

8-2012

Locating and Protecting Facilities Subject to Random Disruptions and Attacks

Hugh Medal

University of Arkansas, Fayetteville

Follow this and additional works at: <http://scholarworks.uark.edu/etd>

 Part of the [Industrial Engineering Commons](#), [Information Security Commons](#), and the [Operational Research Commons](#)

Recommended Citation

Medal, Hugh, "Locating and Protecting Facilities Subject to Random Disruptions and Attacks" (2012). *Theses and Dissertations*. 448.
<http://scholarworks.uark.edu/etd/448>

This Dissertation is brought to you for free and open access by ScholarWorks@UARK. It has been accepted for inclusion in Theses and Dissertations by an authorized administrator of ScholarWorks@UARK. For more information, please contact scholar@uark.edu, ccmiddle@uark.edu.

LOCATING AND PROTECTING FACILITIES
SUBJECT TO RANDOM DISRUPTIONS AND ATTACKS

LOCATING AND PROTECTING FACILITIES
SUBJECT TO RANDOM DISRUPTIONS AND ATTACKS

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy in Industrial Engineering

By

Hugh Richard Medal
North Dakota State University
Bachelor of Science in Industrial Engineering and Management, 2006
University of Arkansas
Master of Science in Industrial Engineering, 2008

August 2012
University of Arkansas

Abstract

Recent events such as the 2011 Tohōku earthquake and tsunami in Japan have revealed the vulnerability of networks such as supply chains to disruptive events. In particular, it has become apparent that the failure of a few elements of an infrastructure system can cause a system-wide disruption. Thus, it is important to learn more about which elements of infrastructure systems are most critical and how to protect an infrastructure system from the effects of a disruption. This dissertation seeks to enhance the understanding of how to design and protect networked infrastructure systems from disruptions by developing new mathematical models and solution techniques and using them to help decision-makers by discovering new decision-making insights.

Several gaps exist in the body of knowledge concerning how to design and protect networks that are subject to disruptions. First, there is a lack of insights on how to make equitable decisions related to designing networks subject to disruptions. This is important in public-sector decision-making where it is important to generate solutions that are equitable across multiple stakeholders. Second, there is a lack of models that integrate system design and system protection decisions. These models are needed so that we can understand the benefit of integrating design and protection decisions. Finally, most of the literature makes several key assumptions: 1) protection of infrastructure elements is perfect, 2) an element is either fully protected or fully unprotected, and 3) after a disruption facilities are either completely operational or completely failed. While these may be reasonable assumptions in some contexts, there may exist contexts in which these assumptions are limiting. There are several difficulties with filling these gaps in the literature. This dissertation describes the discovery of mathematical formulations needed to fill these gaps as well as the identification of appropriate solution strategies.

This dissertation is approved for
recommendation to the
Graduate Council

Dissertation Director:

Edward A. Pohl, Ph.D.

Dissertation Committee:

Manuel D. Rossetti, Ph.D. (Co-Director)

Scott J. Mason, Ph.D.

Russell D. Meller, Ph.D.

Chase E. Rainwater, Ph.D.

©2012 by Hugh R. Medal
All Rights Reserved

Dissertation Duplication Release

I hereby authorize the University of Arkansas Libraries to duplicate this thesis when needed for research and/or scholarship.

Agreed _____
Hugh Richard Medal

Refused _____
Hugh Richard Medal

Acknowledgments

Although this dissertation lists only one author, there are many people who helped me significantly during the process of completing this work.

First, I would like to thank my committee co-chairs, Ed Pohl and Manuel Rossetti. Both of them were very patient and consistently helpful. I never doubted that they sincerely wanted me to succeed. I am grateful to Dr. Pohl for funding my dissertation research through a Department of Homeland Security grant. I also want to acknowledge the outstanding care he has for his students. He often asks how we are doing and even stops by to visit us before he leaves for the day. Dr. Rossetti also truly has his students' best interests in mind. He was consistently available to talk about my research and I benefited from many of our conversations.

