } r, \delta_{j\omega} = 1, \delta_{j'\omega} = 0, \forall j' \in J(i,r); \\ 0, & \text{otherwise.} \end{cases}$

By construction, $(\mathbf{X}', \mathbf{Y}')$ is a feasible solution to (RFL-SCE). In particular, for any customer $i \in \mathcal{I}$ and any scenario $\omega \in \Omega$, either there exists exactly one $j \in \mathcal{J}$ such that $j = j(i,r)$ for some r, and $\delta_{j\omega} = 1, \delta_{j'\omega} = 0, \forall j' \in J(i,r)$, or there exists no $j \in \mathcal{J}$ such that $j =$

$j(i,r)$ for some $r, \delta_{j\omega} = 1$. Hence, there exists exactly one $j \in \mathcal{J} \cup \{0\}$ such that $Y_{ij\omega} = 1, \forall i \in \mathcal{I}, \omega \in \Omega$, which implies that (5.1b) hold.

We next show that $(\mathbf{X}', \mathbf{Y}')$ achieves the same objective value as $(\mathbf{X}, \mathbf{Y}, \mathbf{Z})$. We denote $\Phi(\mathbf{X}, \mathbf{Y}, \mathbf{Z})$ and $\Psi(\mathbf{X}', \mathbf{Y}')$ as the objectives of (RFL-STA) and (RFL-SCE), respectively, and $\Omega(i,r) = \{\omega \in \Omega : Y'_{ij(i,r)\omega} = 1\}$. We have the following result

$$\begin{aligned}
\Psi(\mathbf{X}', \mathbf{Y}') &= \sum_{j\in\mathcal{J}} f_j X'_j + \sum_{i\in\mathcal{I}} \sum_{j\in\mathcal{J}\cup\{0\}} \sum_{\omega\in\Omega} \mu_i d_{ij} Y'_{ij\omega} p_\omega \\
&= \sum_{j\in\mathcal{J}} f_j X_j + \sum_{i\in\mathcal{I}} \sum_{r\in\mathcal{R}_i} \mu_i d_{ij(i,r)} \sum_{\omega\in\Omega(i,r)} p_\omega \\
&= \sum_{j\in\mathcal{J}} f_j X_j + \sum_{i\in\mathcal{I}} \sum_{r\in\mathcal{R}_i} \mu_i d_{ij(i,r)} \sum_{J:J(i,r)\subseteq J, j(i,r)\notin J} p_J^S.
\end{aligned}$$

Applying Equations (4.7) yields

$$\begin{aligned}
\sum_{J:J(i,r)\subseteq J, j(i,r)\notin J} p_J^S &= \sum_{J:J(i,r)\subseteq J, j(i,r)\notin J} \sum_{J_1:J\subseteq J_1} (-1)^{|J_1|-|J|} \left[\prod_{J_2:J_2\cap J_1\neq\emptyset} q_{k_{J_2}} \right] \\
&= \sum_{J:J(i,r)\subseteq J, j(i,r)\notin J} \sum_{J_1:J\subseteq J_1} (-1)^{|J_1|-|J|} \mathcal{A}(J_1) \\
&= \sum_{J:J(i,r)\subseteq J} C_J \mathcal{A}(J),
\end{aligned}$$

where $\mathcal{A}(J) = \prod_{J_2:J_1\cap J\neq\emptyset} q_{k_{J_1}}$ and C_J is the ultimate coefficient of $\mathcal{A}(J)$, which are

$$C_J = \begin{cases} 1, & \text{if } J = J(i,r); \\ -1, & \text{if } J = J(i,r) \cup \{j(i,r)\}; \\ \sum_{J'\subseteq J\setminus J(i,r)} (-1)^{|J\setminus J(i,r)|-|J'|} = \sum_{n=0}^{|J\setminus J(i,r)|} (-1)^n \binom{|J\setminus J(i,r)|}{n} = 0, & \text{otherwise .} \end{cases}$$

Therefore, we have

$$\begin{aligned}
\Psi(\mathbf{X}', \mathbf{Y}') &= \sum_{j\in\mathcal{J}} f_j X_j + \sum_{i\in\mathcal{I}} \sum_{r\in\mathcal{R}_i} \mu_i d_{ij(i,r)} \sum_{J:J(i,r)\subseteq J, j(i,r)\notin J} p_J^S \\
&= \sum_{j\in\mathcal{J}} f_j X_j + \sum_{i\in\mathcal{I}} \sum_{r\in\mathcal{R}_i} \mu_i d_{ij(i,r)} \left[\mathcal{A}(J(i,r)) - \mathcal{A}(J(i,r)\cup\{j(i,r)\}) \right]
\end{aligned}$$

$$
\begin{aligned}
&= \sum_{j\in\mathcal{J}} f_j X_j + \sum_{i\in\mathcal{I}} \sum_{r\in\mathcal{R}_i} \mu_i d_{ij(i,r)} \left[\prod_{J: J\cap J(i,r)\neq\emptyset} q_{k_J} - \prod_{J: J\cap (J(i,r)\cup\{j(i,r)\})\neq\emptyset} q_{k_J} \right] \\
&= \sum_{j\in\mathcal{J}} f_j X_j + \sum_{i\in\mathcal{I}} \sum_{r\in\mathcal{R}_i} \mu_i d_{ij(i,r)} \sum_{l=r}^{r_i(r)-1} \prod_{l'=1}^{l-1} q_{k(i,l')} \left(1 - q_{k(i,l)}\right) \\
&= \sum_{j\in\mathcal{J}} f_j X_j + \sum_{i\in\mathcal{I}} \sum_{r=1}^{R+1} \mu_i d_{ij(i,r)} \prod_{l=1}^{r-1} q_{k(i,l)} \left(1 - q_{k(i,r)}\right) \\
&= \sum_{j\in\mathcal{J}} f_j X_j + \sum_{i\in\mathcal{I}} \sum_{j\in\mathcal{J}\cup\{0\}} \sum_{k\in\mathcal{K}\cup\{0\}} \sum_{r=1}^{R+1} \mu_i d_{ij} Y_{ikjr} Z_{ikjr} \\
&= \Phi(\mathbf{X}, \mathbf{Y}, \mathbf{Z}),
\end{aligned}
$$

which implies that the optimal solution to (RFL-SCE) is a lower bound to (RFL-STA).

Conversely, we map an optimal solution to (RFL-SCE) to a feasible solution to (RFL-STA). Given an optimal solution $(\mathbf{X}, \mathbf{Y})$ to (RFL-SCE), without loss of generality, we assume that each customer always visits its closest open facility for service, and if there exist more than one facility with equal distance, we break the tie by choosing the facility based on index: let $J^* = \{j \in \mathcal{J} : X_j = 1\}$, for each customer i, let $j^i_1, j^i_2, \cdots, j^i_{|J^*|+1}$ be an ordering of the facilities in $J^* \cup \{0\}$ such that for all $2 \leq r \leq |J^*| + 1$, $d_{ij^i_{r-1}} \leq d_{ij^i_r}$ and if $d_{ij^i_{r-1}} = d_{ij^i_r}$, $j^i_{r-1} < j^i_r$. Since a facility is functioning if and only if at least one of its connected stations is operating, we let $K^* = \{k \in \mathcal{K} : \exists j \in J^*, l_{jk} = 1\}$, and $r^i_n = |\{k \in K^* : \exists n_1 < n, l_{j^i_{n_1}} = 1\}|$ be the total number of stations that are connected to at least one facility in $\{j^i_1, \cdots, j^i_{n-1}\}$, we know that facility j^i_n is visited by i if only if all facilities in $\{j^i_1, \cdots, j^i_{n-1}\}$ are unavailable (i.e, all the r^i_n stations are disrupted). Then we define two sequences of facilities and stations respectively as: (i) $j(i,1), j(i,2), \cdots, j(i,|K^*|+1)$ such that $j(i,r) \in \{j : r_j < r \leq r_{j+1}\}$; and (ii) $k(i,1), k(i,2), \cdots, k(i,|K^*|+1)$ such that $k(i,r) \in K(i,r) = \{k : l_{j(i,r)k} = 1, l_{j(i,l)k} = 0, \forall l < r, j(i,l) \neq j(i,r)\}$, and if $K(i,r-1) = K(i,r), k(i,r-1) < k(i,r)$. We construct a solution $(\mathbf{X}', \mathbf{Y}', \mathbf{Z}')$ as follows

(i) $X'_j = X_j$;

(ii) $Y'_{ikjr} = \begin{cases} 1, & \text{if } j = j(i,r), k = k(i,r), d_{ij} \leq d_{i0}; \\ 0, & \text{otherwise;} \end{cases}$

(iii) $Z'_{ikjr} = \begin{cases} \left(1 - q_{k(i,r)}\right) \prod_{l=1}^{r-1} q_{k(i,l)}, & \text{if } j = j(i,r), k = k(i,r), d_{ij} \le d_{i0}; \\ 0, & \text{otherwise.} \end{cases}$

By examining the constraint sets in (RFL-STA), we observe that $(\mathbf{X}', \mathbf{Y}', \mathbf{Z}')$ is a feasible solution to (RFL-STA). We next show that $(\mathbf{X}', \mathbf{Y}', \mathbf{Z}')$ achieves the same objective value as $(\mathbf{X}, \mathbf{Y})$. We let $\mathcal{R}_i = \{r_1^i + 1, r_2^i + 1, \cdots, r_{|J^*|+1}^i + 1\}$, for each $r \in \mathcal{R}_i$, we let $r_i(r) \in \{r' : r' \in \mathcal{R}_i, r' > r, r' \le r'', \forall r'' > r\}$, and $\Omega(i,r) = \{\omega \in \Omega : \delta_{j(i,r)\omega} = 1, \delta_{j(i,l)\omega} = 0, \forall l < r, j(i,l) \ne j(i,r)\}$

$$
\begin{aligned}
\Phi(\mathbf{X}', \mathbf{Y}', \mathbf{Z}') &= \sum_{j \in \mathcal{J}} f_j X'_j + \sum_{i \in \mathcal{I}} \sum_{j \in \mathcal{J} \cup \{0\}} \sum_{k \in \mathcal{K} \cup \{0\}} \sum_{r=1}^{R+1} \mu_i d_{ij} Y'_{ikjr} Z'_{ikjr} \\
&= \sum_{j \in \mathcal{J}} f_j X_j + \sum_{i \in \mathcal{I}} \sum_{r=1}^{R+1} \mu_i d_{ij(i,r)} \prod_{l=1}^{r-1} q_{k(i,l)} \left(1 - q_{k(i,r)}\right) \\
&= \sum_{j \in \mathcal{J}} f_j X_j + \sum_{i \in \mathcal{I}} \sum_{r \in \mathcal{R}_i} \mu_i d_{ij(i,r)} \sum_{l=r}^{r_i(r)-1} \prod_{l'=1}^{l-1} q_{k(i,l')} \left(1 - q_{k(i,l)}\right) \\
&= \sum_{j \in \mathcal{J}} f_j X_j + \sum_{i \in \mathcal{I}} \sum_{r \in \mathcal{R}_i} \mu_i d_{ij(i,r)} \sum_{J : J(i,r) \subseteq J, j(i,r) \notin J} p_J^S \\
&= \sum_{j \in \mathcal{J}} f_j X_j + \sum_{i \in \mathcal{I}} \sum_{r \in \mathcal{R}_i} \mu_i d_{ij(i,r)} \sum_{\omega \in \Omega(i,r)} p_\omega \\
&= \sum_{j \in \mathcal{J}} f_j X_j + \sum_{i \in \mathcal{I}} \sum_{j \in \mathcal{J} \cup \{0\}} \sum_{\omega \in \Omega} \mu_i d_{ij} Y_{ij\omega} p_\omega \\
&= \Psi(\mathbf{X}, \mathbf{Y}).
\end{aligned}
$$

Therefore, the optimal solution to (RFL-STA) is also a lower bound to (RFL-SCE), implying that the optimal solutions to (RFL-STA) and (RFL-SCE) are exactly the same. This completes our proof. □

5.5.2 Proof of Proposition 12

Proof. Suppose, for a contradiction, that $(\mathbf{X}, \mathbf{Y}, \mathbf{Z})$ is optimal to (RFL-STA) but violates Property II, i.e., there exist i, j, k_1, k_2, r such that $X_j = 1, l_{k_1 j} = l_{k_2 j} = 1, q_{k_1} \le q_{k_2}, Y_{ik_2jr} = 1, Y_{ik_1jr'} = 0, \forall r' < r$. We will show that by replacing k_2 with k_1 the objective of (RFL-STA) will decrease. We simply construct a different solution $(\mathbf{X}', \mathbf{Y}', \mathbf{Z}')$ as follows:

(i) $X'_j = X_j$;

$$\text{(ii)}\ \ Y'_{hkls} = \begin{cases} 1, & \text{if } (h,k,l,s) = (i,k_1,j,r); \\ 0, & \text{if } (h,k,l,s) = (i,k_2,j,r); \\ Y_{hkls}, & \text{otherwise;} \end{cases}$$

$$\text{(iii)}\ \ Z'_{hkls} = \begin{cases} \frac{1-q_{k_1}}{1-q_{k_2}} Z_{hk_2ls}, & \text{if } (h,k,l,s) = (i,k_1,j,r); \\ 0, & \text{if } (h,k,l,s) = (i,k_2,j,r); \\ \frac{q_{k_1}}{q_{k_2}} Z_{hkls}, & \text{if } s > r; \\ Z_{hkls}, & \text{otherwise.} \end{cases}$$

By construction, $(\mathbf{X}', \mathbf{Y}', \mathbf{Z}')$ is a feasible solution to (RFL-STA). We use $\mathbf{\Phi}(\mathbf{X}, \mathbf{Y}, \mathbf{Z})$ to denote the objective value of (RFL-STA) associated with $(\mathbf{X}, \mathbf{Y}, \mathbf{Z})$, assume that $Y_{ik^{ir}j^{ir}r} = Y_{i00R^i} = 1$, it follows that

$$\begin{aligned} \mathbf{\Phi}(\mathbf{X}, \mathbf{Y}, \mathbf{Z}) - \mathbf{\Phi}(\mathbf{X}', \mathbf{Y}', \mathbf{Z}') &= \mu_i d_{ij} Z_{k_2jr} + \sum_{s=r+1}^{R^i} \mu_i d_{il^{is}} Z_{k^{is}l^{is}s} - (\mu_i d_{ij} Z'_{k_1jr} + \sum_{s=r+1}^{R^i} \mu_i d_{il^{is}} Z'_{k^{is}l^{is}s}) \\ &= \frac{\mu_i Z_{k_2jr}}{1-q_{k_2}} \Bigg[d_{ij}(1-q_{k_2}) + q_{k_2} \sum_{s=r+1}^{R^i} d_{il^{is}} \prod_{s'=r+1}^{s-1} q_{k^{is'}} (1-q_{k^{is}}) \\ &\qquad - \left(d_{ij}(1-q_{k_1}) + q_{k_1} \sum_{s=r+1}^{R^i} d_{il^{is}} \prod_{s'=r+1}^{s-1} q_{k^{is'}} (1-q_{k^{is}}) \right) \Bigg] \\ &= \frac{\mu_i Z_{k_2jr}}{1-q_{k_2}} (q_{k_2} - q_{k_1}) \left(-d_{ij} + \sum_{s=r+1}^{R^i} d_{il^{is}} \prod_{s'=r+1}^{s-1} q_{k^{is'}} (1-q_{k^{is}}) \right) \end{aligned}$$

As $q_{k^{iR^i}} = 0$, and $d_{il^{is}} \le d_{il^{i,s+1}}$ from Proposition 11, we have

$$\begin{aligned} \sum_{s=R_i-1}^{R^i} d_{il^{is}} \prod_{s'=r+1}^{s-1} q_{k^{is'}} (1-q_{k^{is}}) &\ge d_{il^{iR^i-1}} \prod_{s'=r+1}^{R^i-2} q_{k^{is'}} (1-q_{k^{iR^i-1}}) + d_{il^{iR^i-1}} \prod_{s'=r+1}^{R^i-1} q_{k^{is'}} \\ &= d_{il^{iR^i-1}} \prod_{s'=r+1}^{R^i-2} q_{k^{is'}} \end{aligned}$$

By induction, we can conclude that

$$-d_{ij} + \sum_{s=r+1}^{R^i} d_{il^{is}} \prod_{s'=r+1}^{s-1} q_{k^{is'}} (1-q_{k^{is}}) \ge -d_{ij} + d_{il^{ir+1}} \ge 0.$$

Since $\frac{\mu_i Z_{k_2 jr}}{1-q_{k_2}} \geq 0$, $q_{k_2} - q_{k_1} \geq 0$, we have $\mathbf{\Phi}(\mathbf{X}, \mathbf{Y}, \mathbf{Z}) \geq \mathbf{\Phi}(\mathbf{X}', \mathbf{Y}', \mathbf{Z}')$, which implies that $\mathbf{\Phi}(\mathbf{X}, \mathbf{Y}, \mathbf{Z})$ is not optimal. This completes the proof. □

CHAPTER 6:

RELIABLE SENSOR DEPLOYMENT FOR POSITIONING AND SURVEILLANCE VIA TRILATERATION

High-accuracy object positioning has been playing a critical role in various application contexts such as vehicle navigation, activity tracking, regional surveillance, search/rescue missions, etc. In recent years, massive availability of mobile devices has stimulated demand for many location-aware applications. In a typical such system, multiple sensors are installed to provide coverage jointly, and trilateration is one of the most popular mathematical techniques used by the system to geographically positioning or surveil objects. The effectiveness of the system highly depends on the quality (working range and precision level) and quantity of sensor coverage in the local area.[a]

Installed sensors work in combinations to provide sensory converage. Considering possible sensor disruptions, functionality of a sensor combination could be interrelated to that of another combination, which leads to internal correlations among different sensor combinations. In this case, where to deploy sensors, which sensor combinations to use, and in what sequence and probability to use sensor combinations in case of sensor disruptions, are nontrivial questions.

In this chapter, we incorporate the impacts of sensor disruptions into a reliable sensor deployment framework by extending the ideas of assigning backup sensors as well as correlation decomposition via supporting stations. Specifically, we formulate the reliable sensor deployment problem as a compact mixed-integer linear program and develop a customized Lagrangian relaxation based algorithm with several embedded approximation subroutines.

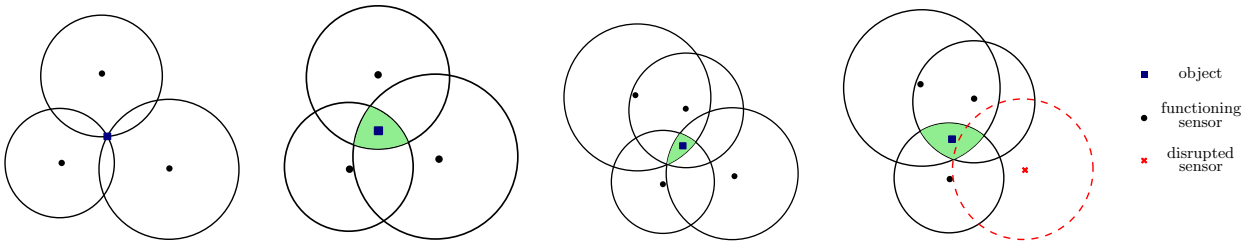
[a]This chapter is based on a published paper, An et al. (2017).

6.1 Introduction

High-accuracy object positioning has been playing a critical role in various application contexts such as: (i) civilian uses, including vehicle navigation, driver guidance, activity tracking; (ii) industrial uses, such as aircraft tracking, regional surveillance, extrasolar planets detection; and (iii) military uses, such as search/rescue missions, missile and projectile guidance. In recent several decades, many object positioning systems with different infrastructures and architectures have been developed. Examples include the Global Positioning System (GPS), cellular phone based systems, computer vision systems (Krumm et al., 2000) – each has its own property, configuration, and reliability. Among them, GPS is the most popular and widely-used system. However, location detection is generally less accurate by GPS inside a blocked space (e.g., inside a building or an underground area) due to attenuation or blockage of satellite or phone signals. In recent years, massive availability of mobile devices has stimulated demand for indoor location-aware applications, including in-building guidance and location-based advertising in shopping malls, elderly navigation in nursing homes, hazardous materials detection in airports, and on-site workforce tracking (policemen, firefighters). In many of these applications, object positioning and tracking are used for surveillance purposes such that situations in all neighborhoods can be covered by tracking some of the objects (people, packages, etc) in those neighborhoods. In some other contexts, the tracking of objects require that all neighborhoods be covered at all times (e.g., there cannot be blind spots). To enable indoor positioning or surveillance, most mobile devices are equipped with sensors, signal transmitters and receivers in order to collect various types of sensing signals via short-range radio waves, Wi-Fi, Bluetooth, or magnetic fields. These signals/data are then used to compute the location of the target objects. For example, Filonenko et al. (2013) presented an accurate indoor positioning approach that combines ultrasound from mobile phones with time-difference-of-arrival asynchronous trilateration methods.

Trilateration is one of the most popular mathematical techniques used by many systems to geographically determine the position of an object. Based on simultaneous distance measurements from three known sites, it essentially solves a geometry problem of finding the intersection of three spheres (Thomas and Ros, 2005; Doukhnitch et al., 2008). For example, in an indoor Wi-Fi positioning system, received signal strengths from all existing Wi-Fi access points are gathered and

converted to distance measurements using existing signal propagation models (each positioning system may have its own distance calculation method). Then, a trilateration algorithm is used to pinpoint the object's location, typically based on distances data from three different sensors, as shown in Fig. 6.1(a). In effect, as shown in Fig. 6.1(b), due to signal scattering and blockage, the collected or calculated distance information may be inaccurate, and the error (illustrated by the size of the shaded area) normally increases with the distance between the sensor and the object. In such cases, a location cannot be precisely identified from three distance measures. Navidi et al. (1998) used statistical methods to quantify the uncertainty in trilateration results from the potential error of distance measurements. It is worth noting that the positioning error can be reduced if accurate information from additional sensors is used, as indicated by the smaller green area in Fig. 6.1(c). Therefore, the effectiveness of an indoor positioning system highly depends on the quality (working range and precision level) and quantity of sensor coverage in the local area. Nevertheless, the number and location of sensors should be carefully determined, because high-precision sensors could be expensive to deploy, the system architecture becomes more complex, and the trilateration algorithm requires more computation time when more sensors are used. We would like to achieve the best positioning accuracy with a reasonable investment in sensor installations.