Additionally, I am indebted to the remainder of my dissertation committee: Scott Mason, Russell Meller, and Chase Rainwater. Dr. Mason helped by encouraging me to convey my research in a way that real decision-makers can understand. He also encouraged me to improve my ability to move projects to completion. Dr. Meller taught me several skills that were useful in writing this dissertation. First, he taught me how to systematically read a paper, a skill that was useful in my research. Second, he helped me become a better writer by promoting excellence in writing and also teaching me how to structure a research paper. Finally, he showed me that industrial engineering research is mainly about answering a question, rather than developing a model or algorithm. Dr. Rainwater provided significant guidance on Chapter 4 in addition to helping me with other modeling and algorithmic issues. I was very impressed by his willingness to help me without receiving anything in return.

Several other staff at the University of Arkansas provided significant help. Bob Haslam, director of the Quality Writing Center, helped me improve my writing, specifically helping me improve the presentation of Chapters 4 and 5 of this dissertation. Further, the staff at the Arkansas High Performance Computing Center answered many questions I had about running my code on their servers; David Chaffin and Jeff Pummill were especially helpful.

I am also thankful for many friends in the Department of Industrial Engineering that have

assisted me throughout this process. Especially notable are Jen Pazour, Ridvan Gedik, and Steve Sharp. I would especially like to thank Vijith Varghese for his unwavering friendship in Christ during the past six years. Specifically, his support of my research was encouraging.

I also want to thank my wife, Leanna. She is a model of the wife described in Proverbs 31; far more precious than jewels (Proverbs 31:10)¹. She was very patient during this whole process and was supportive by being excited when I met milestones, listening to my ideas, and encouraging me to work hard. In fact, she even proofread much of this dissertation.

Finally, and most importantly, I want to thank God, my Father. He gave me life (Psalm 139:13) and meets my daily needs (Matthew 6:33). He provides me with strength (Isaiah 40:29) and peace of mind (John 13:27). And he also provided me with eternal life in his Son Jesus Christ (1 John 5:11)! In summary, God provided me with everything I needed to complete this dissertation.

¹All scripture references are from The Holy Bible, English Standard Version Copyright ©2001 by Crossway Bibles, a division of Good News Publishers.

Contents

1	Introduction	1
2	Literature Review	7
2.1	Introduction	7
2.2	Definitions, Classification Scheme, and Scope	10
2.2.1	Definitions	10
2.2.2	Classification Scheme	14
2.2.3	Scope and Related Work	17
2.3	Descriptive Models	18
2.4	General Modeling Techniques	19
2.5	Design Models	21
2.5.1	Facility Networks: Facility Failures	23
2.5.1.1	Expected Value Risk Measure	24
2.5.1.2	Worst Case Risk Measure	28
2.5.2	Facility Networks: Arc Failures	30
2.5.3	Complex Networks	31
2.5.3.1	Expected Value Risk Measure	32
2.5.3.2	Worst Case Risk Measure	32
2.5.3.3	Survivability Risk Measure	32
2.5.3.4	Robustness Risk Measure	33
2.5.3.5	Risk Metric	33
2.5.4	Future Work	34
2.6	Risk Reduction Models	35
2.6.1	Set of Elements	41
2.6.1.1	Conditional Expected Value Risk Measure	41
2.6.1.2	Defending Against Random Incidents and Strategic Attacks	43
2.6.2	Simple Networks	44
2.6.3	Facility Networks: Facilities Fail	45
2.6.3.1	Expected Value Risk Measure	45
2.6.3.2	Worst Case Risk Measure	46
2.6.4	Complex Networks	49
2.6.4.1	Expected Value Risk Measure	49
2.6.4.2	Conditional Expected Value Risk Measure	50
2.6.4.3	Worst Case Risk Measure	50
2.6.4.4	Survivability Risk Measure	52
2.6.4.5	Robustness Risk Measure	52
2.6.4.6	Risk Metric	53
2.6.4.7	Multiple Risk Measures	53
2.6.5	Future Work	53
2.7	Conclusions	54
2.7.1	Future Work: Imbalances	56
2.7.2	Future Work: Big Picture	57

3	Overview	60
3.1	Gaps in the Literature	60
3.2	Modeling Assumptions	63
3.3	Contributions	65
4	On the R-All-Neighbor P-Center Problem	68
4.1	Introduction	68
4.2	Literature Review	69
4.3	Problem Description and Model	72
4.3.1	Three Index Formulation	73
4.3.2	Reformulation of the RANPCP	75
4.4	Binary Search Algorithm	78
4.4.1	Bounds	80
4.4.2	Greedy Heuristic	81
4.5	Computational Experimentation	81
4.5.1	Experimental Design	82
4.5.2	Comparison of MIP Models	82
4.5.3	MIP Models vs. Binary Search	83
4.5.4	Scalability of Binary Search Algorithm	85
4.6	Tradeoff Analysis	86
4.6.1	Penalties for Only Considering a Single Objective	86
4.6.2	Multi-objective Analysis: Max Closest Distance Vs. Max r^{th} Closest Distance	87
4.6.3	Multi-objective Analysis: Cost Vs. Maximum r^{th} Closest Distance	90
4.6.4	Saturation Point	90
4.7	Example with Managerial Insights	92
4.8	Conclusions and Future Work	95
5	Integrated Facility Location and Hardening	98
5.1	Introduction	98
5.2	Literature Review	100
5.3	Problem Description and Models	103
5.3.1	Describing the Most Disruptive Facility Disruption Scenario	105
5.3.2	Single-Level Model	106
5.3.3	MIP Model	107
5.4	Solution Methodology	108
5.4.1	Auxiliary Problem	109
5.4.2	Binary Search Algorithm	110
5.5	Special Cases	111
5.5.1	Location and Hardening	111
5.5.2	Non-Integrated Methods	111
5.5.2.1	Location Then Hardening (LTH)	112
5.5.2.2	Location Then Hardening: Ideal Proportion (LTHI)	112
5.6	Experimental Results	113
5.6.1	Benefit of Integration	114

5.6.2	Benefit of Hardening	115
5.6.3	Relative Objective Function Increases for Only Considering a Single Objective	117
5.6.4	Computational Experimentation	119
5.6.4.1	Integrated Method vs. Sequential Method	120
5.6.4.2	Computational Performance of Binary Search Algorithm	121
5.7	Managerial Insights	123
5.7.1	Relative Objective Function Increases for Only Considering a Single Objective	124
5.7.2	Budget Vs. Post-Disruption Radius	125
5.8	Summary and Future Work	126
5.8.1	Summary	126
5.8.2	Future Work	128
6	Imperfect Protection of Multi-State Supply Facilities	129
6.1	Introduction	129
6.2	Notation and Problem Description	134
6.2.1	Preliminary Notation	134
6.2.2	Problem Description	136
6.3	Model	138
6.3.1	Additional Notation	138
6.3.2	Model	138
6.4	Solution Method	142
6.4.1	L-Shaped Method	142
6.4.2	Local Search Heuristic	144
6.5	Experimentation	146
6.5.1	Comparison with Other Models	147
6.5.2	Sensitivity Analysis	150
6.5.2.1	Budget	150
6.5.2.2	Number of Allocation Levels and Capacity States	152
6.5.3	Computational Performance	154
6.5.3.1	Experimental Set-Up	154
6.5.3.2	Local Search: Solution Quality	154
6.5.3.3	L-Shaped vs. Local Search	155
6.5.3.4	Local Search: Scalability	156
6.6	Conclusion	157
6.6.1	Future Work	158
7	Conclusions	160
7.1	Summary	160
7.2	Major Insights Gained	161
7.3	Recommendations for Future Work	163
7.3.1	Increasing Model Fidelity	163
7.3.1.1	All-Hazards Protection	163
7.3.1.2	Multiple Time Periods	164

7.3.1.3	Sequential Decision-Making	165
7.3.1.4	Cascading Failures	165
7.3.1.5	Interdependent Infrastructures	165
7.3.2	New Applications	166
7.3.2.1	Interdiction of Infectious Disease Outbreaks	166
7.3.2.2	Interdiction of Crime Hot-Spots	167
A	Examples of Network Disruptions	168
B	Extra Material for Chapter 4	171
B.1	Datasets Used in Experimentation	171
B.2	Effect of Weights on Binary Search Algorithm	172
B.3	Saturation	173
C	Extra Material for Chapter 5	177
C.1	Proofs	177
C.2	Other Models Tested	178
C.3	Comparison of Auxiliary Problems for Integrated MFHLP Binary Search	179
C.4	Binary Search Algorithm	179
C.5	Greedy Heuristic for Auxiliary Sub Problem	180
C.6	Hardening Model	182
C.7	Datasets Used	183
C.8	Comparison of sequential and integrated methods	185
	Bibliography	186