(a) ideal case (b) three sensors (c) multiple sensors (d) disruption scenario

Figure 6.1: Position error illustration in trilateration.

In reality, installed sensors are subject to operational disruptions from time to time due to technological defects, adverse environmental conditions, deliberate sabotages, etc. If a sensor is disrupted, no useful information can be collected, and the effectiveness of object positioning may be affected. This is illustrated in Fig. 6.1(d): if three sensors are needed to "cover" an object or an area via trilateration, when a first-choice sensor is disrupted, a more remote sensor will be used and yet this may yield a larger error (as compared to the scenario in Fig. 6.1(b)). Therefore, the

impacts of sensor disruptions should be considered in a reliable sensor deployment framework such that the overall expected error across the normal non-failure scenario and various sensor failure scenarios is minimized. There have been several research efforts on reliable facility location in the context of logistics networks (Snyder, 2006; Cui et al., 2010; Li and Ouyang, 2010; Xie et al., 2018a). More recently, Li and Ouyang (2011, 2012) investigated reliable traffic sensor deployment in a discrete transportation network to estimate OD flow volume, congestion state, and path travel time. Adjacent sensors along each flow path pair up to monitor the road segment in between, and yet sensors could be disrupted with site-dependent probabilities. Existing studies, therefore, have considered reliable facility/sensor deployment when sensors work either independently or in pairs. To the best of our knowledge, no studies have considered reliable sensor location problem in the context of trilateration, where at least three sensors are required to work together to position a target (or to cover an area).

This paper aims to fill this gap by incorporating the impacts of sensor disruptions into a reliable sensor deployment framework for positioning or surveillance via trilateration. An object is positioned based on distance measurements received from a combination of at least three sensors. Since various sensor combinations could share some common unreliable sensors, failure of a combination could be directly related to that of another combination. This leads to internal correlation among the functionality of sensor combinations. In this case, where to deploy sensors, which combinations of sensors to use, and in what sequence and probability to use backup combinations in case of disruptions, are nontrivial questions. It remains an open challenge to optimize sensor deployment locations that maximize the overall system-wide surveillance or positioning benefits under the risk of site-dependent sensor failures. In this paper, we address the problem by combining and extending the ideas of assigning back-up sensors (Li and Ouyang, 2010, 2011, 2012) as well as correlation decomposition via supporting stations (Li et al., 2013; Xie et al., 2015, 2018a). A compact mixed-integer mathematical model is developed to determine the optimal sensor location, sensor level assignment and combination selection plans. A customized solution algorithm based on Lagrangian relaxation and branch-and-bound is developed, together with an embedded approximation subroutine for sub-problems. A series of hypothetical and empirical case studies are conducted to illustrate the applicability and performance of the proposed methodology.

The remainder of this chapter is organized as follows. Section 6.2 introduces the methodology

we develop for the reliable sensor deployment problem, including the effectiveness measurements and the model formulation. Section 6.3 presents the solution algorithm. Section 6.4 demonstrates the applicability of the model and solution algorithm on a series of examples.

6.2 Model Formulation

We consider an area (e.g., airport, shopping mall) which contains a set of spatial neighborhoods $I := \{i\}$ that need surveillance coverage. In airports, such neighborhoods can be security check gates, boarding gates and restaurants where accidents are more likely to occur due to crowds' gathering. Each point $i \in I$ attracts v_i customers per day. Let J be the set of candidate locations for potential sensor installations. At most one sensor can be installed at each location $j \in J$ at a construction cost f_j. Let d_{ij} denote the distance from surveillance neighborhood i to sensor location j. A sensor located at j could be disrupted with a probability of p_j. For a neighborhood $i \in I$, the sensors are assigned with different backup levels. We assume the receiver (can be the mobile device/object itself) always uses N, where $N \geq 3$, sensors with the lowest backup levels to calculate the position of the object. Without loss of generality, for modeling convenience, N dummy sensors (located at $|J|+1, \cdots, |J|+N$) are added to the system to ensure there are always at least N sensors available even under the worst case scenario in which all sensors are disrupted. Let $\tilde{J}$ be the set of dummy sensors and $\mathcal{J} = J \cup \tilde{J}$ be the set of all sensors. The dummy sensors incur 0 installation cost and are not subject to failure, but make no contribution to object positioning. Let K be the set of candidate sensor combinations to locate customers. Each combination $k \in K$ contains exactly N sensors (including the dummy ones) and could monitor i with accuracy e_{ik}. Let α be the monetary value of sensing accuracy. We introduce incidence matrix $\left\{a_{kj}\right\}$ to represent the mapping relationships between combinations and sensors, where $a_{kj} = 1$ if combination k contains sensor j, or 0 otherwise. The maximum number of combinations is $\sum_{t=0}^{N} \binom{|J|}{t}$, where t indicates that t regular sensors and $N - t$ dummy sensors are used in the combination.

As such, the receiver/object will search from the sensor with the lowest backup level until N sensors have been found. The key decision variables $\mathbf{X} := \{X_j\}$ determine sensor locations, where $X_j = 1$ if a sensor is installed at location j or $X_j = 0$ otherwise. For each surveillance neighborhood, the installed sensors are assigned to it at different levels. Variables $\mathbf{Z} := \{Z_{ijr}\}$ determine the relative sensor levels, where binary variable $Z_{ijr} = 1$ if sensor j is installed and is assigned with level

r to neighborhood i, or 0 otherwise; $\mathbf{Y} := \{Y_{ikr}\}$ denote the sensor combination assignment to the customers, where $Y_{ikr} = 1$ if neighborhood i uses combination k whose highest level element sensor has level r, or 0 otherwise. Note that a combination k corresponds to only one level r, while there may exist multiple combinations corresponding to the same level r. The backup levels are initially assigned to the sensors. A level r, if associated with a sensor combination k, indicates the highest level of any sensor contained in k; it can be uniquely determined from the backup levels of sensors that are assigned to an object, i.e., $\{Z_{ijr}\}$. $\mathbf{P} := \{P_{ikr}\}$ are quasi-probability variables where P_{ikr} defines the probability to use combination k to monitor neighborhood i if $Y_{ikr} = 1$, and is a state variable if $Y_{ikr} = 0$.

This sensor location problem (SLP) can now be formulated as the following mixed-integer nonlinear program:

$$\text{(SLP)} \quad \min_{\mathbf{X},\mathbf{Y},\mathbf{Z},\mathbf{P}} \quad \sum_{j\in J} f_j X_j - \alpha \sum_{i\in I}\sum_{k\in K}\sum_{r=1}^{|\mathcal{J}|} v_i e_{ik} P_{ikr} Y_{ikr} \tag{6.1a}$$

$$\text{s.t.} \quad \sum_{r=1}^{|\mathcal{J}|} Z_{ijr} \le X_j, \ \forall j \in J, i \in I, \tag{6.1b}$$

$$\sum_{r=1}^{|\mathcal{J}|} Z_{ijr} \le 1, \ \forall j \in J, i \in I, \tag{6.1c}$$

$$\sum_{r=1}^{|\mathcal{J}|} Z_{ijr} = 1, \ \forall j \in \tilde{J}, i \in I, \tag{6.1d}$$

$$Z_{ijr} = Z_{i,j+1,r+1}, \ \forall r = 1,2,\cdots,|\mathcal{J}|-1, j \in \tilde{J}\backslash(|J|+N), i \in I, \tag{6.1e}$$

$$\sum_{j\in J} Z_{ijr} + \sum_{s=1}^{r} Z_{i,|J|+1,s} = 1, \ \forall r = 1,\cdots,|\mathcal{J}|, i \in I, \tag{6.1f}$$

$$Y_{ikr} \le \frac{1}{N}\sum_{j\in\mathcal{J}}\sum_{s=1}^{r} a_{kj} Z_{ijs}, \ \forall k \in K, r = 1,\cdots,|\mathcal{J}|, i \in I, \tag{6.1g}$$

$$Y_{ikr} \le \sum_{j\in\mathcal{J}} a_{kj} Z_{ijr}, \ \forall k \in K, r = 1,\cdots,|\mathcal{J}|, i \in I, \tag{6.1h}$$

$$P_{ikr} = \sum_{j\in\mathcal{J}} \left(p_j\right)^{1_{[j\in J]}} Z_{ijr} P_{ik,r-1}, \ \forall k \in K, r = 1,\cdots,|\mathcal{J}|, i \in I, \tag{6.1i}$$

$$P_{ik0} = \prod_{j\in J} \left(1-p_j\right)^{a_{kj}} \left(p_j\right)^{-a_{kj}}, \ \forall k \in K, i \in I, \tag{6.1j}$$

$$X_j, Z_{ijr}, Y_{ikr} \in \{0,1\}, \ \forall k \in K, j \in \mathcal{J}, r = 1,\cdots,|\mathcal{J}|, i \in I. \tag{6.1k}$$

The objective function (6.1a) presents the expected system cost including the sensor installation cost and the expected total inaccuracy penalty, where $P_{ikr}Y_{ikr}$ is the probability that combination k is used by neighborhood i at the r-th level. Constraints (6.1b) enforce that customers can only use installed sensors. Constraints (6.1c) indicate that for a certain surveillance neighborhood, each regular sensor can only be assigned to at most one level. Constraints (6.1d) ensure for a certain surveillance neighborhood, each dummy sensor must be assigned to it at a certain backup level. The same dummy sensor could be assigned to other surveillance neighborhoods at different levels. Constraints (6.1e) postulate that if a dummy sensor $j \in \tilde{J}$ is assigned to surveillance neighborhood i at level r, then dummy sensor $j+1$ must be assigned to i at level $r+1$. Constraints (6.1f) require that at each level r, a surveillance neighborhood i either uses a regular sensor, or it has used the first dummy sensor at level $s \leq r$. Constraints (6.1g) enforce that combination k is available to surveillance neighborhood i only if the N sensors in k are all installed. Constraints (6.1h) require that combination k is available to surveillance neighborhood i when its highest level element serves at level r. Constraints (6.1i) and (6.1j) recursively define the assignment probability P_{ikr} for $Y_{ikr} = 1$ to happen, where the indicator function $1_{[\cdot]} = 1$ when condition $[\cdot]$ holds, or 0 otherwise. Please note that P_{ikr} does not have physical meaning when $Y_{ikr} = 0$ and its value will not affect the value of the objective function. Given that the lower level sensors are used earlier, a combination k is used if and only if its element sensors are all functioning, and the other constructed sensors which has level lower than the highest level in k are all disrupted. The derivation of P_{ikr} is shown as follows.

Proposition 14. *The assignment probability P_{ikr} (for $Y_{ikr} = 1$ to happen) can be calculated recursively by* (6.1i)*, given its initial state value defined by* (6.1j)*.*

Proof. We substitute P_{ikr-1} in the right hand side of (6.1i) by a function of P_{ikr-2} and repeat this procedure for the new right hand side value until $r = 0$. The quasi-probability of neighborhood i using combination k at level r can be written as

$$P_{ikr} = \frac{\prod_{j\in J}\left(1-p_j\right)^{a_{kj}}}{\prod_{j\in J} p_j^{a_{kj}}} \prod_{s\leq r}\left[\sum_{j\in\tilde{J}} Z_{ijs}\left(p_j\right)^{1_{[j\in J]}}\right]. \tag{6.2}$$

When combination k contains a dummy sensor, the indicator function excludes the dummy sen-

sor from the P_{ikr} calculation so as to prevent the case $P_{ikr} = 0, \forall r$, from happening. If $Y_{ikr} = 1$, the element sensors in combination k must be all functioning and the other constructed sensors which has a backup level lower than r are all disrupted. In this case, the probability that combination k is used by neighborhood i can thus be written as $\prod_{j\in J}(1-p_j)^{a_{kj}} \prod_{s\le r}\left[\sum_{j\in\mathcal{J}} Z_{ijs}(p_j)^{1-a_{kj}}\right]$, which can be shown to equal (6.2) after some simple algebraic calculation. Namely, the quasi-probability calculated by (6.2) is equivalent to the assignment probability that neighborhood i uses combination k (i.e., when $Y_{ikr} = 1$).

Since $Y_{ik\bar{r}} = 0, \forall \bar{r} \neq r$ must hold if $Y_{ikr} = 1$, the corresponding $P_{ik\bar{r}}$ only serves as a state variable (not a real probability), which will facilitate the calculation of the actual assignment probability P_{ikr} given that $Y_{ikr} = 1$. However, the value of the non-probability variable $P_{ik\bar{r}}$ will not affect SLP as $Y_{ik\bar{r}}P_{ik\bar{r}} = 0$ when $Y_{ik\bar{r}} = 0$ in the objective function (6.1a).

This recursive formula (6.1i), together with the initial value P_{ik0} specified by (6.1j) jointly define the quasi-probability in (6.2), which is also equivalent to the actual assignment probability P_{ikr} for $Y_{ikr} = 1$ to happen. □

In (SLP), the surveillance neighbourhood i can choose any installed sensors and assign them to various levels flexibly to minimize the inaccuracy penalty. However, at optimality, each neighbourhood i will use all installed sensors, and the backup level of each sensor solely depends on its relative distance to the neighbourhood (i.e., irrelevant to its failure probability), as proved in the following proposition.

Proposition 15. *In any optimal solution* $\{\mathbf{X}, \mathbf{Z}, \mathbf{Y}\}$*, for each surveillance neighborhood* i*, an installed sensor must be assigned to a backup level, and a nearer sensor must be assigned to an earlier level; i.e. the following two properties must hold (i) if* $X_j = 1$*, then* $\sum_{r=1}^{|\mathcal{J}|} Z_{ijr} = 1$*; (ii) if* $Z_{ij_1r} = Z_{ij_2r+1} = 1$ *for some* i, r*, then* $d_{ij_1} \le d_{ij_2}$*.*

Proof. We first prove property (i): suppose sensors J' are installed and are used to monitor neighbourhood i with an accuracy contribution of C_1. Suppose the newly installed sensor j where $j \notin J'$ is assigned to the last level $|J'|+1$ (i.e., $Z_{ij,|J'|+1} = 1$) and j together with J' provide a total accuracy contribution of C_2. By doing so, all combinations generated by J' and their probabilities to happen are unchanged in the new system $j \cup J'$. The sensor j brings new combinations and thus additional accuracy contribution. (i.e., $Y_{ik,|J'|+1} \neq 0$ for some k). Hence we have $C_2 \ge C_1$. Since assigning

sensor j to the last level $|J'|+1$ is only a feasible solution to the level assignment problem for sensors $j \cup J'$, C_2 is a lower bound to the maximum accuracy C_3 provided by sensors $j \cup J'$. Hence we have $C_3 \geq C_2 \geq C_1$. As such, a sensor must be assigned at a certain level to monitor the neighbourhoods once it is installed.

We now prove the second property (ii) by contradiction. Suppose there exist i, j_1, j_2 and r such that $Z_{ij_1r} = Z_{ij_2r+1} = 1$ and $d_{ij_1} > d_{ij_2}$. We consider a combination k_{Aj_1} where j_1 is its highest level element sensor (j_1 has level r) and A represent the other element sensors in combination k. Assume that the probability for sensors A to be functioning and all the remaining sensors assigned at levels 1 to $r-1$ to be disrupted is P_{r-1}. The expected accuracy contribution by k_{Aj_1} can be written as $e_{ik_{Aj_1}} P_{r-1}(1-p_{j_1})$. We consider another combination k_{Aj_2} constituted by sensors $A \cup j_2$. Given that j_2 is assigned to level $r+1$, the expected accuracy contribution by k_{Aj_2} is $e_{ik_{Aj_2}} P_{r-1} p_{j_1}(1-p_{j_2})$. Then the expected accuracy contribution associated with these two backup levels r and $r+1$ is $C_4 = \sum_A e_{ik_{Aj_1}} P_{r-1}(1-p_{j_1}) + e_{ik_{Aj_2}} P_{r-1} p_{j_1}(1-p_{j_2})$. Consider another solution $\{\mathbf{X}, \mathbf{Z}', \mathbf{Y}'\}$ where $Z'_{ij_2r} = Z'_{ij_1r+1} = 1$. The associated expected accuracy contribution is $C_5 = \sum_A e_{ik_{Aj_2}} P_{r-1}(1-p_{j_2}) + e_{ik_{Aj_1}} P_{r-1} p_{j_2}(1-p_{j_1})$. Since $e_{ik_{Aj_1}} < e_{ik_{Aj_2}}$ holds according to $d_{ij_1} > d_{ij_2}$, simple algebra shows that $C_5 > C_4$ and the expected accuracy contribution for assignments $\mathbf{Y}$ at other levels are exactly the same in these two solutions. As such, the latter solution $\{\mathbf{X}, \mathbf{Z}', \mathbf{Y}'\}$ is better than $\{\mathbf{X}, \mathbf{Z}, \mathbf{Y}\}$, which poses a contradiction. This completes the proof. □

The current model is nonlinear due to the existence of nonlinear terms $P_{ikr}Y_{ikr}$ in (6.1a) and $Z_{ijr}P_{ikr-1}$ in (6.1i). Linearization techniques introduced by Sherali and Alameddine (1992) (similar to those in Li and Ouyang (2012)) can be applied: i.e., we replace each $P_{ikr}Y_{ikr}$ and $Z_{ijr}P_{ikr-1}$ by new continuous variables W_{ikr} and V_{ikjr}, respectively, and enforce their equivalence by adding the following sets of constraints where M_k is the maximum value of $\mathbf{P_{ikr}}$ with $M_k = \prod_{j \in J}(1-p_j)^{a_{kj}} p_j^{-a_{kj}}$.

$$W_{ikr} \leq P_{ikr} + M_k\,(1 - Y_{ikr})\,,\ \forall k \in K, r = 1, \cdots, |\mathcal{J}|, i \in I, \tag{6.3a}$$

$$W_{ikr} \geq P_{ikr} + M_k\,(Y_{ikr} - 1)\,,\ \forall k \in K, r = 1, \cdots, |\mathcal{J}|, i \in I, \tag{6.3b}$$

$$W_{ikr} \leq M_k Y_{ikr},\ \forall k \in K, r = 1, \cdots, |\mathcal{J}|, i \in I, \tag{6.3c}$$

$$W_{ikr} \geq -M_k Y_{ikr},\ \forall k \in K, r = 1, \cdots, |\mathcal{J}|, i \in I, \tag{6.3d}$$

$$V_{ikjr} \le P_{ikr-1} + M_k \left(1 - Z_{ijr}\right),\ \forall k \in K, j \in \mathcal{J}, r = 1, \cdots, |\mathcal{J}|, i \in I, \tag{6.3e}$$

$$V_{ikjr} \ge P_{ikr-1} + M_k \left(Z_{ijr} - 1\right),\ \forall k \in K, j \in \mathcal{J}, r = 1, \cdots, |\mathcal{J}|, i \in I, \tag{6.3f}$$

$$V_{ikjr} \le M_k Z_{ijr},\ \forall k \in K, j \in \mathcal{J}, r = 1, \cdots, |\mathcal{J}|, i \in I, \tag{6.3g}$$

$$V_{ikjr} \ge -M_k Z_{ijr},\ \forall k \in K, j \in \mathcal{J}, r = 1, \cdots, |\mathcal{J}|, i \in I. \tag{6.3h}$$

The original (SLP) is then transformed into the following mixed integer linear program, which we call the linearized sensor location problem (LSLP). It remains an NP hard problem, but small instances can be readily solved by existing solvers (such as CPLEX).

$$\text{(LSLP)} \quad \min_{\mathbf{X},\mathbf{Y},\mathbf{Z},\mathbf{P},\mathbf{W},\mathbf{V}} \quad \sum_{j \in J} f_j X_j - \alpha \sum_{i \in I} \sum_{k \in K} \sum_{r=N}^{|\mathcal{J}|} v_i e_{ik} W_{ikr} \tag{6.4a}$$

$$\text{s.t.} \quad (6.1\text{b}) - (6.1\text{h}), (6.1\text{j}) - (6.1\text{k}), (6.3\text{a}) - (6.3\text{h}),$$

$$P_{ikr} = \sum_{j \in \mathcal{J}} p_j^{1_{[j \in J]}} V_{ikjr},\ \forall k \in K, r = 1, \cdots, |\mathcal{J}|, i \in I, \tag{6.4b}$$

$$W_{ikr} \ge 0, V_{ikjr} \ge 0,\ \forall k \in K, j \in \mathcal{J}, r = 1, \cdots, |\mathcal{J}|, i \in I. \tag{6.4c}$$

Owning to the formidable size of variables in the model, solving (LSLP) by commercial solvers is still not an easy job. CPLEX fails to obtain a feasible solution for a small size network even after several hours of computation. In the following section, more sophisticated solution approaches are developed to overcome such computational difficulties.