1 Introduction

On March 11, 2011, an earthquake occurred near Japan, triggering a tsunami which caused mass destruction on Japan's northeastern coast. This destruction included the loss of many lives, the damage of a large amount of property, and a near nuclear disaster. In addition, this destruction affected the supply chains of many Japanese-based manufacturers. The earthquake damaged sensitive equipment at Renesas Electronics Corporation, which was making microchip controllers at just one factory north of Tokyo, causing them to halt production. Because several Japanese automakers used Renesas as their only source for microchip controllers, automakers were forced to halt car production for up to six months (Kim, 2012).

In response to this disruption, the automaker Toyota vowed to be better prepared next time. In particular, they worked to reduce the vulnerability of their suppliers by asking them to either spread production or hold extra stock. "Our plan is to manage risk while at the same time reducing costs," said vice president Shinichi Sasaki (Kim, 2012).

In this dissertation, OR methodologies are used to help organizations such as Toyota "manage risk while at the same time reducing costs." Specifically, models are developed for designing and protecting networks that are subject to disruptions. In military and supply chain contexts, people have referred to three stages of decision-making: operational (short-term), tactical (medium-term), and strategic (long-term). This dissertation focuses on the strategic level. The strategic level includes decisions made before a disruption occurs such as (1) where to pre-position relief supplies, (2) how to design systems so that they are less vulnerable to disruptions, (3) how to conduct capacity planning in view of a potential disruption, and (4) how to plan for workforce training to build manpower capacity in preparation for a disruption.

This dissertation focuses on strategic decisions because the impact of a disruption can be significantly reduced with sound decisions at the strategic level. Indeed, the anecdote, "an ounce of prevention is worth a pound of cure," applies to network disruptions. Studies have shown that a \$1 investment in preparedness can reduce the cost of responding to a disruption by \$7 (Healy and

Malhotra, 2009).

There are many categories of infrastructure in networks such as supply chains and transportation networks: roads, interchanges, supply facilities, transportation hubs, distribution centers, etc. This dissertation focuses on one category of infrastructure: supply facilities. Supply facilities were chosen because they are critical to supply chains. For example, if a road or interchange is destroyed, there are often other available routes. However, as the tsunami in Japan illustrated, when a supply facility is disrupted companies are often left without other options for procuring supplies, as was the case with Toyota.

The decision of locating supply facilities is often faced in the government sector and the private sector. For example, city governments face the problem of where to locate schools, hospitals, police stations, and fire stations. In the private sector, retailers decide where to locate retail stores, airlines decide where to locate hub airports, and cell phone providers decide where to locate broadcast towers. Facility location decisions are often aided by mathematical models, which prescribe the best locations for facilities. However, the classic facility location models optimize long-run performance of the facility location configuration and assume that facilities are always available for service. This may be a realistic assumption in some cases, in some situations. However, in some contexts, facility unavailability is so prevalent or causes such a large disruption that they should not be ignored in facility location models. In this section we present some recent findings on the modeling of facilities subject to failure.

The location and protection of facilities is important for the mission of the Office of Critical Infrastructure protection of the Department of Homeland Security (DHS). DHS has listed 18 sectors that are considered critical infrastructure sectors. While all of these sectors utilize facilities, we have identified several of these sectors in which the operation of facilities is critical: Energy, Information Technology, Postal and Shipping, Communications, Transportation Systems, Emergency Services, and Water. Table 1.1 lists examples of critical facilities for each of the 7 sectors that we identified. Facilities are also important for commerce. Intermodal terminals, warehouses, factories, and retail stores are critical to domestic commerce. Global commerce is dependent on ports

Table 1.1: Sectors with critical facilities

Sector	Critical facilities
Energy	power plants and sub station
Information Technology	internet switching stations
Postal and Shipping	processing facilities
Communications	cellular towers
Transportation Systems	airports, ports, intermodal terminals
Emergency Services	hospitals, fire stations
Water	waste water treatment facilities

in United States and throughout the world. Facilities are also important for disaster response. For example, the Strategic National Stockpile includes strategically located facilities throughout the United States stocked with pharmaceutical supplies that may be needed to respond to a bio-terror attack on a large US city.