6.3 Solution Algorithm

6.3.1 Lagrangian Relaxation

In (LSLP), the sensor location variables $\mathbf{X}$ are correlated with the sensor level assignment variables $\mathbf{Z}$ by constraints (6.1b), which is further correlated with the sensor combination assignment variables $\mathbf{Y}$ through constraints (6.1g) and (6.1h). Such correlation complicates the model and makes the computation challenging. Moreover, a great amount of continuous variables are introduced for linearization, which adds to the computation burden significantly. In the following, we will work with the original (SLP) directly to tackle the problems through various relaxation and approximation techniques. To decouple the correlation between $\mathbf{X}$ and $\mathbf{Z}$, we relax constraints

(6.1b) in (SLP) and add them to objective function (6.1a) with nonnegative Lagrangian multipliers $\mu = \left\{\mu_{ij}, \forall i \in I, j \in J\right\}$. The relaxed problem becomes:

$$\text{(RSLP)} \quad \min_{\mathbf{X},\mathbf{Y},\mathbf{Z},\mathbf{P}} \quad \sum_{j\in J}(f_j - \sum_{i\in I}\mu_{ij})X_j - \alpha \sum_{i\in I}\sum_{k\in K}\sum_{r=N}^{|\mathcal{J}|} v_i e_{ik} P_{ikr} Y_{ikr} + \sum_{i\in I}\sum_{j\in J}\mu_{ij}\sum_{r=1}^{|\mathcal{J}|} Z_{ijr} \tag{6.5a}$$

$$\text{s.t.} \quad \text{(6.1c)} - \text{(6.1k)}.$$

Given μ, the optimal solution of (6.5a) provides a lower bound to the original (SLP) problem. After the above relaxation, the (RSLP) reduces to two parts, which can be solved separately. The part involving $\mathbf{X}$,

$$\min_{X_j\in\{0,1\}} \sum_{j\in J}\left(f_j - \sum_{i\in I}\mu_{ij}\right)X_j,$$

can be solved by simple inspection; i.e., given any $\{\mu_{ij}\}$, we can easily find the optimal $\mathbf{X}$ as follows:

$$X_j = \begin{cases} 1 & \text{if } f_j - \sum_{i\in I}\mu_{ij} < 0, \\ 0 & \text{otherwise.} \end{cases}$$

The part involving $\mathbf{Z}$ and $\mathbf{Y}$ can be further separated into individual sub-problems, one for each neighborhood i. For ease of notation, we omit the subscripts i in Z_{ijr}, Y_{ikr} and P_{ikr}. The sub-problem (RSLP_i) with respect to neighborhood i is:

$$(\text{RSLP}_i) \quad \min -\alpha \sum_{k\in K}\sum_{r=N}^{|\mathcal{J}|} v_i e_{ik} P_{kr} Y_{kr} + \sum_{j\in J}\mu_{ij}\sum_{r=1}^{|J|} Z_{jr} \tag{6.6a}$$

$$\text{s.t.} \quad \sum_{r=1}^{|\mathcal{J}|} Z_{jr} \le 1, \ \forall j \in J, \tag{6.6b}$$

$$\sum_{r=1}^{|\mathcal{J}|} Z_{jr} = 1, \ \forall j \in \tilde{J}, \tag{6.6c}$$

$$Z_{jr} = Z_{j+1,r+1}, \ \forall r = 1, \cdots, |\mathcal{J}| - 1, j \in \tilde{J}\backslash|J| + N, \tag{6.6d}$$

$$\sum_{j\in J} Z_{jr} + \sum_{s=1}^{r} Z_{|J|+1,s} = 1, \ \forall r = 1, \cdots, |\mathcal{J}|, \tag{6.6e}$$

$$Y_{kr} \leq \frac{1}{N} \sum_{j \in \mathcal{J}} \sum_{s=1}^{r} a_{kj} Z_{js}, \ \forall k \in K, r = 1, \cdots, |\mathcal{J}|, \tag{6.6f}$$

$$Y_{kr} \leq \sum_{j \in \mathcal{J}} a_{kj} Z_{jr}, \ \forall k \in K, r = 1, \cdots, |\mathcal{J}|, \tag{6.6g}$$

$$P_{kr} \leq \sum_{j \in \mathcal{J}} p_j Z_{jr} P_{kr-1}, \ \forall k \in K, r = 1, \cdots, |\mathcal{J}|, \tag{6.6h}$$

$$P_{k0} = \prod_{j \in J} (1 - p_j)^{a_{kj}} (p_j)^{-a_{kj}}, \ \forall k \in K, \tag{6.6i}$$

$$Z_{jr}, Y_{kr} \in \{0, 1\}, \ \forall k \in K, j \in \mathcal{J}, r = 1, \cdots, |\mathcal{J}|. \tag{6.6j}$$

(RSLP$_i$) can be linearized the same way as (SLP) by adding (6.3a)-(6.3h). It is well-known that the optimal objective value of the above (RSLP) for any given μ provides a lower bound to the original (SLP) problem. According to Proposition 15, a nearer sensor must be assigned to an earlier level at optimum. Based on this property, we can find an upper bound to the original (SLP) quickly through fixing the optimal sensor location decisions $\mathbf{X}$ obtained from the relaxed problem (RSLP) and assigning neighbourhoods accordingly. For each neighbourhood i, we sort all constructed sensors (i.e., $X_j = 1$) in ascending order of d_{ij} and assign each sensor with a level r equal to its rank in distance (i.e., $Z_{ijr} = 1$ if sensor j is installed to be the r^{th} nearest sensor to neighborhood i). Based on the level assignment of the installed sensors (the value of $\mathbf{Z}$), we enumerate all possible combinations $\mathbf{Y}$ to get their total accuracy contribution.

In the remainder of the Lagrangian relaxation solution framework, we use standard sub-gradient technique (Fisher, 2004) to update the multipliers ; i.e.,

$$\mu_{ij}^{n+1} = \mu_{ij}^{n} + s_j^n \left(\sum_{r=1}^{|\mathcal{J}|} Z_{ijr}^n \leq X_j^n \right), \tag{6.7}$$

$$s_j^n = \frac{\xi^n \left(g^* - g_D(\mu^n) \right)}{\left\| \sum_{r=1}^{|\mathcal{J}|} Z_{ijr}^n - X_j^n \right\|^2}, \tag{6.8}$$

where μ_{ij}^n represents a multiplier in the n^{th} iteration, s_j^n is the step size, ξ^n is a scalar and g^* and $g_D(\mu^n)$ are the best upper bound and the current lower bound, respectively. If the Lagrangian relaxation algorithm fails to find a solution with small enough gap in a certain number of iterations, we embed it into a branch-and-bound (B&B) framework to further close the gap.

However, solving the mixed integer program (RSLP_i) repeatedly for each neighborhood and across Lagrangian relaxation iterations could still be time-consuming. As such, an approximation approach is developed to quickly identify lower bounds to the relaxed sub-problems (RSLP_i).

6.3.2 Approximation of P_{kr}

Equations (6.6h) show that P_{kr} depends on P_{kr-1} and Z_{jr}, which builds connections across the decision variables and brings difficulties in solving the sub-problem. Similar to Cui et al. (2010), we approximate the variable probability P_{kr} with fixed numbers. For each combination k with its highest level element sensor assigned at level r, we select the regular sensors which are not in k and are closer to the monitored neighborhood than its most remote sensor in k. Let the number of qualified regular sensors be κ, where $\kappa < |J|$. We rank those κ regular sensors based on their failure probabilities and let $j_1, j_2, \cdots, j_\kappa$ be an ordering of the sensors such that $p_{j_1} \geq p_{j_2} \geq \cdots \geq p_{j_\kappa}$. For $N \leq r \leq N + \kappa$, we define one set of variables $\beta_{kr} = \prod_{j' \in J} (1 - p_{j'})^{a_{kj'}} \prod_{l=1}^{r-N} p_{j_l}$. While for $r < N$ or $r > N + \kappa$, we set $\beta_{kr} = 0$. Replacing P_{kr} with β_{kr}, we can modify the (RSLP_i) as:

$$(\text{DRSLP}_i) \quad \min -\alpha \sum_{k \in K} \sum_{r=N}^{|\mathcal{J}|} v_i e_{ik} \beta_{kr} Y_{kr} + \sum_{j \in J} \mu_{ij} \sum_{r=1}^{|J|} Z_{jr} \tag{6.9a}$$

$$\text{s.t.} \quad (6.6\text{b}) - (6.6\text{g}), (6.6\text{j}).$$

Proposition 16. *The solution to ($DRSLP_i$) provides a lower bound to the relaxed subproblem ($RSLP_i$).*

Proof. (DRSLP_i) is constructed through replacing P_{kr} with β_{kr} and removing constraints (6.6h)-(6.6i). As removing constraints enlarges the feasible region of (RSLP_i), it will never increase the objective value of this minimization problem. The effect of replacing P_{kr} with β_{kr} in the objective function is studied under two scenarios where $Y_{kr} = 1$ or $Y_{kr} = 0$. If $Y_{kr} = 0$, the value of β_{kr} won't affect the optimal objective value as $\beta_{kr} Y_{kr} = P_{kr} Y_{kr} = 0$. When $r < N$, $Y_{kr} = 0$ must hold since the most remote sensor in k must be assigned at a higher level than N. When $r > N + \kappa$, we also have $Y_{kr} = 0$ since a sensor can't be assigned to a level higher than the total number of sensors who are closer than it. When $Y_{kr} = 1$, the probability of using combination k is $P_{kr} = \prod_{j \in J} (1 - p_j)^{a_{kj}} \prod_{s \leq r} \left[\sum_{j \in \mathcal{J}} Z_{ijs} (p_j)^{1 - a_{kj}} \right]$ according to the proof to Proposition 1. P_{kr}

calculates the probability to have the N sensors in k working and $r-N$ regular sensors disrupted. Based on the construction of β_{kr} where $N \leq r \leq N+\kappa$, β_{kr} provides an upper bound to P_{kr} if $Y_{kr}=1$. Therefore, $\beta_{kr}Y_{kr}$ must be an upper bound to $P_{kr}Y_{kr}$ for any $k \in K, r = N, \cdots, |\mathcal{J}|$ and the optimal objective value of (DRSLP$_i$) is a lower bound to the optimal objective value of (RSLP$_i$). □

6.3.3 Approximation of Y_{kr}

In this section, constraints (6.6f) and (6.6g) are replaced by a simple equality formula to decouple the connection between Z_{jr} and Y_{kr}. For a neighborhood i, the Lagrangian multiplier μ_{ij} in (6.9a) can be interpreted as an extra installation cost of sensor j and the first term $\sum_{k\in K}\sum_{r=N}^{|\mathcal{J}|} v_i e_{ik}\beta_{kr}Y_{kr}$ represents the total accuracy contribution of the installed sensors to the system. Given that the N dummy sensors are always installed and assigned to the highest levels, we let the regular sensors be installed sequentially from level 1 to level $|J|$. Let binary variables $\{y_{krt} : \forall k, r\}$ be the combination assignments when t regular sensors are installed. $y_{krt}=1$ if combination k is used ($Y_{kr}=1$) given t regular sensors are installed. As such, the total accuracy can be decomposed into $|J|$ portions, one for each level t. The tth portion calculates the additional benefits contributed by installing a sensor j at level t.

The accuracy contribution of all sensors, i.e. the first term in (6.9a) omitting the constant v_i, can be reformulated as:

$$AC = \sum_{k\in K}\sum_{r=N}^{|\mathcal{J}|} e_{ik}\beta_{kr}Y_{kr} = \sum_{r=N}^{N}\sum_{k\in K} e_{ik}\beta_{kr}y_{kr0} + \sum_{t=1}^{|J|}\left[\sum_{r=N}^{N+t}\sum_{k\in K} e_{ik}\beta_{kr}y_{krt} - \sum_{r=N}^{N+t-1}\sum_{k\in K} e_{ik}\beta_{kr}y_{kr,t-1}\right], \tag{6.10}$$

where $\sum_{r=N}^{N}\sum_{k\in K} e_{ik}\beta_{kr}y_{kr0}$ represents the accuracy contribution of the N dummy sensors, which is 0 by definition; $\sum_{r=N}^{N+t}\sum_{k\in K} e_{ik}\beta_{kr}y_{krt}$ states the accuracy level of the system with N dummy sensors and t regular sensors; difference of the two terms in the parentheses represents the accuracy improvement by adding one regular sensor at level t given that $t-1$ regular sensors are already installed. If we expand the summation terms in (6.10), the intermediate accuracy level $\sum_{r=N}^{N+t}\sum_{k\in K} e_{ik}\beta_{kr}y_{krt}$ for any t where $t < |J|$ will be cancelled out. As such, AC will be simplified as $AC = \sum_{r=N}^{N+|J|}\sum_{k\in K} e_{ik}\beta_{kr}y_{krt} = \sum_{k\in K}\sum_{r=N}^{|\mathcal{J}|} e_{ik}\beta_{kr}Y_{kr}$, which mathematically proves the second equivalence in (6.10).

Adding one regular sensor j at level t is equivalent to replacing the dummy sensor $|J|+1$ with j and moving all the dummy sensors upward by one level. The resultant accuracy difference for the two systems with t or $t-1$ regular sensors can be calculated by updating the accuracy level of every combination relating to sensor j. The total system accuracy is reformulated as:

$$AC = \sum_{t=1}^{|J|} \sum_{r=\max\{t,N\}}^{N+t} \sum_{k\in K} \left(e_{ik}\beta_{kr} - (1-p_j) e_{ik'}\beta_{k'r} \right) 1_{[J_k \setminus J_{k'}=j \&\& b_{kj}=1]} y_{krt}, \tag{6.11}$$

where J_k represents the set of regular sensors in combination k and parameter $b_{kj}=1$ if j is the most remote regular sensor in k and is 0 otherwise. In the indicator function $1_{[\cdot]}$, $J_k \setminus J_{k'} = j$ identifies the updated combination k who has one additional regular sensor j comparing with an existing combination k'; $b_{kj}=1$ forces this sensor j to be the most remote regular sensor in k. Inserting j at level t brings new combinations – the term $e_{ik}\beta_{kr}1_{[J_k \setminus J_{k'}=j \&\& b_{kj}=1]}$ calculates the accuracy contribution of a new combination k which uses j as its most remote regular sensor. Moreover, inserting j at level t changes the assignment level of all dummy sensors. An existing combination k' who contains dummy sensors in the old system (with $t-1$ regular sensors) will be used in the new system (with t regular sensors) only when sensor j is disrupted – the term $(1-p_j)e_{ik'}\beta_{k'r}1_{[J_k \setminus J_{k'}=j \&\& b_{kj}=1]}$ in (6.11) calculates the contribution deduction due to the probability decrease of using combination k'.

Fig. 6.2 illustrates the decomposition process in order to calculate the total accuracy level of the system with 3 dummy sensors and $|J|$ regular sensors. Sensors $a, b, c, \cdots, j$ are sequentially added to the system to calculate their contribution. For example, contribution of sensor c is equal to the difference in system accuracy when $t=3$ or $t=2$. Let the element sensors-combination index be defined as: abc-1, abD_1-2, acD_1-3, bcD_1-4, aD_1D_2-5, bD_1D_2-6, cD_1D_2-7, $D_1D_2D_3$-8. The system accuracy when $t=3$ and $t=2$ respectively are

$$AC|_{t=3} = \sum_{r=N}^{N+3} \sum_{k\in K} e_{ik}\beta_{kr}y_{kr3} = e_1\beta_{13} + e_2\beta_{24} + e_3\beta_{34} + e_4\beta_{44} + e_5\beta_{55} + e_6\beta_{65} + e_7\beta_{75} + e_8\beta_{86},$$

$$AC|_{t=2} = \sum_{r=N}^{N+2} \sum_{k\in K} e_{ik}\beta_{kr}y_{kr2} = e_2\beta_{23} + e_5\beta_{54} + e_6\beta_{64} + e_8\beta_{85}.$$

The contribution of inserting sensor c at level $t = 3$ is

$$AC|_{t=3} - AC|_{t=2} = [e_1\beta_{13} - e_2(\beta_{23} - \beta_{24})] + [e_3\beta_{34} - e_5(\beta_{54} - \beta_{55})] + [e_4\beta_{44} - e_6(\beta_{64} - \beta_{65})] + [e_7\beta_{75} - e_8(\beta_{85} - \beta_{86})], \quad (6.12)$$

where each combination k, $k = 1, 3, 4$ or 7, has element sensor c as its most remote regular sensor, namely $b_{kc} = 1$; the combinations paired up in brackets (1 and 2; 3 and 5; 4 and 6; 7 and 8) have the same regular sensors except for sensor c, namely $J_k \backslash J_{k'} = c$. According to the construction of β_{kr}, the paired probabilities in the parentheses satisfy $\beta_{24} = p_c\beta_{23}, \beta_{55} = p_c\beta_{54}, \beta_{65} = p_c\beta_{64}$ and $\beta_{86} = p_c\beta_{85}$. Substituting $\beta_{k'r+1}$ by $p_c\beta_{k'r}$ in the parentheses, we can simplify (6.12) as follows:

$$AC|_{t=3} - AC|_{t=2} = [e_1\beta_{13} - e_2\beta_{23}(1 - p_c)] + [e_3\beta_{34} - e_5\beta_{54}(1 - p_c)] + [e_4\beta_{44} - e_6\beta_{64}(1 - p_c)] + [e_7\beta_{75} - e_8\beta_{85}(1 - p_c)]. \quad (6.13)$$

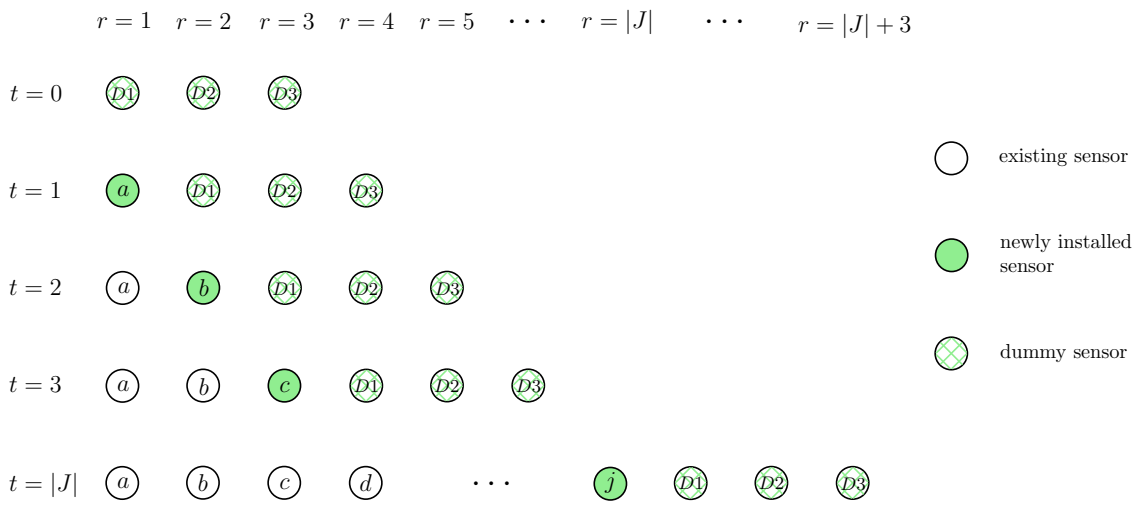


Figure 6.2: Sensor contribution decomposition.

For any t and r satisfying $1 \le t \le |J|$, $\max\{t, N\} \le r \le N + t$, a combination k fulfilling $1_{[J_k \backslash J_{k'} = j \&\& b_{kj} = 1]} y_{krt} = 1$ must have j as its most remote regular sensor, have its most remote sensor assigned at level r and thus have $r - t$ dummy sensors. Hence we only need to choose $N - 1 - (r - t)$ regular sensors from the $t - 1$ alternatives to get a qualified k. We denote the maximum number of such updated combinations by $n_{tr} = \binom{t-1}{N-1-r+t}$. We also let $C_{ijrk} =$

$\left(e_{ik}\beta_{kr} - (1-p_j)e_{ik'}\beta_{k'r}\right)1_{[J_k \setminus J_{k'}=j \&\& b_{kj}=1]}$. For each $1 \le t \le |J|$, $\max\{t,N\} \le r \le N+t, 1 \le j \le |J|$, let $k_1, k_2, \cdots, k_{|K|}$ be an ordering of the coefficients C_{ijrk} such that $C_{ijrk_1} \ge C_{ijrk_2} \ge \cdots \ge C_{ijrk_{|K|}}$. We define $\gamma_{ijt} = \sum_{r=\max\{t,N\}}^{N+t} \sum_{l=1}^{n_{tr}} C_{ijrk_1}$. Based on the construction of γ_{ijt}, $\gamma_{ijt}Z_{jt}$ provides an upper bound to the accuracy improvement from inserting sensor j at level t. Replacing AC by its upper bound $\sum_{t=1}^{|J|}\sum_{j\in J}\gamma_{ijt}Z_{jt}$, (DRSLP$_i$) further reduces to the following simple assignment problem (TRSLP$_i$), which can be solved by the Hungarian algorithm.

$$\text{(TRSLP}_i\text{)} \quad \min -\sum_{t=1}^{|J|}\sum_{j\in\mathcal{J}} v_i\gamma_{ijt}Z_{jt} + \sum_{r=1}^{|J|}\sum_{j\in\mathcal{J}}\mu_{ij}Z_{jr} \tag{6.14a}$$

$$\text{s.t.} \quad \text{(6.6b)} - \text{(6.6e)},$$

$$Z_{jr} \in \{0,1\},\ \forall j \in \mathcal{J}, r = 1, \cdots, |\mathcal{J}|. \tag{6.14b}$$

Proposition 17. *The solution to (TRSLP$_i$) provides a lower bound to the relaxed sub-problem (RSLP$_i$).*

Proof. (DRSLP$_i$) is constructed through replacing $\sum_{k\in K}\sum_{r=N}^{|\mathcal{J}|} v_i e_{ik}\beta_{kr}Y_{kr}$ with the approximation $\sum_{r=1}^{|J|}\sum_{j\in\mathcal{J}} v_i\gamma_{ijr}Z_{jr}$ and removing constraints (6.6f)-(6.6g). As removing constraints enlarges the feasible region of (DRSLP$_i$), it will never increase the objective value of this minimization problem. Based on the construction of γ_{ijr}, $\sum_{r=1}^{|J|}\sum_{j\in\mathcal{J}} v_i\gamma_{ijr}Z_{jr}$ provides an upper bound to $\sum_{k\in K}\sum_{r=N}^{|\mathcal{J}|} v_i e_{ik}\beta_{kr}Y_{kr}$. Therefore, the optimal objective value of (TRSLP$_i$) is a lower bound to the optimal objective value of (DRSLP$_i$). Together with the result in Proposition 16, the solution of (TRSLP$_i$) is a lower bound to the relaxed sub-problem (RSLP$_i$). □

6.4 Numerical Examples

To demonstrate the applicability of the proposed models and algorithms, we apply them to a series of hypothetical grid networks as well as a more realistic Wi-Fi Access Point (AP) network in Terminal 5 of the Chicago O'Hare Airport. The proposed solution algorithms are programmed in C++ and run on a 64-bit Intel i7-3770 computer with 3.40 GHz CPU and 8G RAM. The linearized LSLP are tackled by commercial solver CPLEX 12.4 using up to 4 threads. We set the overall solution time limit to be 3600 seconds.

6.4.1 Hypothetical Grid Networks

A 2×3 rectangle grid network and six $n \times n$ square grid networks for $n \in \{3, 4, 5, 6, 7, 8\}$ are generated to represent various hypothetical study regions. In the square grid networks, each network contains $(n-1)^2$ cells. The four corners of each cell represent the candidate sensor locations, adding to a total number of n^2 candidate sensor locations. The centroid of each cell is constructed to be a surveillance neighborhood, adding to $(n-1)^2$ neighborhoods. The network layouts are shown in Fig. 6.3. We omit the surveillance neighborhoods in some of the larger networks (i.e., from 5×5 to 8×8) for cleaner figure presentation. The edge length of each cell is set to 1. The customer demand of each neighborhood i is $v_i = 10$, the value of α is 1, and the fixed sensor installment cost is 10. The value of coverage is 1. The site-dependent failure probability of sensor location j is assumed to vary from 0.1 to 0.2 based on its Euclidean distance to the center of the study region. The sensor(s) located nearest to the center have the highest failure probability of 0.2, the sensor(s) located farthest away have the lowest probability of 0.1. The failure probability of a sensor in the middle linearly decreases with the distance to the center. Each combination uses $N = 3$ sensors. Combination accuracy is computed based on $e_{ik} = \sum_{j \in J} \frac{a_{kj}}{(d_{ij})^2 + \epsilon}, \forall i \in I, k \in K$, where d_{ij} is the Euclidean distance and ϵ is a small positive number. The reliable sensor location problems are solved by two approaches: (i) CPLEX directly applied to the mixed-integer linear program LSLP and (ii) Lagrangian relaxation based branch-and-bound method with approximation algorithm (LR+B&B+Approx.). Table 6.1 summarizes and compares the results from the two approaches.

As one can observe from the table, the solution time and solution quality rapidly deteriorate with the network size, due to the significant increase in the number of integer variables $\mathbf{Y}$ and $\mathbf{Z}$. CPLEX could only find the optimal solution to the specifically constructed small rectangle network. In the second case, CPLEX identified a feasible solution but failed to find a lower bound, despite its rather small network size. For the other larger networks, CPLEX ran out of memory and could not provide a feasible solution or a lower bound. In contrast, optimal solutions to the first 6 cases were obtained by the LR+B&B+Approx. approach within 3 minutes. For the 8×8 network, there is a residue gap of 2.32% after 1 hour of computation.

In Fig. 6.3, the installed sensors in the best solutions from the LR+B&B+Approx. approach are marked green. We can observe that more sensors are installed in order to monitor a larger

Table 6.1: Algorithm performance comparison for the 7 hypothetical cases.

	Sensor network	Neighborhood network	No. of sensors	Final UB	Final LB	Final gap (%)	CPU time (s)
CPLEX	2×3	1×2	2	-1.31	-1.30	0	1.6
	3×3	2×2	4	-14.01	fail	100	3600
	4×4	3×3	-	-	-	fail	3600
	...	...	-	-	-	fail	3600
	8×8	7×7	-	-	-	fail	3600
LR+B&B	2×3	1×2	2	-1.31	-1.31	0	0.1
	3×3	2×2	4	-24.28	-24.28	0	0.1
	4×4	3×3	5	-77.38	-77.38	0	0.4
	5×5	4×4	8	-150.78	-150.78	0	0.8
	6×6	5×5	14	-243.48	-243.48	0	38
	7×7	6×6	21	-360.99	-360.99	0	181
	8×8	7×7	29	-489.07	-500.41	2.32	3600

region. In the first three cases, the installed sensors are clustered in the center of the study region mainly owning to their short distances to all the surveillance neighborhoods, which provides better accuracy with a limited number of sensors. In the four larger cases, it is interesting to observe that the sensors are installed symmetrically along the diagonal lines. Moreover, no sensor is installed immediately next to the boundary, while all nearby candidate locations (e.g., slightly closer to the region center) are selected. Those properties indicate the possibility to decompose a larger yet symmetrical network into several smaller ones to obtain the sensor deployment effectively. Take the 8×8 network for example, if the sensor at coordinate (1, 1) is installed (assuming the bottom left sensor is located at the origin (0, 0)), then we can automatically install the sensor at (6, 6), which could significantly speed up solution process. As such, the proposed algorithm could possibly handle an even larger symmetrical network efficiently.

Fig. 6.4 illustrates how the sensor combinations are used by the customers in neighborhood $i = 1$ (i.e., indicated by the dark star in Fig. 6.3) in the 3-by-3 case. The installed sensors 4, 5, 6, 8 are assigned to levels 1 - 4 based on distance, while the dummy sensors are assigned at levels 5 - 7. Some representative combinations are illustrated in this figure. For example, the shaded combination ($k = u$) will be used to monitor this neighborhood if and only if sensors 5, 6 and 8 are functioning and sensor 4 has been disrupted. The most remote sensor in this combination is 8, which is ranked at level $r = 4$. Hence combination u corresponds to backup level $r = 4$ and it will be used with a probability of $P_{1u4} = p_4(1 - p_5)(1 - p_6)(1 - p_8) = 0.15 \times 0.8 \times 0.75 \times 0.75 = 0.0675$ based on the sensor failure probability settings.

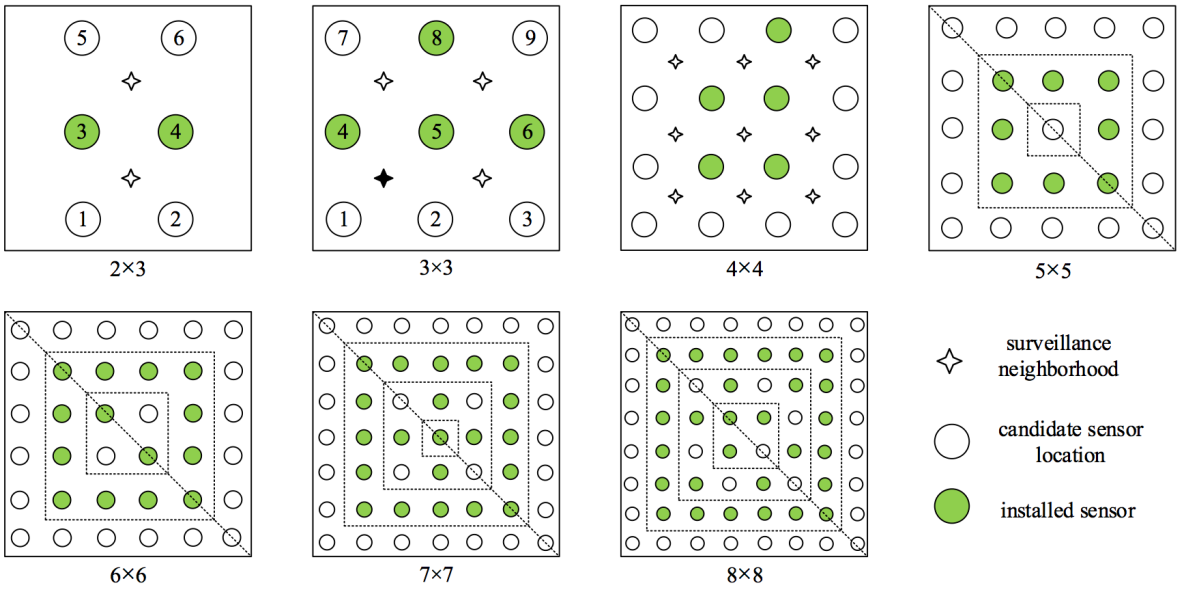


Figure 6.3: Optimal sensor deployment for the 7 hypothetical cases.

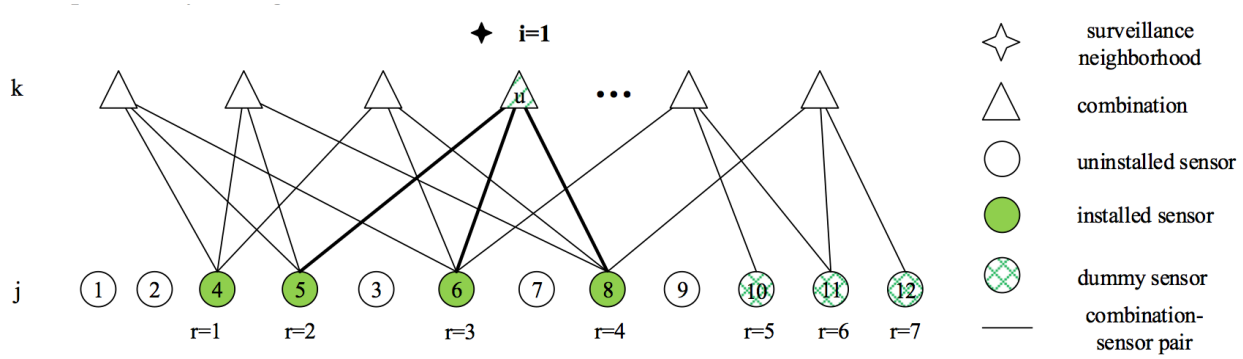


Figure 6.4: Detailed assignment plan of sensor combinations to neighborhood $i = 1$.

6.4.2 Wi-Fi Point Network for O'Hare Terminal 5

The Chicago O'Hare International Airport is one of the busiest airports in the world. In June 2016 alone, a total of 7,329,084 travelers passed through the airport (CDA, 2016). Boingo, the O'Hare Airport's Wi-Fi provider, has pioneered a new "S.M.A.R.T" network design (Secure, Multi-platform, Analytics-Driver, Responsive and Tiered) which allows increased access point density for location-based services like queue management, advertising, and passenger guidance. Such a system is expected to deliver valuable business intelligence and actionable insights to enable high-quality passenger service.

In this case study, we select the departure level of Terminal 5 to investigate Wi-Fi Access

Point deployment for better location-based services. Terminal 5 contains Concourse M, which is used for all international arrivals and part of the international departures (those of most non-US carriers). We select 52 heavy-traffic venues inside the terminal, including 21 gates, 10 restaurants, 13 shops, 6 airline lounges and 1 security check point, as key surveillance neighborhoods; see Fig. 6.5. Average hourly surveillance demand at each neighborhood is assumed to be proportional to the local passenger flow per the monthly statistics report of the Chicago Department of Aviation (CDA, 2016). The terminal is further divided into square cells with an edge length of 10 meters[b]. The corners of every square cell are considered candidate sensor/AP locations. There are 222 candidate locations in total. Boingo uses Cisco's AP systems with chipsets featuring 802.11ac standard, with an installation cost of about US$200 each (maintenance cost or other capital cost is not considered). The received signal strength (RSS) follows a logarithm function of distance (Shchekotov, 2014), and hence we assume a combination of sensors will yield an accuracy measure of $e_{ik} = \sum_{j \in J} 22 a_{kj} \log_{10}\left(\frac{40}{d_{ij}+\epsilon}\right), \forall i \in I, k \in K$, where d_{ij} is the Euclidean distance in meters, and 40 (meters) is the effective range of a Cisco AP. Each combination uses $N = 3$ sensors.

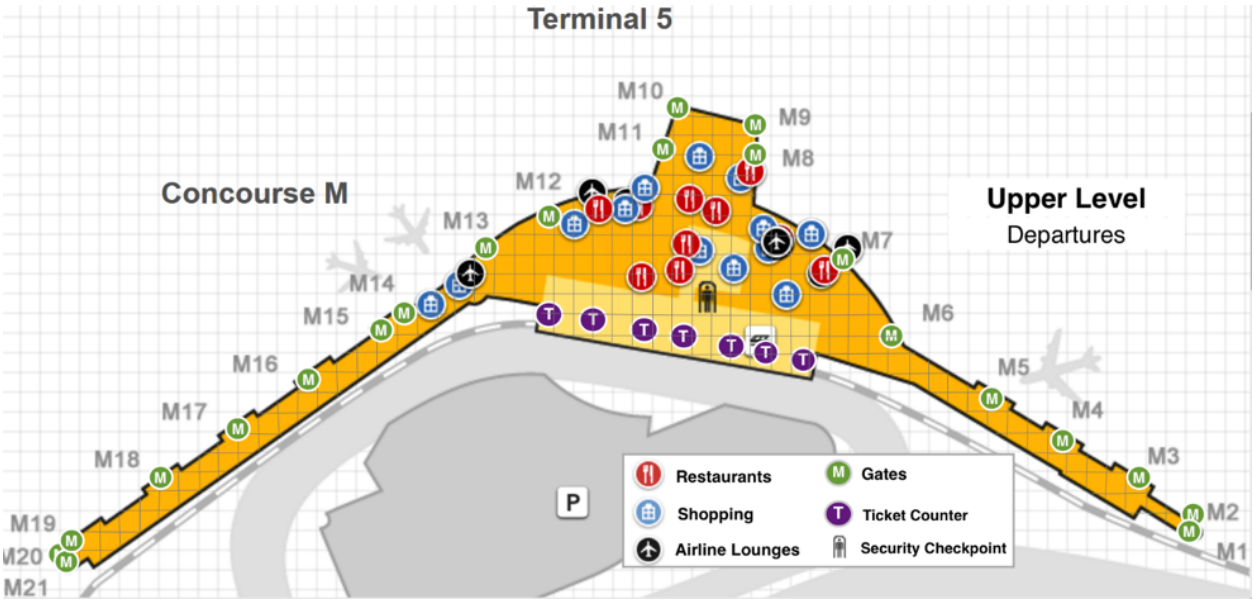


Figure 6.5: O'Hare terminal 5 map (Source: http://www.flychicago.com/OHare/EN/AtAirport/map).

We consider site-independent, yet low, median and high levels of sensor disruption probabilities;

[b]Generally, access points should be separated by at least 10 feet in order to reduce adjacent channel interference, and it is recommended that APs are mounted at 30-40 feet (or approximately 10 meters) from one another (https://supportforums.adtran.com/docs/DOC-7257).

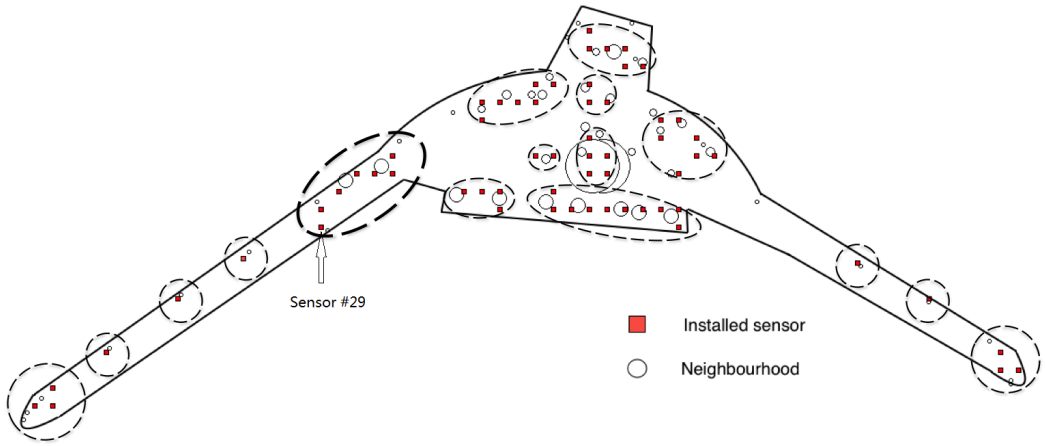
i.e., $p_j = p \in \{0.01, 0.2, 0.5\}, \forall j$. The system performance measures under these scenarios are presented in Table 6.2. All results are obtained from the proposed LR+B&B+Approx algorithm within 3600 seconds. Overall, a higher sensor failure probability leads to a fewer number of installed sensors as well as a significant deterioration in the best objective value (i.e., the final UB). The residue gap also increases slightly with the failure probability. The value of α reflects the tradeoff between the positioning accuracy e_{ik} and the unit sensor installation cost f_j. Very often the value of α may be subject to speculation and interpretation. We thus conduct sensitivity analysis over α while keeping the same formula for e_{ik}, and when $f_j = 200, p_j = 0.01$. When α increases from 0.025 to 0.4, the number of installed sensors increases drastically from 17 to 111, and the objective function drops by about two orders of magnitude. These results indicate that the benefits of deploying more sensors far outweigh the installation costs in the O'hare case study.

Table 6.2: Performance measures for the O'Hare Airport case.

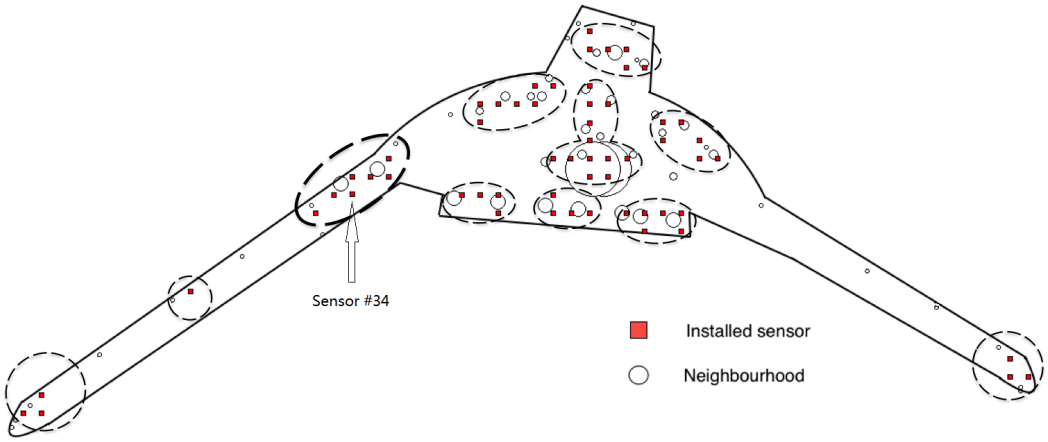
Failure prob	α	No. of candidate sensors	No. of neighbor -hoods	No. of installed sensors	Final UB	Final LB	Final gap(%)	CPU time(s)
0.01	0.1	222	52	61	-32306.5	-33795.4	4.6	3600
0.2	0.1	222	52	58	-27720.0	-29147.4	5.1	3600
0.5	0.1	222	52	56	-18761.6	-19786.5	5.5	3600
0.01	0.025	222	52	17	-3417.3	-3739.3	8.5	3600
0.01	0.050	222	52	34	-11732.2	-12355.9	5.0	3600
0.01	0.2	222	52	92	-79741.0	-82682.9	3.6	3600
0.01	0.4	222	52	111	-180132.0	-186298.0	3.3	3600

The optimal sensor locations for the three cases are shown in Fig. 6.6. The solid-line circles represent the surveillance neighborhoods, and their size indicates the volume of surveillance demand. The installed sensors are marked by shaded squares. They can be roughly clustered into groups, as shown by the dotted ellipses, in which the distances between any adjacent sensors do not exceed 15 meters – i.e., these sensors are likely to provide effective backups to each other. Under a low failure probability, the installed sensors can be clustered into 16 groups, and the sensor nearest to every surveillance neighborhood is always installed. This forms a rather dispersed sensor network overall. In the two wings of the airport, 5 isolated sensors are installed in order to monitor their most adjacent neighborhoods, although these sensors only make marginal contributions to other neighborhoods.