The classic facility location model includes a set of demand points, which require service from a facility. A budget exists to locate facilities that will provide service to all of the demand points. After the facilities are located, each demand point is assigned to the facility that is closest to it. Thus, facilities are located in order to optimize some service measure. In the p -center model, the service measure is the maximum distance from any demand point to its closest located facility.

In this example the p -center model is used to choose the locations of the facilities. Thus, the objective is to minimize the maximum distance from any demand point to its closest located facility. The objective of minimizing the maximum distance is often used when locating public-sector facilities such as hospitals and fire stations in order to provide equitable service to all of the demand points.

The solution to the p -center model is shown in Figure 1.1a. A set of demand points representing the capitals of the lower-48 states and the District of Columbia are shown in black dots. This is the classic USCities dataset (Daskin, 1995), which is based on data from the 1990 US census. Facilities can be located at any of the demand point locations. The population of the state capital is used as a proxy for the amount of demand required by that point. The cost of living in the state capital is used

as a proxy for the cost of locating a facility at that point. The distance between each pair of points is the Great Circle distance. In the solution, facilities are located at Harrisburg, Lansing, Oklahoma City, Jackson, Salt Lake City, Augusta, Boise City, and Cheyenne. The maximum distance from a demand point to its assigned facility is 294, which is the distance from Carson City to the facility located at Boise City.

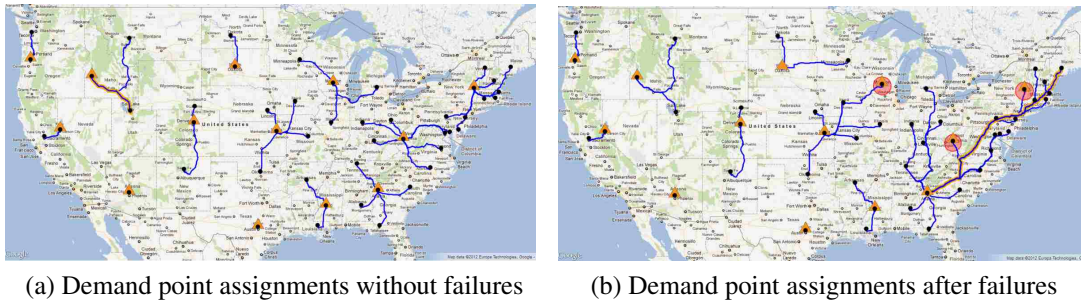


Figure 1.1: Facilities located to optimize performance without facility failures

The solution to the p -center model, shown in Figure 1.1a, is based on the assumption that facilities are always available. However, facilities sometimes become unavailable due to natural disasters, terrorist attacks, labor strikes, or man-made accidents. In our modeling of facility failures, we assume that after facilities fail, demand points are assigned to their closest facility that is still operating. We also assume that a facility is always in one of two states: available or unavailable.

Figure 1.1b displays the failure of the 3 facilities whose failure causes the greatest impact. Note that because these three facilities have failed, demand points must be reassigned to their closest located facility. The radius is now 1071. We call the maximum distance after facility failures the *post-disruption radius*. Conversely, we call the maximum distance when no facilities have failed the *non-disruption radius*.

Figure 1.2 shows the solution generated by a model that optimizes the post-disruption radius that is introduced in Chapter 4. This model, called the r -all-neighbor p -center (RANPCP) model,

prescribes the location of one more facility than the p -center model. In the RANPCP solution, eight facilities are located at least as far east as Topeka, KS, compared to six in the p -center solution. These facilities make the network less vulnerable to the failures in the northeast shown in Figure 1.1b.

Figure 1.2a shows the RANPCP solution without failures. Since the RANPCP optimizes the post-failure radius, the non-failure radius increases from 294 (the radius for the p -center solution) to 439, a 49% increase. Because the RANPCP model located more facilities in the east, the radius is now in the west.

Figure 1.2b shows the RANPCP solution after failures. Since the RANPCP optimizes the post-failure radius, the post-failure radius decreases from 1071, the post-failure radius for the p -center solution, to 718 (the distance from Tallahassee to the facility located at Jefferson City), a 33% decrease. Because the RANPCP model located more facilities in the east, the impact of a disruption in the east is not as great.