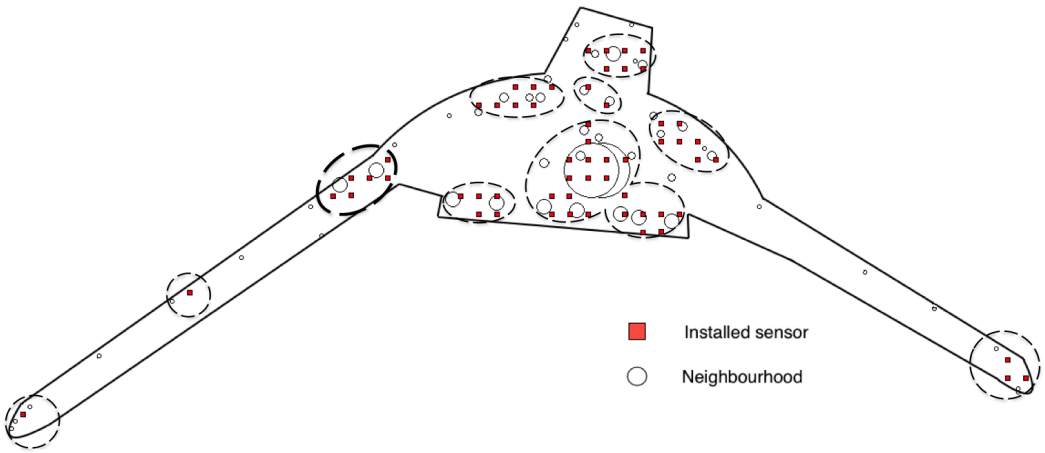
When the sensor disruption probability increases to 0.2 and 0.5, the number of sensor groups



(a) $p = 0.01$



(b) $p = 0.2$



(c) $p = 0.5$

Figure 6.6: Optimal sensor locations under (a) low, (b) median, and (c) high sensor disruption probabilities.

drops to 12 and 10, respectively, and fewer isolated sensors are installed. Sensors within a group tend to become more clustered so as to better back each other up. This is clearly illustrated, for example, by the highlighted group (see the bold ellipse). Meanwhile, sensors also tend to cluster around the center of the concourse where demand is the heaviest. For example, 10 sensors are clustered within 20 meters from at the security checkpoint when $p = 0.5$, while there are only 7 when $p = 0.2$ and 5 when $p = 0.01$. In summary, under higher failure probability, the model tends to yield a higher degree of sensor clustering especially around the heavy-demand neighborhoods, while at the same time a smaller total number of sensors would be installed especially around the less crowded neighborhoods.

A closer look at the sensor deployment in the highlighted group (bold ellipse) reveals some interesting points. When p increases from 0.01 to 0.2, sensor #29 is removed from the low demand neighborhood while sensor #34 is added to the high demand neighborhood. Such changes can be explained by the marginal costs and marginal benefits of these sensors. In the case of $p = 0.2$, if we add sensor #29 back, the marginal coverage benefit is \$185.7, which is lower than its installation cost \$200. On the other hand, if we remove sensor #34, the coverage accuracy loss is \$252.1 when $p = 0.2$, which is higher than \$200. This result can be generalized. When the disruption probability increases, the sensors become less reliable, and more sensors will be needed to maintain the same coverage accuracy. In high-demand neighborhoods, the net marginal benefit of installing an extra "back-up" sensor (e.g., to maintain the accuracy) may be high enough to outweigh the installation cost. We hence may observe an increase in the sensor number near those neighborhoods. In low-demand neighborhoods, however, the net marginal benefit of adding a sensor may not justify its cost, and we will therefore expect reduction of sensors. In other words, the spatial distribution of sensors tends to be more clustered near high-demand neighborhoods under high disruption probabilities, but at the same time more sparse near low-demand neighborhoods. The total number of sensors across all neighborhoods may not exhibit a monotonic relationship with the value of p.

CHAPTER 7:

RELIABLE NETWORK DISTRICTING WITH CONTIGUITY, BALANCE, AND COMPACTNESS CONSIDERATIONS

Another important extension of the relible facility location research is reliable network districting problems, which aim at partitioning a network into districts under some operational considerations, and assign the demands in the partitioned districts to suppliers/facilities for actual service. In the reliable network districting problems, various operational criteria for districting (e.g., contiguity, balance, compactness) and the reliability of service providers (caused by internal or external factors) are simultaneously considered.[a]

In this chapter, we formulate the reliable network districting problem as mixed-integer optimization models using the location-assignment based modeling approaches. A series of modeling techniques are adopted to address multiple districting criteria: (i) *contiguity*: each district must be contiguous; (ii) *balance*: total nodal weights are balanced across districts; (iii) *compactness*: each district is compact in shape;. The reliability of facilities is incorporated by introducing district demand re-assignments. Customized solution approaches including constructive and neighborhood search heuristics, set-cover based lower bound estimation are designed to efficiently solve the mathematical models.

We apply the proposed methodology to an empirical railroad call center design problem. Railroad companies rely on good call centers to reliably handle incoming crew/resource call demands so as to maintain efficient operations and customer services in their networks. The full-scale case

[a]This chapter is based on a published paper, Xie and Ouyang (2016).

study demonstrates the performance and applicability of our methodology, and helps draw various managerial insights.

7.1 Introduction

Districting is a well-known problem in the operations research literature. It aims at partitioning a geographical space into sub-districts under various criteria and constraints. Depending on the specific application context, operational criteria may include the district contiguity, district compactness, workload balance, socio-economic homogeneity, etc. In the literature, probably the most intensively studied problem is regarding political districting, which divides a jurisdiction area (e.g., a state or a region) into electoral constituencies such that the political candidates from each area are elected to a parliamentary assembly. The "one man-one vote" principle requires that all districts contain approximately the same number of candidates/voters to avoid benefiting a certain party or candidate. There are several other applications of districting problems, which include: (i) service districting, referring to the design of districts for social facilities like schools, hospitals, fire stations; (ii) sales market districting, which subdivides the market areas of companies into multiple regions of responsibility; and (iii) distribution districting, which designs the pickup and delivery districts in logisitics context.

All the traditional districting problems simply focus on partitioning a network/area into districts under some operational considerations (Hess et al., 1965; Garfinkel and Nemhauser, 1970; Blais et al., 2003; Ferland and Guénette, 1990; Hess and Samuels, 1971). However, most of these studies ignore the fact that in many real-world districting applications, each of the partitioned district is associated with a supplier/facility. For example, in the school districting problem, all students in a district should be assigned to the corresponding school in that particular district. In the postal service districting problem, pickup/delivery locations in one district are typically visited sequentially by some specific postman/vehicle. Hence, many of the districting applications also involve facility location and demand assignment decisions. To the best of our knowledge, only very limited work has been done to combine the facility location and districting considerations. Moreover, the fact that facilities are subject to possible disruptions due to technical and personnel reasons emphasizes the necessity of further incorporating the reliability issues into the facility districting framework. Therefore, we aim at developing an innovative reliable facility districting

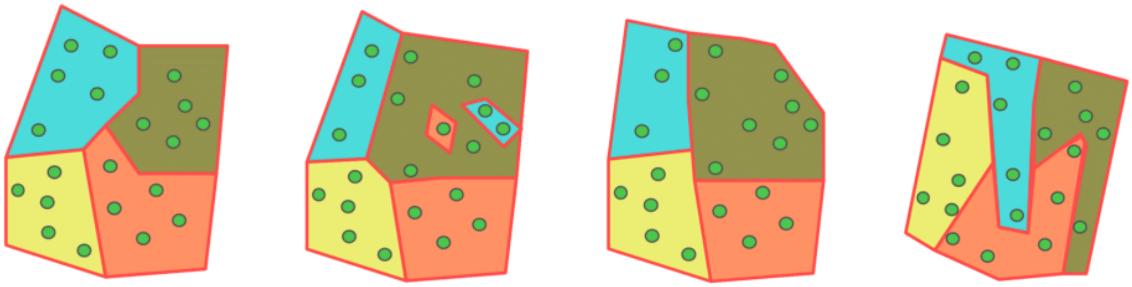
framework to incorporate the districting considerations into the reliable facility location models. It is worth noting that with reliability issues considered, many of the traditional modeling methods for addressing various operational districting criteria (e.g., district contiguity, workload balancing) would no longer work, or at least require modifications. Hence, we need to design new/adpated customized methods/techniques to address these various criteria.

In light of these challenges, in this chapter, we formulate the reliable network districting problem as mixed-integer optimization models using the location-assignment based modeling approaches. A series of modeling techniques (e.g., network flow constraints) are adopted to address multiple districting criteria: (i) contiguity: each district must be contiguous; (ii) compactness: each district is compact in shape; (iii) balance: demand is balanced across districts. The reliability of facilities is incorporated by introducing district re-assignments. Note that when facilities are subject to disruptions, the expected demand assigned to each facility across all possible facility failure scenarios should be considered. Customized solution approaches including constructive and neighborhood search heuristics, set-cover based lower bound estimation are developed to efficiently solve the mathematical models. Several numerical examples including a series of hypothetical test cases and an empirical full-scale railroad application are conducted to demonstrate the performance and applicability of our methodology. Various managerial insights are also drawn.

The remainder of this chapter is organized as follows. Section 7.2 introduces various operational districting criteria. Section 7.3 formulates the reliable districting problem into a mixed-integer optimization model using location-assignment based modeling approach. Section 7.4 designs customized heuristics to efficiently solve the model, and Section 7.5 presents a set-cover based estimation for the lower bound to provide optimality information. In Section 7.6, results for a series of hypothetical numerical examples and a full-scale railroad applications are presented.

7.2 Districting Criteria

As that defined in Mehrotra et al. (1998), a districting plan is a partitioning of a given graph/network (consisting of nodes and links) into a predetermined number of partitions such that the nodes in each parition induce a subgraph, which we call a district. Typically, districting problems involve the consideration of various operational criteria and constraints, including district contiguity, weight balance, and district compactness.



(a) Plan I (b) Plan II (c) Plan III (d) Plan IV

Figure 7.1: Graphical illustration for various districting criteria.

7.2.1 Contiguity

A district is contiguous if it is possible to travel between any two points in the same district without traversing any other district. In graphical terms, the subgraph corresponding to the district must be connected, i.e., there exists a path between any two nodes in the subgraph. Many districting applications require that each district in the plan is contiguous. For example, in the railroad call center design problem, the spatial district served by each caller desk should be contiguous so as to satisfy a number of practical operational requirements, e.g., administrative autonomy for resource/crew reallocation and train traffic management.

Figure 7.1(a) shows a districting plan (Plan I) in which all districts are contiguous, while in the Plan II represented by Figure 7.1(b), the districts colored with red and blue are apparently not contiguous as traveling from node 1 to node 2 (or from node 3 to node 4) requires going through the district in brown color.

7.2.2 Balance

In the given graph/network, each node is typically associated with a certain amount of weight, which can be either demand, workload, or population of the corresponding unit the node represents. Many districting plans are optimized to balance the total nodal weights across all districts. For example, in the political districting applications, the districts should have nearly equal populations to adhere to the one-person, one-vote principle. In the railroad call center design problems, the expected workload should be well balanced across caller desks, so that no desks are too much more occupied than others and bear excessive work pressure.

Comparing the districting plans shown in Figure 7.1(a) and Figure 7.1(c), we can easily observe that Plan I is more balanced than Plan III with respect to the number of nodes falling inside them. Specifically, in Plan I, all four districts have the same number of nodes, while in Plan III, the number of nodes in the four districts vary (i.e., 3,4,5,8).

7.2.3 Compactness

The compactness of districts is posed to prevent the formation of oddly-shaped districts. Intuitively, a compact district is supposed to have circular or square shape rather than being long and thin or snakelike. In districting problems, various forms of compactness standards have been imposed to achieve certain geographic configurations. For example, in political districting, measure of compactness is essentially enforced so as to disallow plans that were deemed to be gerrymandered. While in the railroad call center design applications, each spatial district is expected to be compact in shape so as to avoid high crew operating and transportation/logistics costs inside odd-shaped districts.

Figure 7.1(d) shows a plan IV with districts of elongated or curved shapes. To visit all nodes within each district, the traveling routes in Plan IV could be very awkward and may induce high transportation costs.

7.3 Location-Assignment Model Formulation

In this section, we formulate the reliable districting problem into a mixed-integer linear optimization program using the location-assignment based modeling approach. A series of modeling techniques (e.g., network flow constraints) are adopted to address the required practical criteria and to incorporate the reliability consideration. Customized algorithms consisting of a constructive heuristic and a neighborhood search procedure are developed to efficiently solve the mathematical model.

The districting problem is aiming at partitioning a given undirected network $\mathcal{G} = (\mathcal{I}, \mathcal{E})$ (with node set $\mathcal{I}$ and edge set $\mathcal{E}$) into a fixed number M of districts that jointly cover all the nodes in $\mathcal{I}$, and to assign these districts to M facilities for service. Binary parameter $\delta_{i_1 i_2}$ indicates whether two nodes $i_1 \in \mathcal{I}$ and $i_2 \in \mathcal{I}$ are adjacent in network $\mathcal{G}$; i.e., $\delta_{i_1 i_2} = 1$ if edge $(i_1, i_2) \in \mathcal{E}$, or $\delta_{i_1 i_2} = 0$ otherwise. Each node in the graph, $i \in \mathcal{I}$, generates a certain quantity of demand D_i (e.g., crew dispatch and emergency calls in the call center design problem). Let $\mathcal{M} = \{1, 2, \cdots, M\}$ be the set

of district indices. The nodes in each district, $I_m, m \in \mathcal{M}$, must be assigned to one facility so that the facility is primarily serving the demand from that district.

With the location-assignment based modeling approach, each node is included in one district, and then assigned to one facility. The corresponding formulation is presented as follows:

$$\text{(Location-Assignment)} \quad \min \quad \sum_{m \in \mathcal{M}} f(y_{im}) \tag{7.1a}$$

$$\text{s.t.} \quad \sum_{m \in \mathcal{M}} y_{im} = 1, \ \forall i \in \mathcal{I}, \tag{7.1b}$$

$$\text{constraints for contiguity,} \tag{7.1c}$$

$$\text{constraints for balance,} \tag{7.1d}$$

$$\text{constraints for compactness,} \tag{7.1e}$$

$$\text{constraints for reliable assignment,} \tag{7.1f}$$

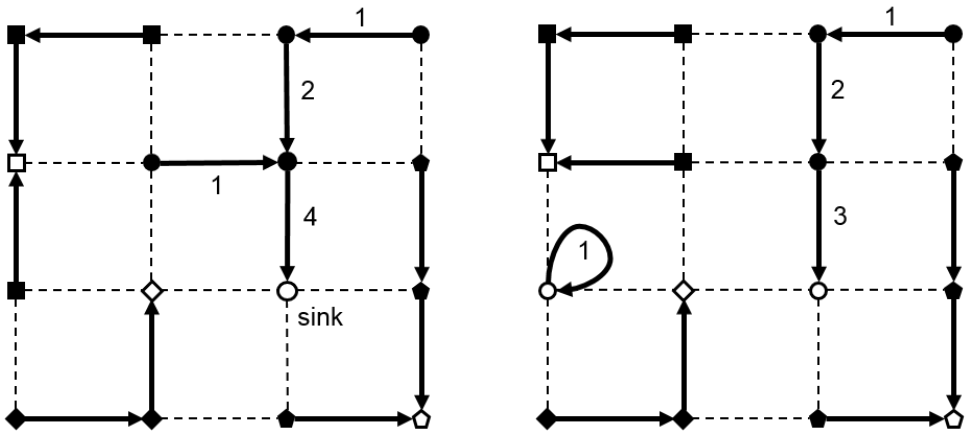
where binary variables y_{im} represent whether node $i \in \mathcal{I}$ is assigned to district $m \in \mathcal{M}$. The objective can be expressed as a function of the assignment variables y_{im}, and by solving the model (Location-Assignment), we can obtain each district $m \in \mathcal{M}$ as the subgraph induced by nodes $\mathcal{I}_m = \{i : y_{im} = 1\}$. Constraints (7.1c)–(7.1f) formulate the various districting criteria, with details stated as follows.

Contiguity

In many real-world scenarios, each district is required to be contiguous, i.e., it is possible to travel between any two points in the same district without having to traverse any other district. For a given graph $\mathcal{G} = (\mathcal{I}, \mathcal{E})$, there exist an exponential number $2^{|\mathcal{I}|}$ of possible districts. To avoid enumerating all feasible districts, we develop a network-flow based technique to ensure the district contiguity criterion.

For each of the M facilities corresponding to the districts $\mathcal{M}$, we imagine that the network $\mathcal{G}$ contains a corresponding "proxy" sink node that each can absorb any amount of incoming flow. We also assume that one unit of virtual flow is generated at each node $i \in \mathcal{I}$, and it will flow through the network to finally reach one of the M sink nodes. If each district contains exactly one sink node to receive all the virtual flow that has originated within the district, and if we can force that

no flow is allowed between different districts, then we have successfully partitioned $\mathcal{G}$ into several mutually disjoint but connected districts. Figure 7.2(a) illustrates a feasible network partition with four districts, and an associated virtual flow pattern. In the figure, each arrow indicates the virtual flow between nodes, and the number near the arrow gives the flow volume. The nodes in each partitioned district have the same shape (e.g., circle, square), while the sink node of this district is hollow. As shown, a partition is feasible if all virtual flow from a district could be collected and routed to its sink without passing any nodes outside of the district. Figure 7.2(b), in contrast, illustrates an infeasible partition, because there is no way to collect and route virtual flow from all "circle" nodes to a single sink node in this district, i.e., the district is not contiguous.



(a) Feasible partition (b) Infeasible partition

Figure 7.2: Network-flow based technique to model contiguity criterion.

For any $i \in \mathcal{I}, m \in \mathcal{M}$, we let variable $y_{im} = 1$ if node i is in district m, or $y_{im} = 0$ otherwise. In addition, we let $x_{im} = 1$ if node i is the sink node of district m, or $x_{im} = 0$ otherwise. We further let $z^m_{i_1 i_2} = 1$ if nodes i_1 and i_2 are adjacent in network G and both of them are in district m, or $z^m_{i_1 i_2} = 0$ otherwise. The virtual link flow from any node i_1 to any node i_2 is denoted as $f^m_{i_1 i_2}, \forall i_1, i_2 \in \mathcal{I}$. The contiguity of district $m \in \mathcal{M}$ is then ensured by the following constraints:

$$\sum_{m \in \mathcal{M}} y_{im} = 1, \ \forall i \in \mathcal{I}, \tag{7.2a}$$

$$\sum_{i \in \mathcal{I}} x_{im} = 1, \ \forall m \in \mathcal{M}, \tag{7.2b}$$

$$x_{im} \leq y_{im}, \ \forall i \in \mathcal{I}, m \in \mathcal{M}, \tag{7.2c}$$

$$z^m_{i_1 i_2} \le \delta_{i_1 i_2} \cdot \frac{y_{i_1 m} + y_{i_2 m}}{2},\ \forall i_1, i_2 \in \mathcal{I}, m \in \mathcal{M}, \tag{7.2d}$$

$$f^m_{i_1 i_2} \le (|\mathcal{I}| - |\mathcal{M}|)\, z^m_{i_1 i_2},\ \forall i_1, i_2 \in \mathcal{I}, m \in \mathcal{M}, \tag{7.2e}$$

$$y_{i_1 m} + \sum_{i_2 \in \mathcal{I}} f^m_{i_2 i_1} \ge \sum_{i_2 \in \mathcal{I}} f^m_{i_1 i_2},\ \forall i_1 \in \mathcal{I}, m \in \mathcal{M}, \tag{7.2f}$$

$$y_{i_1 m} + \sum_{i_2 \in \mathcal{I}} f^m_{i_2 i_1} \le \sum_{i_2 \in \mathcal{I}} f^m_{i_1 i_2} + (|\mathcal{I}| - |\mathcal{M}|)\, x_{i_1 m},\ \forall i_1 \in \mathcal{I}, m \in \mathcal{M}. \tag{7.2g}$$

Constraints (7.2a) enforce that each node $i \in \mathcal{I}$ belongs to exactly one district. Constraints (7.2b) make sure that each district has exactly one sink node. Constraints (7.2c) ensure that the sink node of each district should be inside the district due to the contiguity requirement. Constraints (7.2d) and (7.2e), together with integrality of the respective decision variables, require that virtual flow only exists on links within each district, while $|\mathcal{I}| - |\mathcal{M}|$ is the maximum possible flow volume on a link. Constraints (7.2f) and (7.2g) stipulate flow conservation at nodes in the network, except for those sink nodes (i.e., any node $i \in \mathcal{I}$ with $x_{im} = 1$).

Reliable Assignment

Due to exogenous reasons (such as adverse weather, power outage, etc.), each facility is subject to independent disruptions with an identical probability q, which could be defined as the fraction of time in a relatively long horizon (e.g., a year) during which a facility is in a disrupted state (e.g., due to adverse weather or labor issues). Once a facility fails, all the demands originally assigned to it are either served by another functioning backup facility, or the service is lost. For each district $m \in \mathcal{M}$, we plan a series of backup plans that involve up to $R \ge 1$ other facilities. We call its rth choice of facility as its level-r choice, while the level-0 choice is its original associated facility m. The district receives service from its level-r facility if its level-0, $\cdots$, level-$(r-1)$ choices have all become unavailable, which occurs with a corresponding probability of $(1-q)q^r$. To the very extreme, if all of the district's $R+1$ assigned facilities (including the level-0 choice) have failed, which occurs with a probability of q^{R+1}, every unit of demand in the district will incur a penalty of W_{missing}. Without loss of generality, we construct a dummy facility and assign all "lost" demand to this dummy facility.

We let variable $Y^r_{mn} = 1$ if district $m \in \mathcal{M}$ is assigned to facility $n \in \mathcal{M}$ at level r, or $Y^r_{mn} = 0$ otherwise. Similarly, let $Y^r_{m0} = 1$ if district $m \in M$ is assigned to the dummy facility at level r, or

$Y^r_{m0} = 0$ otherwise. The following constraints enforce such back-up assignments to facilities:

$$\sum_{n\in\mathcal{M},n\neq m} Y^r_{mn} + Y^r_{m0} = 1,\ \forall m \in \mathcal{M}, r = 1, 2, \cdots, R, \tag{7.3a}$$

$$Y^r_{m0} \leq Y^{r+1}_{j0},\ \forall m \in \mathcal{M}, r = 1, 2, \cdots, R, \tag{7.3b}$$

$$Y^{R+1}_{m0} = 1,\ \forall m \in \mathcal{M}. \tag{7.3c}$$

Here, constraints (7.3a) ensure that at each backup level, each district is assigned to either a regular facility or the dummy facility. Constraints (7.3b) enforce that if a district is assigned to the dummy facility at some level r, it is assigned to the dummy facility at all higher levels $r+1, \cdots, R+1$. Constraints (7.3c) guarantee that all districts are assigned to the dummy facility at level $R+1$.

Even in a facility failure scenario, the continuity requirement mandates that any districts (re-)assigned to a functioning facility should still be a connected graph. To ensure this, we enforce a strong requirement that if district m_1 can be re-assigned to a backup facility corresponding to district m_2, then districts of indices m_1 and m_2 must be adjacent to each other, i.e., there exist two nodes $i_1 \in \mathcal{I}_{m_1}$ and $i_2 \in \mathcal{I}_{m_2}$ such that $\delta_{i_1 i_2} = 1$. With this restriction, if a district has a limited number of neighbors, we may not be able to assign a district $m \in \mathcal{M}$ to a regular facility at some level $r \leq R$; instead, the district will be re-assigned to the dummy facility with probability q^r. As such, the probability that j is assigned to the dummy facility is calculated as $\sum_{s=r}^{R}(1-q)q^s + q^{R+1} = q^r$, which equals to the true probability for service loss. We define $w^{m_1 m_2}_{i_1 i_2} = 1$ if nodes i_1 and i_2 belong respectively to districts m_1 and m_2, or $w^{m_1 m_2}_{i_1 i_2} = 0$ otherwise. Furthermore, we define $l_{m_1 m_2} = 1$ if districts m_1 and m_2 are connected, and 0 otherwise. Then, the contiguity requirement can still be enforced under facility disruption and district reassignment scenarios by the following additional constraints and the respective integralities of decision variables:

$$w^{m_1 m_2}_{i_1 i_2} \leq \delta_{i_1 i_2} \cdot \frac{y_{i_1 m_1} + y_{i_2 m_2}}{2},\ \forall i_1, i_2 \in \mathcal{I}, m_1, m_2 \in \mathcal{M}, \tag{7.4a}$$

$$l_{m_1 m_2} \leq \sum_{i_1 \in I} \sum_{i_2 \in I} w^{m_1 m_2}_{i_1 i_2},\ \forall m_1, m_2 \in \mathcal{M}, \tag{7.4b}$$

$$\sum_{r=1}^{R} Y^r_{m_1 m_2} \leq l_{m_1 m_2},\ \forall m_1, m_2 \in \mathcal{M}. \tag{7.4c}$$

Constraints (7.4a) and (7.4b) determine whether two districts $m_1 \in \mathcal{M}$ and $m_2 \in \mathcal{M}$ are adjacent

to each other. Constraints (7.4c) enforce that a district can only be re-assigned to another facility corresponding to an adjacent district.

Workload Balance

We aim at balancing the expected total demands assigned to each facility across all normal and disruption scenarios. It is easy to see that such expected demand for a regular facility $m \in \mathcal{M}$, X_m, can be derived as follows:

$$X_m = \sum_{i \in \mathcal{I}} y_{im} D_i (1-q) + \sum_{n \in \mathcal{M}} \sum_{r=1}^{R} Y_{nm}^r \sum_{i \in \mathcal{I}} y_{im} D_i (1-q) q^r, \ \forall m \in \mathcal{M}. \tag{7.5}$$

The first and second terms are respectively the expected cost of serving the demand in district m via facility m and all possible re-assignments. Similarly, the expected demand assigned to the dummy facility can be expressed as follows:

$$X_0 = \sum_{m \in \mathcal{M}} Y_{m0}^{R+1} \sum_{i \in \mathcal{I}} y_{im} D_i q^{R+1} + \sum_{m \in \mathcal{M}} \sum_{r=1}^{R} Y_{m0}^r \sum_{i \in \mathcal{I}} y_{im} D_i (1-q) q^r. \tag{7.6}$$

A simple way to balance the workload is to minimize the maximum value of $\{X_m\}_{m \in \mathcal{M}}$, i.e., $X_{\max} := \max_{m \in \mathcal{M}} X_m$. Note that the following constraints hold:

$$X_{\max} \geq X_m, \ \forall m \in \mathcal{M}. \tag{7.7}$$

Compactness

The compactness of districts prevents the formation of oddly-shaped districts, and high compactness indicates that any district should be circular or square in shape rather than being elongated. In our reliable network districting problem, due to possible disruptions of facilities, a demand may be serviced by different facilities at different distances/costs. Therefore, we define the "expected" compactness of a district m based on two levels of spatial hierarchy: (i) the "median"-type total weighted distance for all nodal demand within district m to its choice-0 desk (i.e., the sink within the district), and (ii) the expected distance from the "own" sink node of district m to all other "backup" sink nodes at different choice levels. For a district $m \in \mathcal{M}$, we define $v_{i_1 i_2}^m = 1$ if $i_1 \in \mathcal{I}_m$ and i_2 is the sink node of district j, or $v_{i_1 i_2}^m = 0$ otherwise. We also denote the network shortest

path distance between nodes i_1 and i_2 as $d_{i_1 i_2}$. The shortest path distance between the sink nodes of two districts m_1 and m_2, which is a decision variable that depends on the corresponding sink locations, is denoted as $\hat{d}_{m_1 m_2}$. We know the following must hold:

$$v^m_{i_1 i_2} \geq y_{i_1 m} + x_{i_2 m} - 1, \forall i_1, i_2 \in \mathcal{I}, \tag{7.8a}$$

$$\hat{d}_{m_1 m_2} \geq d_{i_1 i_2} \left(x_{i_1 m_1} + x_{i_2 m_2} - 1 \right), \ \forall i_1, i_2 \in \mathcal{I}, m_1, m_2 \in \mathcal{M}. \tag{7.8b}$$

Constraints (7.8a) determine whether a node $i_1 \in \mathcal{I}$ and a sink $i_2 \in \mathcal{I}$ are in the same district. Constraints (7.8b) compute the distance between the sink nodes of two districts m and n. Then the compactness measure for district m is calculated as

$$C_m = \sum_{i_1 \in \mathcal{I}} \sum_{i_2 \in \mathcal{I}} d_{i_1 i_2} v^m_{i_1 i_2} + \alpha \sum_{n \in \mathcal{M}} \sum_{r=1}^{R} (1-q) q^r Y^r_{mn} \hat{d}_{mn}, \ \forall m \in \mathcal{M}, \tag{7.9}$$

where α is a weight parameter. A larger value of C_m indicates a less compact district m.

Model formulation

Now, the reliable network districting problem (RND) can be formulated as the following mixed-integer programming model:

$$\text{(RND)} \quad \min \quad W_{\text{balance}} \cdot X_{\max} + W_{\text{missing}} \cdot X_0 + W_{\text{compact}} \cdot \sum_{m \in \mathcal{M}} C_m \tag{7.10a}$$

$$\text{s.t.} \quad (7.2\text{a}) - (7.2\text{g}), (7.3\text{a}) - (7.3\text{c}), (7.4\text{a}) - (7.4\text{c}),$$

$$(7.5), (7.6), (7.7), (7.8\text{a}) - (7.8\text{b}), (7.9),$$

$$y_{im}, x_{im}, z^m_{i_1 i_2}, w^{m_1 m_2}_{i_1 i_2}, l_{m_1 m_2}, v^m_{i_1 i_2}, Y^r_{m_1 m_2}, Y^r_{m0} \in \{0,1\}, f^m_{i_1 i_2}, \hat{d}_{m_1 m_2} \geq 0,$$

$$\forall i, i_1, i_2 \in \mathcal{I}, m, m_1, m_2 \in \mathcal{M}, r = 1, 2, \cdots, R, \tag{7.10b}$$

where W_{balance}, W_{compact} and W_{missing} are the relative weight coefficients for the maximum facility workload $X_{\max}$, the compactness measure C_m, and the dummy facility workload X_0, respectively. The objective function (7.10a) presents the expected system cost including the "cost" for workload balancing, the "cost" for compactness, and the penalty for demand loss. We assume that $W_{\text{missing}} > W_{\text{balance}}$, otherwise the problem becomes trivial such that all districts will be assigned to the dummy

desk at level 1 and no further backup plan is needed. In addition to Constraints (7.2a) – (7.9), Constraints (7.10b) enforce the integrity and non-negativity of all decision variables.

The formulation (RND) contains several nonlinear terms $Y^r_{mn}y_{im}$, $Y^r_{m0}y_{im}$ and $Y^r_{mn}\hat{d}_{mn}$ in several sets of constraints. We linearize them by applying a variant of the technique introduced by Sherali and Alameddine (1992), whereas $Y^r_{mn}y_{im}$, $Y^r_{m0}y_{im}$ and $Y^r_{mn}\hat{d}_{mn}$ are replaced by new continuous variables U^{ir}_{mn}, U^{ir}_{m0} and U^r_{mn}, respectively. Their equivalence is enforced by adding the following new constraints.

$$U^{ir}_{mn} \geq Y^r_{mn} + y_{im} - 1, \ \forall m,n \in \mathcal{M}, i \in \mathcal{I}, r = 1,2,\cdots,R, \tag{7.11a}$$

$$U^{ir}_{mn} \leq Y^r_{mn}, \ \forall m,n \in \mathcal{M}, i \in \mathcal{I}, r = 1,2,\cdots,R, \tag{7.11b}$$

$$U^{ir}_{mn} \leq y_{im}, \ \forall m,n \in \mathcal{M}, i \in \mathcal{I}, r = 1,2,\cdots,R, \tag{7.11c}$$

$$U^{ir}_{m0} \geq Y^r_{m0} + y_{im} - 1, \ \forall m \in \mathcal{M}, i \in \mathcal{I}, r = 1,2,\cdots,R+1, \tag{7.11d}$$

$$U^{ir}_{m0} \leq Y^r_{m0}, \ \forall m \in \mathcal{M}, i \in \mathcal{I}, r = 1,2,\cdots,R+1, \tag{7.11e}$$

$$U^{ir}_{m0} \leq y_{im}, \ \forall m \in \mathcal{M}, i \in \mathcal{I}, r = 1,2,\cdots,R+1, \tag{7.11f}$$

$$U^r_{mn} \geq \hat{d}_{mn} + d_{\max}\left(Y^r_{mn} - 1\right), \ \forall m,n \in \mathcal{M}, r = 1,2,\cdots,R, \tag{7.11g}$$

$$U^r_{mn} \leq \hat{d}_{mn}, \ \forall m,n \in \mathcal{M}, r = 1,2,\cdots,R, \tag{7.11h}$$

$$U^r_{mn} \leq d_{\max}Y^r_{mn}, \ \forall m,n \in \mathcal{M}, r = 1,2,\cdots,R, \tag{7.11i}$$

$$U^{ir_1}_{mn}, U^{ir_2}_{m0}, U^{r_1}_{mn} \geq 0, \ \forall m,n \in \mathcal{M}, i \in \mathcal{I}, r_1 = 1,2,\cdots,R, r_2 = 1,2,\cdots,R+1, \tag{7.11j}$$

where $d_{\max}$ is the maximum distance between any two nodes in network $\mathcal{G}$. Note from (7.11a)-(7.11c) that $U^{ir}_{mn} = 1$ if and only if $Y^r_{mn} = y_{im} = 1$; if either Y^r_{mn} or y_{im} equals 0, then $U^{ir}_{mn} = 0$ as well. Hence, these constraints ensure that $U^{ir}_{mn} = Y^r_{mn}y_{im}$ exactly. Similarly, (7.11d)-(7.11f) ensure $U^{ir}_{m0} = Y^r_{m0}y_{im}$. From (7.11g)-(7.11i), and the fact that $\hat{d}_{mn} - d_{\max} \leq 0 \leq \hat{d}_{mn} \leq d_{\max}$, we know that $U^r_{mn} = \hat{d}_{mn}$ if $Y^r_{mn} = 1$, or 0 otherwise. These constraints exactly function in the same way as $U^r_{mn} = \hat{d}_{mn}Y^r_{mn}$. The model formulation (RND) is then transformed into the following linearized reliable network districting problem (LRND):

$$\text{(LRND)} \quad \min \quad W_{\text{balance}} \cdot X_{\max} + W_{\text{missing}} \cdot X_0 + W_{\text{compact}} \cdot \sum_{m \in \mathcal{M}} C_m \tag{7.12a}$$

$$\text{s.t.} \quad (7.2a)-(7.2g), (7.3a)-(7.3c), (7.4a)-(7.4c),$$

$$(7.7), (7.8a)-(7.8b), (7.9), (7.11a)-(7.11j),$$

$$X_m = \sum_{n\in\mathcal{M}}\sum_{r=1}^{R}\sum_{i\in\mathcal{I}} U_{nm}^{ir} D_i(1-q)q^r + \sum_{i\in\mathcal{I}} y_{im} D_i(1-q),\ \forall m \in \mathcal{M}, \tag{7.12b}$$

$$X_0 = \sum_{m\in\mathcal{M}}\sum_{r=1}^{R}\sum_{i\in\mathcal{I}} U_{m0}^{ir} D_i(1-q)q^r + \sum_{m\in\mathcal{M}}\sum_{i\in\mathcal{I}} U_{m0}^{iR+1} D_i q^{R+1}, \tag{7.12c}$$

$$C_m = \sum_{i_1\in\mathcal{I}}\sum_{i_2\in\mathcal{I}} D_{i_1 i_2} v_{i_1 i_2}^m + \sum_{n\in\mathcal{M}}\sum_{r=1}^{R} U_{mn}^r (1-q)q^r,\ \forall m \in \mathcal{M}. \tag{7.12d}$$

7.4 Solution Algorithm

The mixed-integer linear program (LRND) could be potentially solved by commercial solvers such as CPLEX or Gurobi. However, as we will show in the case studies in Section 7.6, the computational burden is greatly exacerbated by the network-flow based constraints and the reliable assignment strategy, especially when the network size is relatively large. It takes a large amount of computation time for the solvers to obtain even a feasible solution (usually with a quite poor quality). In light of this challenge, in this section, we develop a customized heuristic algorithm to obtain solutions in a reasonable amount of time. The algorithm contains two parts: (i) a constructive heuristic to obtain an initial solution, and (ii) a neighborhood search procedure to improve the solution.

7.4.1 Constructive Heuristic

We first propose a constructive heuristic to obtain an initial solution to (LRND). We define the deterministic version (built facilities are always functioning) of the linearized network districting problem as follows:

$$\text{(DLRND)} \quad \min \quad W_{\text{balance}} \cdot X_{\max} + W_{\text{compact}} \cdot \sum_{m\in\mathcal{M}} C_m \tag{7.13a}$$

$$\text{s.t.} \quad (7.2a)-(7.2g),$$

$$X_m = \sum_{i\in\mathcal{I}} y_{im} D_i,\ \forall n \in \mathcal{M}, \tag{7.13b}$$

$$X_{\max} \geq X_m,\ \forall m \in \mathcal{M}, \tag{7.13c}$$

$$v_{i_1 i_2}^m \geq y_{i_1 m} + x_{i_2 m} - 1,\ \forall i_1, i_2 \in \mathcal{I}, m \in \mathcal{M}, \tag{7.13d}$$

$$C_m = \sum_{i_1 \in \mathcal{I}} \sum_{i_2 \in \mathcal{I}} d_{i_1 i_2} v^m_{i_1 i_2}, \ \forall m \in \mathcal{M}. \tag{7.13e}$$

In this problem, the operational criteria/requirements including district contiguity, workload balance, and district compactness are all considered, so it is likely to generate an initial solution with a reasonably good quality. The detailed steps of the constructive heuristic to solve (DLRND) are described as follows:

Step 1. Solve a median problem where M sinks are initially located at selected nodes in the network $\mathcal{G}$. Note that the districts in the solution of the median problem are not necessarily contiguous due to workload balancing.

Step 2. Starting with the M initial sink nodes, we expand them into a set $\mathcal{M}$ of contiguous districts as follows:

(i) Compute the average demand of each district as $\bar{D} = \sum_{i \in \mathcal{I}} D_i / M$, and set a demand threshold $\widehat{D}$ as a function of $\bar{D}$, e.g., $\widehat{D} = \beta \bar{D}$, or $\widehat{D} = \bar{D} + \text{Constant}$;

(ii) For each district $m \in \mathcal{M}$ with demand $\sum_{i \in \mathcal{I}_m} D_i$, if there exists one node i such that i is adjacent to m and $\sum_{i' \in \mathcal{I}_m} D_{i'} + D_i \leq \widehat{D}$, we expand district m by attaching node i and the associated links to it;

(iii) If no district can be further expanded without violating the workload limit, but there remain some nodes that have not been included in any district, do the following. For each remaining node i, we try to attach it to each of its adjacent districts and evaluate the resulting workload balance; select the attachment that yields the least workload balance violation.

Step 3. Given the set $\mathcal{M}$ of districts generated in Step 2, we check for any pair of districts m_1 and m_2, if they are adjacent and the difference between their demands exceeds a specified threshold $\widetilde{D}$ (e.g., $\widetilde{D} = \gamma \bar{D}$), we run (DLRND) to re-partition the network with districts other than m_1 and m_2 fixed.

This process continues until there exists no such pair of districts, or when some other termination criteria are met, e.g., the maximum number of iterations is reached.

7.4.2 Neighborhood Search

The solution obtained by the constructive heuristic does not consider the possible facility disruptions. To incorporate the reliability issue into the solution, we develop a neighborhood search heuristic. In the solution space, a neighborhood of a districting solution (i.e., the partition) is defined as another partition with two of the adjacent districts redesigned (i.e., the nodes in the two districts are reassigned to yield two new districts) while all other districts remain the same.

Specifically, given the set $\mathcal{M}$ of districts, if districts m_1 and m_2 are to be redesigned, we let $L = \mathcal{M}\backslash\{m_1, m_2\}$, fix the following variables, and solve (LRND) again to re-partition the network:

(i) $x_{im}, y_{im}, z^m_{i_1 i_2}, f^m_{i_1 i_2}, v^m_{i_1 i_2},\ \forall i, i_1, i_2 \in \mathcal{I}, m \in \mathcal{L}$;

(ii) $w^{m_1 m_2}_{i_1 i_2},\ \forall m_1, m_2 \in \mathcal{M}$;

(iii) $l_{m_1 m_2}, \hat{d}_{m_1 m_2},\ \forall m_1, m_2 \in \mathcal{M}$;

After solving the (LRND) model, if the new solution is feasible and yields an improvement to the objective, it will be accepted and the partition solution will be updated. Then the set of possible neighborhood moves is updated and the neighborhood search process starts again. The searching procedure continues until some stopping criteria are met, or if all possible neighborhood moves have been enumerated without yielding any improvements.

The above algorithm re-designs two neighboring districts at a time. In general, any number of districts can be redesigned together in one search move. Since the assignment of demand in each district is interrelated with decisions in many other districts, re-partitioning multiple districts in one move may lead to a better solution or help avoid a local optimum. However, in so doing, the corresponding (LRND) in each move takes a longer time to solve. So we set a max time for each move, taking into account the tradeoff between the improvement of solution quality and the computation time in each move.

7.5 Error Estimation

The customized heuristic algorithm provides a feasible solution to (LRND). In this section, we use set-cover based modeling techniques to construct a lower bound to (LRND), so as to obtain the optimality of the heuristic solution and demonstrate its good quality.

7.5.1 Initial Lower Bound

We split the objective of (LRND) into three components and calculate the lower bound to each component, which in total give an initial integrated lower bound.

(i) Since $W_{\text{balance}} < W_{\text{missing}}$, if a district m has an adjacent district n, we can prove that the corresponding facility n is always chosen by demands in district m at a lower level than the emergency facility. Since each district has at least one adjacent district, then we can obtain a lower bound to $X_{\max}$ as

$$\widehat{X_{\max}} \geq \frac{\sum_{i\in\mathcal{I}} D_i(1-q^2)}{M}. \tag{7.14}$$

(ii) Each district $m \in \mathcal{M}$ is forced to be assigned to the dummy facility at one level $r \in [1, 2, \cdots, R+1]$, and contribute $\sum_{i\in\mathcal{I}_m} D_i q^r$ to X_0. So the minimum value of X_0 can be expressed as

$$\widehat{X_0} \geq \sum_{i\in\mathcal{I}} D_i q^{R+1}. \tag{7.15}$$

(iii) Given the compactness measure (7.9), since each district $m \in \mathcal{M}$ has at least one adjacent district, we have $\sum_{r=1}^{R}(1-q)q^r Y^r_{mn}\hat{d}_{mn} \geq (1-q)q^r\hat{d}_{mn_m}$ where n_m is the district with sink node closest to the sink node of district m. Therefore, we can calculate the lower bound of $\sum_{m\in\mathcal{M}} C_m$ as

$$\begin{aligned}\widehat{\sum_{m\in\mathcal{M}} C_m} = \min & \sum_{m\in\mathcal{M}}\sum_{i_1\in\mathcal{I}}\sum_{i_2\in\mathcal{I}} D_{i_1 i_2} v^m_{i_1 i_2} + \alpha \sum_{m\in\mathcal{M}} \hat{d}_m(1-q)q \\ & \text{s.t. } (7.2\text{a}) - (7.2\text{g}), (7.8\text{a}) - (7.8\text{b}), \\ & \hat{d}_m = \sum_{n\in\mathcal{M}} \hat{d}_{mn} u_{mn},\ \forall m \in \mathcal{M}, \\ & \sum_{n\in\mathcal{M}} u_{mn} = 1,\ \forall m \in \mathcal{M},\end{aligned} \tag{7.16}$$

where $\hat{d}_m$ is the minimum distance between the sink node of district $m \in \mathcal{M}$ and any other district $n \neq m$, and u_{mn} indicates whether the sink of district n is closest to the sink of district m, $u_{mn} = 1$ if it is, 0 otherwise.

Combining (i) (ii) and (iii) gives the overall lower bound to (LRND) and the corresponding error/gap estimation as

$$\text{LB-INIT}_{(\text{LRND})} = W_{\text{balance}} \cdot \widehat{X_{\max}} + W_{\text{missing}} \cdot \widehat{X_0} + W_{\text{compact}} \cdot \widehat{\sum_{m \in \mathcal{M}} C_m}. \tag{7.17}$$

$$\text{Error}_{(\text{LRND})} = \frac{\text{UB}_{(\text{LRND})} - \text{LB}_{(\text{LRND})}}{\text{UB}_{(\text{LRND})}} \times 100\% \tag{7.18}$$

Note that since the lower bounds to the three cost components are not necessarily achieved simultaneously, the initial lower bound $\text{LB-INIT}_{(\text{LRND})}$ may not be tight. In the next two sections, we will use a set-cover based modeling techniques to further improve it.

7.5.2 District Filtering

To apply set-cover based modeling approaches, we have to first generate a set of possible district options. Given network $\mathcal{G} = (\mathcal{I}, \mathcal{E})$, the maximum number of all possible districts can be as large as $O(2^{|\mathcal{I}|})$, which is exponentially large. Therefore, it is extremely difficult, if not impossible, to enumerate all of the possible districts. We instead filter districts on a "possible to be optimal" basis by excluding those districts that are impossible to be in the optimal solution. The algorithm to generate the set of feasible district options can be summarized as the following steps.

Step 1 We use the heuristic algorithm proposed in Section 7.4 to obtain the upper bound UB_{LRND}, which is no less than the optimal objective value.

Step 2 For any district $m \in \mathcal{M}$, assume that its demand is D_m and minimum compactness measure is C_m (which can be calculated beforehand), we can compute the lower bound to the compactness of the other $M - 1$ districts easily as C_m^{others}.

Step 3 If the lower bound to the corresponding objective gives

$$W_{\text{balance}} \cdot D_m(1 - q) + W_{\text{compact}} \cdot (C_m + C_m^{\text{others}}) + W_{\text{missing}} \cdot \widehat{X_0} > \text{UB}_{\text{LRND}},$$

we claim that district m is impossible to be in the optimal solution, and thus exclude it from the set of possible districts.

In this way, a very large portion of the district options is eliminated from the initial set. The

pseudo-code for filtering the district options is summarized as Algorithm 2.

7.5.3 Improved Lower Bound

After filtering the district options, we obtain a set $\mathcal{J}$ of districts which is of limited size. Further, by observing that the cost components with X_0 and C_m are typically restricting each other, i.e., when X_0 is small, each district tends to have more neighbors and thus is less compact in shape with larger value of C_m. Therefore, we use a set-cover model to integrate these two components and improve the initial lower bound. The model formulation is presented as follows

$$\text{(LRND-LB-IMPR)} \quad \min \quad W_{\text{missing}} \cdot X_0 + W_{\text{compact}} \cdot \sum_{j \in \mathcal{J}} C_j Z_j$$

$$\text{s.t.} \quad \sum_{j \in \mathcal{J}} Z_j = M \tag{7.19a}$$

$$\sum_{j \in \mathcal{J}} a_{ij} Z_j = 1, \ \forall i \in \mathcal{I} \tag{7.19b}$$

$$\sum_{r=1}^{R} Y_j^r \leq \sum_{k \in \mathcal{J}} l_{jk} Z_k, \ \forall j \in \mathcal{J} \tag{7.19c}$$

$$Y_j + \sum_{s=1}^{r} Y_{j0}^s = Z_j, \ \forall j \in \mathcal{J}, r \leq R \tag{7.19d}$$

$$\sum_{r=1}^{R+1} Y_{j0}^r = Z_j, \forall j \in \mathcal{J}, \tag{7.19e}$$

$$X_0 = \sum_{j \in \mathcal{J}} \sum_{r=1}^{R+1} Y_{j0}^r d_j q^r, \tag{7.19f}$$

$$Z_j, Y_j^r, Y_{j0}^r \in \{0,1\}, \ \forall j \in \mathcal{J}, r = 1, 2, \cdots, R+1. \tag{7.19g}$$

Constraints (7.19a) enforce that exactly M districts are selected from all the district options. Constraints (7.19b) ensure that each customer is contained in exactly one district, i.e., selected districts are non-overlapping and cover the entire network. Constraints (7.19c) indicate that customers can only be assigned to facilities corresponding to those selected districts. Constraints (7.19d)–(7.19e) pose requirements on district assignments, ensuring that customers are first assigned to regular facilities, then to the emergency facility at the highest level. Constraints (7.19f) calculate the value of X_0. Given the optimal value of (LRND-LB-IMPR) as $\text{OPT}_{\text{(LRND-LB-IMPR)}}$, we have the improved

Algorithm 2 Generate the set of feasible district options

DistrictGeneration()

1. **for** $i \in \mathcal{I}$ **do**
2. sourceHeap← createHeap(i), stack← $\emptyset$, districts← $\emptyset$
3. stack.push(DFSNode(i, sourceHeap, districts))
4. **while** stack is not empty **do**
5. node ← stack.pop()
6. **if** $W_{\text{weight}} \cdot$ node.demand $\cdot(1-q) + W_{\text{compact}} \cdot$ node.compact $\leq \overline{W}$ **then**
7. **if** node.hash is in districts.hashList **then**
8. districts.setList.add(node.set), district.hashList.add(node.hash)
9. **end if**
10. **if** node.unvisited is not empty **then**
11. nextNode = node.unvisited.extractmin()
12. **if** node.unvisited is not empty **then**
13. stack.push(node)
14. nextHeap ← updateHeap(node, next, i,)
15. **else**
16. nextHeap ← createHeap(node, next, i,)
17. **end if**
18. newNode
19. stack.push(newNode)
20. **end if**
21. **end if**
22. **end while**
23. remove i from $\mathcal{G}$
24. **end for**

createHeap(i, visitedNodes)

1. newHeap ← $\emptyset$
2. **for** $n \in \mathcal{E}(i)$ **do**
3. **if** n is not in visitedNodes **then**
4. newHeap.insert(n)
5. **end if**
6. **end for**
7. **return** newHeap

updateHeap(i, heap, visitedNodes)

1. newHeap ← heap
2. **for** $n \in \mathcal{E}(i)$ **do**
3. **if** n is not in newHeap or visitedNodes **then**
4. newHeap.insert(n)
5. **end if**
6. **end for**
7. **return** newHeap

lower bound to (LRND) as

$$\text{LB-IMPR}_{\text{(LRND)}} = W_{\text{balance}} \cdot \widehat{X_{\max}} + \text{OPT}_{\text{(LRND-LB-IMPR)}}. \tag{7.20}$$

7.6 Numerical Examples

We apply the mathematical model and solution approaches to two sets of examples so as to demonstrate their applicability and performance. In the first example, we partition a series of hypothetical square grid networks of various sizes (as shown in Figure 7.3), and compare the performance of different solution approaches with different parameter settings. The second numerical example, on the other hand, involves a full-scale call center design problem for a U.S. Class I railroad company. The proposed model and solution algorithms are programmed in C# and run on a 64-bit Intel i7-3770 computer with 3.40 GHz CPU and 8G RAM. The mixed-integer linear programs, if solved directly, are tackled by commercial solver Gurobi.

7.6.1 Hypothetical Grid Networks

For each value $n \in \{4,5,6,7,8\}$, an $n \times n$ grid network is generated to represent a hypothetical study region with n^2 nodes and $2n(n-1)$ links. The n^2 nodes are indexed as $1,2,\cdots,n^2$ from left to right, and from bottom to top, as shown in Figure 7.3. The length of every edge between two adjacent nodes is set to 1. For node i, the demand is set to $30+20(\text{mod}(i,10)\cdot 10\pi/13)$. The failure probability of each location is assumed to be 0.1. We allow the number of districts $M \in \{5,7\}$, and the maximum assignment level $R \in \{3,4\}$. The penalty values are set as $W_{\text{balance}} = 1.0, W_{\text{compact}} = 2.0$, and $W_{\text{missing}} = 500.0$, respectively.

We use two different solution approaches to solve the reliable districting problems for these grid networks. The first approach directly applies Gurobi to solve the linearized programming models (7.12a) – (7.12d), while the computation time limit is set to 7200s. The second approach uses the proposed heuristic algorithm (i.e., construction and neighborhood search), the max time limit for each search move is set to 60s. Table 7.1 summarizes and compares the results obtained by the two approaches.

Overall, as the network size, the number of partitions, and/or the maximum number of backup levels R increase, the problem becomes more challenging to solve. Gurobi seems to have difficulty

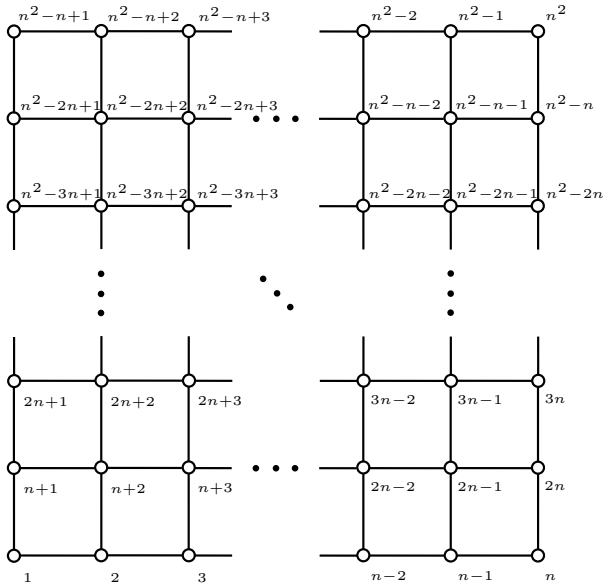


Figure 7.3: $n \times n$ square grid network.

closing the reported optimality gap for all the test cases within the 7200s computation time limit, while for all 8×8 cases, Gurobi even could not find a feasible solution. In contrast, our proposed heuristic approach is able to obtain reasonable solutions in a relatively short amount of computation time. By comparing the results obtained by the two approaches, we observe that their solutions are comparable only for moderate size problems (e.g., $n = 4$, 5 partitions, and R =3 or 4). When the problem size is large, the quality of the Gurobi solution is quite poor, i.e., its reported gap is very large and its best feasible solution has an objective value far exceeding that from the heuristic algorithm. As such, the proposed heuristic algorithm outperforms the Gurobi method in terms of both solution quality and computation time, especially for large-scale problem instances.

For cases with the same network size and R value, as the number of partitions $|J|$ increases, the computation time of our algorithm increases due to the larger number of variables and constraints. The objective value, however, reduces because with a larger number of partitions, the expected workload of each caller desk becomes smaller, and the probability of losing service due to insufficient adjacent districts is smaller. Similarly, a bigger value of R results in a longer computation time (also because of the increase in problem size) and a smaller solution objective, probably due to the lower chance for any demand to lose service.

A closer look at some solutions is provided in Table 7.2, where the solution details for the cases with $n = 8$ network, 7 partitions, and $R = 3$ or 4 are presented. Column 3 lists the nodes in each

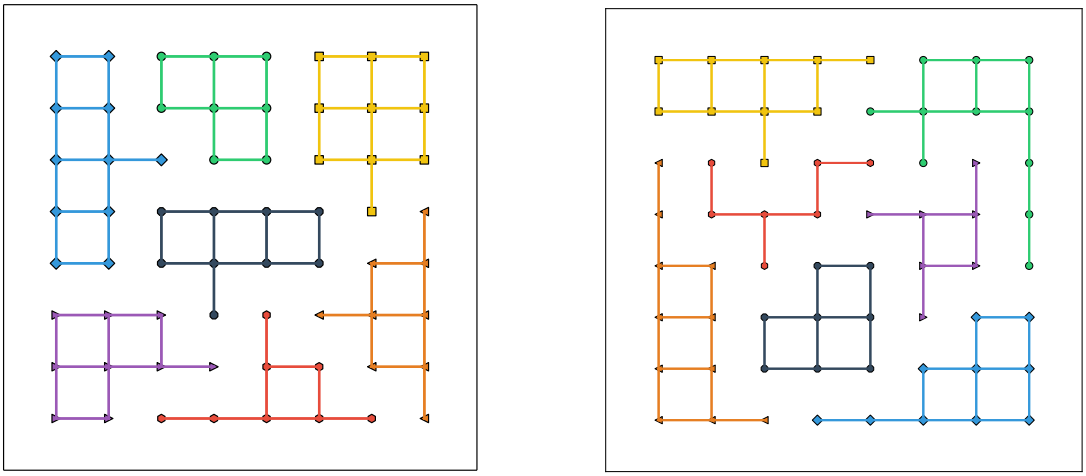
Table 7.1: Algorithm performance comparison for different test cases.

Network size	M	R	Lower bound	Heuristic objective	Heuristic gap (%)	Heuristic time (s)	Gurobi objective	Gurobi gap	Gurobi time (s)
4×4	5	3	141.882	147.601	3.9%	304	147.601	16.2%	7200
	5	4	137.336	144.067	4.7%	374	144.067	21.7%	7200
	7	3	112.588	115.163	2.2%	319	115.278	26.7%	7200
	7	4	99.114	102.642	3.4%	505	119.778	31.9%	7200
5×5	5	3	242.855	250.891	3.2%	277	253.079	28.0%	7200
	5	4	227.425	241.708	5.9%	630	260.021	34.6%	7200
	7	3	187.202	194.268	3.6%	412	214.327	50.6%	7200
	7	4	167.342	174.914	4.3%	856	223.066	66.1%	7200
6×6	5	3	364.477	372.739	2.6%	597	388.893	41.8%	7200
	5	4	332.461	362.608	8.3%	1891	546.601	143.3%	7200
	7	3	285.383	297.416	4.0%	1041	366.211	81.2%	7200
	7	4	247.965	266.872	7.1%	2518	589.741	287.6%	7200
7×7	5	3	522.309	528.513	1.2%	3189	593.627	68.4%	7200
	5	4	454.154	515.274	11.9%	4567	770.790	171.0%	7200
	7	3	412.809	422.193	2.2%	3867	2311.332	740.3%	7200
	7	4	344.654	372.789	7.5%	5845	2097.163	920.4%	7200
8×8	5	3	713.427	727.619	2.0%	3963	–	–	–
	5	4	623.922	714.946	12.7%	4327	–	–	–
	7	3	568.722	575.618	1.2%	4131	–	–	–
	7	4	479.217	521.694	8.1%	4697	–	–	–

Table 7.2: Detailed results of the 8×8 network partitioned into 7 districts.

	District	Nodes	Nodal Demand	X_m	C_m	1	2	3	4
$R = 3$	1	44, 45, 51, 52, 53, 59, 60, 61	259.16	287.06	10.30	4	6	7	–
	2	25, 26, 33, 34, 41, 42, 43, 49, 50, 57, 58	320.01	288.53	18.31	1	6	5	–
	3	8, 15, 16, 22, 23, 24, 31, 32, 40	265.14	288.38	13.50	4	6	7	–
	4	3, 4, 5, 6, 7, 13, 14, 21	237.83	269.17	11.49	3	6	5	–
	5	1, 2, 9, 10, 11, 12, 17, 18, 19	319.90	288.41	12.40	4	6	2	–
	6	20, 27, 28, 29, 30, 35, 36, 37, 38	274.79	283.73	13.31	1	4	3	–
	7	39, 46, 47, 48, 54, 55, 56, 62, 63, 64	312.25	283.59	14.51	3	6	1	–
$R = 4$	1	32, 40, 46, 48, 53, 54, 55, 56, 62, 63, 64	312.45	289.14	20.32	5	2	4	7
	2	4, 5, 6, 7, 8, 14, 15, 16, 23, 24	301.36	274.36	17.41	5	1	6	3
	3	1, 2, 3, 9, 10, 17, 18, 25, 26, 33, 41	303.55	275.89	21.31	6	4	2	7
	4	27, 34, 35, 36, 42, 44, 45	256.02	284.03	10.30	6	5	3	1
	5	22, 30, 31, 37, 38, 39, 47	256.04	288.20	9.39	6	1	4	2
	6	11, 12, 13, 19, 20, 21, 28, 29	238.08	287.95	10.30	4	3	5	2
	7	43, 49, 50, 51, 52, 57, 58, 59, 60, 61	321.57	289.47	16.22	4	1	3	–

district, Column 4 shows the total demand from all nodes in each district, and Column 5 gives the expected workload X_m (re-)assigned to each facility m in normal and disruption scenarios. Column 6 presents compactness metric C_m for each district. Columns 7–10 are the backup desk assignment decisions for each district at various backup levels. We observe that our proposed algorithm is able to achieve all partitioning objectives satisfactorily. Even though the original nodal demand from the districts are not quite even, the expected workloads (re-)assigned to all the desks are very well balanced. Figure 7.4 graphically illustrates the spatial partitions, where each connected subgraph represents one district. In general all the districts have reasonable compactness. Moreover, we observe that when $R = 4$, districts in the partition results are slightly more elongated in shape than those when $R = 3$, so as to have more adjacent districts (and hence, caller desks) as backup options. Specifically, when $R = 3$, all districts but one have exactly 3 adjacent districts, while in the case when $R = 4$, 6 out of the 7 districts have at least 4 adjacent districts.



(a) $R = 3$ (b) $R = 4$

Figure 7.4: The partition results of the 8×8 network under different values of R.

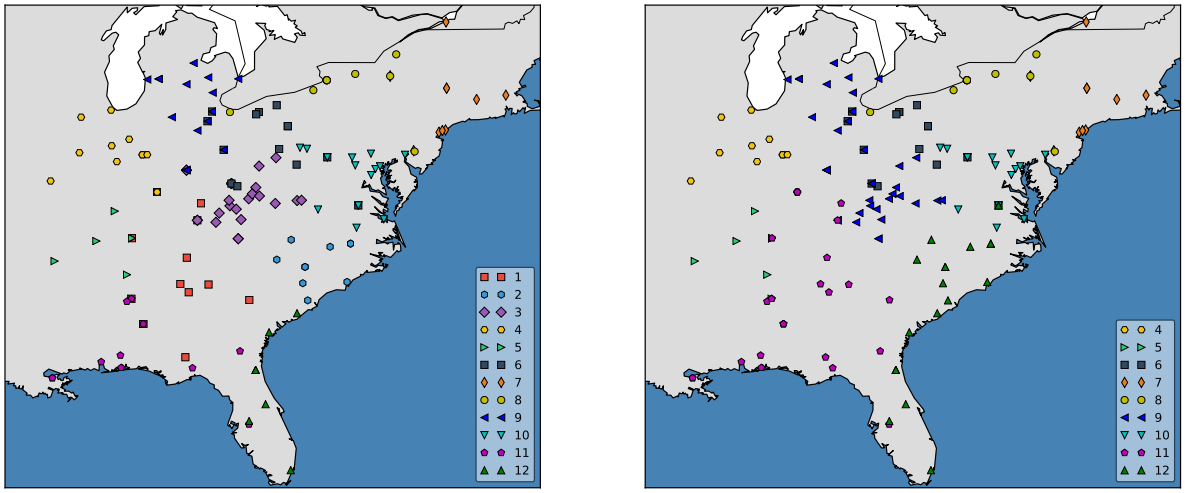
7.6.2 Full-scale Railroad Call Center Design

Next, we apply our methodology to an empirical call center design problem for a U.S. Class I railroad company. Railroad companies regularly receive a large number of calls from their customers and business partners (normally via telephone, intercom, computer or other devices) regarding real-time train operation, service scheduling, and resource/crew arrangements in the field. Such calls have significant implications on the operational efficiency of the railroad system and hence are important

for the success of the railroad companies. The call center serves as the information and decision hub for a railroad company's production and maintenance activities. The planning of the call center plays a critical role in the allocation and utilization of the railroad's manpower, equipment, and other resources. This is particularly the case when railroads are facing operational uncertainties and the risk of disruptions (e.g., due to adverse weather, train accidents, power outage, labor issues). In case of unexpected emergencies, effective communication and efficient resource (re-)allocation across the railroad network must be ensured via the call center such that backup plans for train timetabling, rolling stock and crew scheduling can be carried out in time – The call center itself must be functioning reliably so as to be responsible for real-time emergency management and disaster response. Therefore, a good call center design that can reliably and efficiently handle incoming calls across the railroad system is very critical.

Usually a call center for a railroad company consists of multiple crew caller desks, each of which is responsible for the calls from a particular predefined spatial region (i.e., district). A good call center design should have the following characteristics: (i) all incoming call demand can be handled by a properly assigned caller desk under any circumstance, otherwise significant penalty will occur due to disruption to railroad operations; (ii) the expected workload is well balanced across caller desks, so that no desks are too much more occupied than others; (iii) the spatial district served by a caller desk should be contiguous so as to satisfy a number of practical operational requirements, e.g., administrative autonomy for resource/crew reallocation and train traffic management; and (iv) the spatial district corresponding to one caller desk is compact in shape so as to avoid high transportation/logistics costs inside odd-shaped districts.

As shown in Figure 7.5, the railroad network consists of 157 nodes (the rail yards/supply points of the company), each with a specified demand of crew calls. The detailed data are confidential, but the maximum, minimum, and average of the nodal demand are 3219, 6, and 76, respectively, representing a huge disparity in demand quantity across nodes. The distance between two adjacent nodes is measured along the shortest path in the actual rail track network. The company's call center contains 12 caller desks. Thus the company wants the network to be partitioned into $M = 12$ corresponding districts and assigned to these desks. The maximum number of back levels is $R = 3$, the probability for each desk to fail is $q = 0.1$, and the penalty values are set as $W_{\text{balance}} = 1.0, W_{\text{compact}} = 2.0, W_{\text{missing}} = 500.0$, respectively.



(a) Districting plan with no disruption

(b) Districting plan when desks 1, 2, and 3 fail

Figure 7.5: Results for the full-scale railroad call center design problem.

Table 7.3: The demand, workload, and backup assignments of each district.

Index	1	2	3	4	5	6	7	8	9	10	11	12
Demand	9320	10240	9862	10239	10734	10970	8593	9747	9855	9635	10386	10261
Workload	10199	10160	9961	10182	9847	9995	9478	10012	9988	9639	10195	10175
Level 1	11	12	1	9	4	8	10	7	3	7	2	1
Level 2	5	10	9	3	3	9	8	6	8	8	5	10
Level 3	12	3	6	5	11	10	6	9	6	6	12	2

The commercial solvers cannot solve a problem of this size without running into memory issues. Results from the proposed heuristic algorithm is presented in Figure 7.5(a) and Table 7.3. In Figure 7.5(a), each color (or shape) of nodes corresponds to one district, which is contiguous and compact in shape in the company's network (although the detailed network is not shown). Table 7.3 shows the total nodal demand from each district and the expected workload (re-)assigned to each desk. The expected workloads of the desks are very well balanced, i.e., the biggest and smallest workloads are 10199 and 9478, respectively, with a mere 7.6% difference. Also, the $R = 3$ choices of backup desks of each district are summarized in Table 7.3. For example, nodal demands in district 1 are assigned to desks 11, 5, and 12 at backup levels 1, 2, and 3, respectively. Figure 7.5(b) then shows the assignments of nodal demands to desks when desks 1, 2, and 3 fail. We observe that nodes in the original districts 1, 2, and 3 are reassigned to desks 11 (1st backup of district 1), 12 (1st backup of district 2), and 9 (2nd backup of district 3 since its 1st backup is disrupted already), respectively. The company considers such a result far superior than their current practice, and our

model solution has since been implemented.

Table 7.4: Sensitivity analysis on system parameters M and q.

Number of partitions M	Time (s)	Heuristic objective	$X_{\max}$	X_0	$\sum_{m\in\mathcal{M}} C_m$
8	2381	27335	15075	11.984	3134
10	5117	23271	12065	11.984	2607
12	7200	20883	10199	11.984	2346
14	7200	19019	8703	11.984	2162
16	7200	17709	7595	11.984	2061
Probability q	Time (s)	Heuristic objective	$X_{\max}$	X_0	$\sum_{m\in\mathcal{M}} C_m$
0.025	6983	15019	10065	0.047	2465
0.05	6454	15354	10137	0.749	2421
0.1	7200	20883	10199	11.984	2346
0.2	7045	111044	10048	191.747	2561

Finally, we conduct sensitivity analysis on two key system parameters: the number of partitions M and the disruption probability q. The results are shown in Table 7.4. In terms of algorithm performance, as the number of partitions M increases, the size of the problem becomes larger, and a longer computation time (i.e., up to 7200 seconds) is needed. As for each component of the objective, the maximum desk workload $X_{\max}$ and the compactness measure $\sum_{m\in\mathcal{M}} C_m$ both decrease with M, while X_0 remains the same. This is intuitive. When M is larger, the area size and average demand of each district are smaller, implying too that the districts are more compact. However, since the penalty for missing demand is high, the demand of every district will always be assigned to the dummy desk (pay the penalty) at the highest possible level $R + 1 = 4$ with the smallest possible probability $q^4 = 0.0001$. This contributes to the same value of X_0 under different values of M. On the other hand, as the probability value q increases, the computational time does not change much due to the invariant problem size. The values of $X_{\max}$ and $\sum_{m\in\mathcal{M}} C_m$ also almost remain the same under the same number of partitions M. However, the value of X_0 increases significantly because a larger value of q leads to a higher probability for the demand to lose service (and hence, bear the penalty).

CHAPTER 8:

CONCLUSIONS AND FUTURE RESEARCH DIRECTIONS

8.1 Conclusions

Facility location decisions lie at the center of planning many infrastructure systems to serve spatially distributed customers and maximize service efficiency. Recently, devastating infrastructure damages observed in real world show that infrastructure facilities may be subject to disruptions that compromise system performance, which emphasizes the necessity of taking facility disruptions into consideration during the planning. Furthermore, facility systems often exhibit complex interdependence when: (1) facilities are spatially correlated due to physical connections/interrelations, and (2) facilities provide combinatorial service under cooperation, competition or restrictions. These further complicate the facility location design. Therefore, this dissertation aims at extending the reliable facility location models to tackle all these challenges in designing a reliable interdependent facility system. Specifically, several important and challenging extended topics in the reliable facility location context are investigated, including facility correlations, facility combinations, and facility districting.

First, we study the reliable service systems design problem with considerations of possible network access failures. In many faclity systems where customers pass through certain network access points to visit facilities for service, failures of the network access points could potentially affect the functionality of service facilities, and consequently introduce reliability and correlation issues to the system. We add a layer of stations to represent the network access points, and connect them to facilities to indicate their real-world relationships. With the additional layer of stations, the

original fa- cility system is augmented into an integrated facility-station structure, based on which we develop a compact mixed-integer mathematical model to formulate the reliable system design problem. Customized Lagrangian relaxation based algorithm is designed, and multiple case studies are conducted to test the applicability and performance of the proposed model and algorithm. Numerical results provide a range of managerial insights. For example, site-dependent access point failure probabilities generally lead to higher levels of facility concentration so as to provide more back-up options to the customers.

Next, when there exist no physical stations and facilities are exposed to shared hazards, we develop a systematic methodological framework to model generally correlated facility disruptions. We define three probabilistic representations of a correlated disruption profile (i.e., scenario representation, marginal repre- sentation and conditional representation), show their equivalence, and unify them by pairwise transformations. Then we transform these probabilistic disruption profiles into an equivalent supporting station structure consisting of a compact set of additional virtual supporting stations. Each station fails independently, and by proper connections to the original facilities, such an augmented system can capture any correlations among original facilities. Yet, the independent failures of the stations enable us to reduce the computational complexity for system evaluation and optimization. Several favorable properties of the framework were analytically proven, three numerical case studies with different settings (e.g., disaster patterns, correlation types, input disruption profiles) and a series of sensitivity analyses were conducted to illustrate the methodology implementation, and to draw managerial insights.

Based on the virtual supporting station structure, to optimize facility location design, we further show that the reliable facility location problem under correlated facility disruptions can be formulated into a compact mixed-integer mathematical model, which is equivalent to the traditional scenario-based formulation and is much more compact in size. The proposed model can be solved by customized Lagrangian relaxation algorithms (with customized modules for obtaining upper and lower bounds). Multiple case studies with various network settings and correlation patterns were conducted to test the performance and applicability of the methodology. Superiority of the proposed station structure has also been clearly demonstrated via numerical experiments.

We then extend our research focus to the area of sensor deployment, and propose a reliable sensor location model to maximize the accuracy and effectiveness of object positioning and surveillance.

The model allows sensor failures to occur with site-dependent probabilities. We first formulate the reliable sensor deployment problem as a mixed-integer linear program. We then develop a customized Lagrangian relaxation and branch-and-bound algorithm (with approximation subroutine designed for subproblems) to effectively solve the mathematical model. A series of computational experiments with grid networks of varying sizes demon- strate that the proposed algorithm far outperforms CPLEX in terms of solution quality and computation time. In particular, the Lagrangian relaxation and branch-and-bound algorithm is able to solve median-size networks with up to 64 candidate sensor locations and 49 surveillance neighborhoods within 1 hour. A real-world application for the AP network design for Chicago O'Hare Airport Terminal 5 is presented to demonstrate the applicability of the model, the efficiency of the proposed algorithms, and draw practical insights.

Finally, we study a reliable network districting problem, which aims to partition an undirected network into a fixed number of districts and assign their demands to different facilities. Several operational criteria including district contiguity and compactness, facility service reliability, and demand balance are explicitly addressed in a mixed-integer programming model. Customized solution approaches including constructive and neighborhood search heuristics, set-cover based lower bound estimation are developed to efficiently solve the mathematical models to obtain near-optimum solutions. We conduct several numerical examples (including an empirical case study) with different network sizes and parameter settings to demonstrate the applicability and performance of our methodology. Results of the examples show that the proposed heuristic algorithm outperforms commercial solver Gurobi in terms of providing good quality solutions in a shorter amount of computation time. In addition, managerial insights are drawn from the various numerical examples.

As a summary, the main contributions of this PhD research consist of: (1) establishing a new systematic methodological framework based on quasi-probabilities and supporting stations to describe and decompose facility correlations into succinct mathematical representations, which allow compact mathematical formulations to be developed for planning facility locations under correlated facility disruptions; (2) expanding the reliable facility location modeling framework to allow facilities to provide combinatorial service; e.g., in the context of sensor deployment problems, sensors work in combinations to provide positioning/surveillance service via trilateration procedure; and (3) incorporating reliability concepts into the spatial districting context (e.g., for political, school, service systems), where spatial contiguity, compactness, and demand balance must be ensured, and

developing various types of new customized model formulations and solution approaches. For all these studies, we develop specific mathematical derivations, modeling techniques, optimization formulations, and customized exact/heuristic algorithms. Numerous hypothetical numerical studies are presented to illustrate the performance of our methodologies, and a series of empirical/industrial applications are also conducted to demonstrate the applicability of our research, and to draw managerial insights. The findings and outcomes in the dissertation will serve as the basis for reliable facility location planning in many engineering contexts.

8.2 Future Directions

Further research can be conducted in the following directions.

First, there are many other ways to express correlations among facility disruptions. Sometimes the correlations follow explicit physical laws. For example, under natural disasters, the correlation between any two candidate locations could be specified by a decaying probability of failure "contagion" (e.g., $e^{distance}$) that depends on their relative distance, or be expressed by correlation coefficient between two Bernoulli random variables that indicate whether the two locations are disrupted. Specific model structure and insights might be available at those correlation patterns, and deserve further investigation.

Second, as we intensively adopt the concept of independent supporting stations to capture the effects of facility correlations, it will be interesting as well as important to consider correlated disruptions of the supporting stations. This is because that in many real world contexts (e.g., correlated bridge failures or roadway blockages due to shared hazards), the supporting stations are actually not correlated with each other. Similarly, in chapter 6, sensors that are assumed to be functioning independently, are also possible to be correlated. Therefore, it is important to extend the methodologies developed in this dissertation to incorporate these additional types of correlations.

Additionally, in this dissertation, all studies focus on problems with discrete settings and formulate discrete mathematical models. However, similar problems are frequently formulated in the continuous metric space where system parameters are described by continuous functions. For example, in chapter 6, the distributed demand and sensor installation cost might be described by continuous density functions; similarly, in chapter 7, discrete network could be replaced with a con-

tinuous plane, which calls for general partitioning rules with the consideration of reliability issues in continuous settings. To this end, we leave these challenging possibilities for future research.

Finally, it will also be interesting to apply our methodology to more real-world cases, so as to help policy makers develop engineering and planning guidelines that will lead to more reliable and resilient systems. For example, studies in chapter 3 could provide further guidances on the positioning and utilizing of emergency response resources in many practical contexts; while models and algorithms in chapter 6 can be applied to similar resource allocation and positioning application contexts such as the seismic sensor network configuration for earthquake epicenter calculations, and complex structure health monitoring for bridges and tunnels.

REFERENCES

Ahmed, A., Watling, D., Ngoduy, D., 2014. Significance of sensor location in real-time traffic state estimation. Procedia Engineering 77, 114–122.

An, K., Xie, S., Ouyang, Y., 2017. Reliable sensor location for object positioning and surveillance via trilateration. Transportation Research Part B: Methodological , In press.

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

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

Berman, O., Krass, D., 2011. On n-facility median problem with facilities subject to failure facing uniform demand. Discrete Applied Mathematics 159, 420–432.

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, 332–350.

Berman, O., Krass, D., Menezes, M.B.C., 2009. Locating facilities in the presence of disruptions and incomplete information. Decision Sciences 40, 845–868.

Berman, O., Krass, D., Menezes, M.B.C., 2013. Location and reliability problems on a line: Impact of objectives and correlated failures on optimal location patterns. Omega 41, 766–779.

Blais, M., Lapierre, S.D., Laporte, G., 2003. Solving a home-care districting problem in an urban setting. The Journal of the Operational Research Society 54, 1141–1147.

Bozkaya, B., Erkut, E., Laporte, G., 2003. A tabu search heuristic and adaptive memory procedure for political districting. European Journal of Operational Research 144, 12–26.

Brualdi, R.A., 2004. Introductory Combinatorics, 4/e. Pearson Education India.

Camacho-Collados, M., Liberatore, F., Angulo, J., 2015. A multi-criteria police districting problem for the efficient and effective design of patrol sector. European Journal of Operational Research 246, 674–684.

CDA, 2016. Chicago department of aviation. http://www.flychicago.com/ohare/en/home/Pages/default.aspx. Accessed: 2016-08-15.

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

Church, R.L., Scaparra, M.P., Middleton, R.S., 2004. Identifying critical infrastructure: The median and covering facility interdiction problems. Annals of the Association of American Geographers 94, 491–502.

Clouqueur, T., Phipatanasuphorn, V., Ramanathan, P., Saluja, K.K., 2003. Sensor deployment strategy for detection of targets traversing a region. Mobile Networks and Applications 8, 453–461.

Crummy, B., 2013. Mile-long train carrying crude oil derails, explodes in north dakota. http://usnews.nbcnews.com/_news/2013/12/30/22113442-mile-long-train-carrying-crude-oil-derails-explodes-in-north-dakota?lite. Accessed: 2014-03-17.

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

Danczyk, A., Di, X., Liu, H.X., 2016. A probabilistic optimization model for allocating freeway sensors. Transportation Research Part C: Emerging Technologies 67, 378–398.

Daskin, M.S., 2011. Network and Discrete Location: Models, Algorithms, and Applications. 2nd ed., John Wiley & Sons, New York.

De Stefano, M., Gherlone, M., Mattone, M., Di Sciuva, M., Worden, K., 2015. Optimum sensor placement for impact location using trilateration. Strain 51, 89–100.

Dirac, P.A.M., 1942. Bakerian lecture. the physical interpretation of quantum mechanics. Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences 180, 1–40.

Doukhnitch, E., Salamah, M., Ozen, E., 2008. An efficient approach for trilateration in 3d positioning. Computer Communications 31, 4124–4129.

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

Eisenman, S., Fei, X., Zhou, X., Mahmassani, H., 2006. Number and location of sensors for real-time network traffic estimation and prediction: sensitivity analysis. Transportation Research Record: Journal of the Transportation Research Board 1964, 253–259.

Erdemir, E.T., Batta, R., Spielman, S., Rogerson, P.A., Blatt, A., Flanigan, M., 2008. Location coverage models with demand originating from nodes and paths: Application to cellular network design. European Journal of Operational Research 190, 610–632.

Fei, X., Mahmassani, H.S., 2011. Structural analysis of near-optimal sensor locations for a stochastic large-scale network. Transportation Research Part C: Emerging Technologies 19, 440–453.

Ferland, J.A., Guénette, G., 1990. Decision support system for the school districting problem. Operations Research 38, 15–21.

Feynman, R.P., 1987. Negative probability, in: Hiley, B.J., Peat, D. (Eds.), Quantum Implications: Essays in Honour of David Bohm. Methuen, pp. 235–248.

Filonenko, V., Cullen, C., Carswell, J.D., 2013. Indoor positioning for smartphones using asynchronous ultrasound trilateration. ISPRS International Journal of Geo-Information 2, 598–620.

Fisher, M.L., 2004. The lagrangian relaxation method for solving integer programming problems. Management Science 50, 1861–1871.

Garfinkel, R.S., Nemhauser, G.L., 1970. Optimal political districting by implicit enumeration techniques. Management Science 16, 495–508.

Geetla, T., Batta, R., Blatt, A., Flanigan, M., Majka, K., 2014. Optimal placement of omnidirectional sensors in a transportation network for effective emergency response and crash characterization. Transportation Research Part C: Emerging Technologies 45, 64–82.

Gentili, M., Mirchandani, P., 2012. Locating sensors on traffic networks: Models, challenges and research opportunities. Transportation Research Part C: Emerging Technologies 24, 227–255.

Gold, R., Stevens, L., 2014. U.s. issues emergency testing order to crude oil rail shippers: Move by transportation department response to crude-by-rail accidents. http://on.wsj.com/1WQaKvt. Accessed: 2014-03-17.

Griffiths, D., 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, 637–648.

Gueye, S., Menezes, M.B.C., 2015. General asymptotic and submodular results for the median problem with unreliable facilities. Operations Research Letters 43, 519–521.

He, S., 2013. A graphical approach to identify sensor locations for link flow inference. Transportation Research Part B: Methodological 51, 65–76.

Hess, S.W., Samuels, S.A., 1971. Experiences with a sales districting model: criteria and implementation. Management Science 18, 41–54.

Hess, S.W., Weaver, J.B., Siegfeldt, H.J., Whelan, J.N., Zitlau, P.A., 1965. Nonpartisan political redistricting by computer. Operations Research 13, 998–1006.

Huang, R., Kim, S., Menezes, M.B.C., 2010. Facility location for large-scale emergencies. Annals of Operations Research 181, 271–286.

Krumm, J., Harris, S., Meyers, B., Brumitt, B., Hale, M., Shafer, S., 2000. Multi-camera multi-person tracking for easyliving, in: Proceedings of the Third IEEE International Workshop on Visual Surveillance (VS'2000), pp. 3–10.

Li, X., Ouyang, Y., 2010. A continuum approximation approach to reliable facility location design under correlated probabilistic disruptions. Transportation Research Part B: Methodological 44, 535–548.

Li, X., Ouyang, Y., 2011. Reliable sensor deployment for network traffic surveillance. Transportation Research Part B: Methodological 45, 218–231.

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

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

Liberatore, F., Scaparra, M.P., Daskin, M.S., 2011. Analysis of facility protection strategies against an uncertain number of attacks: The stochastic r-interdiction median problem with fortification. Computers & Operations Research 38, 357–366.

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

Liu, C., Fan, Y., Ordóñez, F., 2009. A two-stage stochastic programming model for transportation network protection. Computers & Operations Research 36, 1582–1590.

Lu, M., Ran, L., Shen, Z.J.M., 2015. Reliable facility location design under uncertain correlated disruptions. Manufacturing & Service Operations Management 17, 445–455.

Mehrotra, A., Johnson, E.L., Nemhauser, G.L., 1998. An optimization based heuristic for political districting. Management Science 44, 1100–1114.

Mirchandani, P.B., Li, J.Q., Long, Y., 2010. Locating a surveillance infrastructure in and near ports or on other planar surfaces to monitor flows. Computer-Aided Civil and Infrastructure Engineering 25, 89–100.

Muyldermans, L., Cattrysse, D., Oudheusden, D.V., Lotan, T., 2002. Districting for salt spreading operations. European Journal of Operational Research 139, 521–532.

Navidi, W., Murphy, W.S., Hereman, W., 1998. Statistical methods in surveying by trilateration. Computational Statistics & Data Analysis 27, 209–227.

NBC News, 2013. Danger on the tracks: Unsafe rail cars carry oil through us towns. http://www.nbcnews.com/news/other/danger-tracks-unsafe-rail-cars-carry-oil-through-us-towns-f8C11082948. Accessed: 2014-03-17.

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, 309–319.

Peng, F., Li, X., Ouyang, Y., 2011. Installation of railroad wayside defect detectors. Transportation Research Record: Journal of the Transportation Research Board 2261, 148–154.

Roa, J.O., Jiménez, A.R., Seco, F., Prieto, J.C., Ealo, J., 2007. Optimal placement of sensors for trilateration: Regular lattices vs meta-heuristic solutions. Springer Berlin Heidelberg, Berlin, Heidelberg. pp. 780–787.

Russell, S., Norvig, P., 2009. Artificial Intelligence: A Modern Approach. 3rd ed., Prentice Hall Press, Upper Saddle River, NJ.

Scaparra, M.P., Church, R.L., 2008a. A bilevel mixed-integer program for critical infrastructure protection planning. Computers & Operations Research 35, 1905–1923.

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

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

Sherman, J., Morrison, W.J., 1950. Adjustment of an inverse matrix corresponding to a change in one element of a given matrix. The Annals of Mathematical Statistics 21, 124–127.

Snyder, L.V., 2006. Facility location under uncertainty: a review. IIE Transactions 38, 547–564.

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

Thomas, F., Ros, L., 2005. Revisiting trilateration for robot localization. IEEE Transactions on Robotics 21, 93–101.

Wang, Y.C., Hu, C.C., Tseng, Y.C., 2005. Efficient deployment algorithms for ensuring coverage and connectivity of wireless sensor networks, in: First International Conference on Wireless Internet (WICON'05), pp. 114–121.

Xie, S., An, K., Ouyang, Y., 2018a. Planning facility location under generally correlated facility disruptions: Use of supporting stations and quasi-probabilities. Transportation Research Part B: Methodological , Revisions under review.

Xie, S., An, K., Ouyang, Y., 2018b. Positioning, Planning and Operation of Emergency Response Resources and Coordination between Jurisdictions. Technical Report. Center for Transportation Studies, University of Minnesota.

Xie, S., Li, X., Ouyang, Y., 2015. Decomposition of general facility disruption correlations via augmentation of virtual supporting stations. Transportation Research Part B: Methodological 80, 64–81.

Xie, S., Ouyang, Y., 2016. Railroad caller districting with reliability, contiguity, balance, and compactness considerations. Transportation Research Part C: Emerging Technologies 73, 65 – 76.

Xie, S., Ouyang, Y., 2018. Reliable facility location design under the risk of network access failures. Transportation Research Part E: Logistics and Transportation Review , Under review.

Xie, X., Dai, J.B., 2014. Sensor placement in the ultrasonic positioning system. International Journal of Management Science and Engineering Research 1, 31–39.

Zou, Y., Chakrabarty, K., 2004. Sensor deployment and target localization in distributed sensor networks. ACM Transactions on Embedded Computing Systems 3, 61–91.