This example shows that locating facilities to optimize post-failure performance can reduce the vulnerability of the network. However, this vulnerability reduction can cause the non-failure performance to become worse. Thus, a tradeoff exists between performance without failures and performance after failures. This dissertation will present methods to optimize the performance after failures and will analyze the tradeoff between performance without failures and performance after failures.

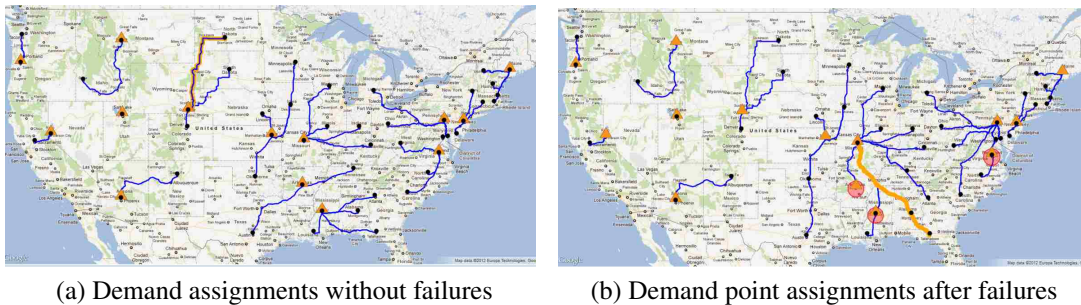


Figure 1.2: Facilities located to optimize post-failure radius

The remainder of this dissertation is as follows. Chapter 2 is a comprehensive review of the literature on designing and protecting networks subject to disruptions. Chapter 3 gives a detailed discussion of the gaps in the literature that are addressed in this dissertation and outlines the remainder of the dissertation, describing its contributions. Chapters 4–6 are the core contribution of this dissertation. Finally, Chapter 7 summarizes the contents of this dissertation, draws conclusions, and recommends several areas of future research.

2 Literature Review¹

Abstract

Networked infrastructures are everywhere in the world, such as in transportation, communication, and water distribution. Because of increased global competition, these networks are not only often geographically dispersed and fast evolving, but also must be relatively cheap to build and operate. These characteristics make some networks fragile and vulnerable to disruptions. This is a problem because many of these networks are crucial to the function of regional and global economies. A body of research has emerged that addresses this problem; it seeks to design more efficient networks and reduce the disruption risk of existing networks. Partly due to the occurrence of disruptive events, such as the 2001 bombing of the World Trade Center towers and Hurricane Katrina in 2005, this body of research has grown substantially in the last decade or so. In this survey, we review articles from this body of literature, place them into categories, and suggest topics for future research. A goal of this paper is to help researchers see the different approaches taken by other researchers from a variety of disciplines (such as physics, engineering, economics).

2.1 Introduction

The term *network* is defined as a collection of entities associated with each other through physical and/or virtual relationships/connections. Networks may be physical, such as those existing in transportation, the worldwide Web, wired and wireless communication, and electrical power, as well as water, oil, and gas distribution. Networks may also be virtual or relationship-based such as social networks, biological networks, and the partnerships that exist in supply chains. As our world becomes more interconnected, networks are becoming more geographically distributed, as evidenced in supply chains and communication systems. Additionally, these networks are becoming more and more critical: most of the world relies heavily on networked systems such as transportation, electrical power, telecommunication, and the Internet.

Not surprisingly, when these network infrastructure elements are disrupted, serious conse-

¹This chapter is based on an article written by the author of this dissertation (Medal et al., 2011b).

quences often occur. For example, the 2004-2005 disruption to the rail network in the Powder River Basin of Wyoming, which was due to rail line and engine failures, resulted in a shortage of coal and electricity price increases of up to 15% for certain regions of the US (News and Information, 2005). Another, more recent, example is the 2010 eruptions of the volcano Eyjafjallajökull in Iceland, which disrupted air travel throughout northern and western Europe for several days (Bolić and Sivčev, 2011). Compounding the problem is that these networks are often interdependent. Thus, when one network is disrupted, consequences are realized in others. These are called *cascading failures*. One of the most well-known examples of a cascading failure is the blackout that occurred in the Northeastern US in 2003. The blackout was caused by the failure of power lines due to contact with trees, exacerbated by a software bug in the energy management system, and ultimately impacting many other important networks such as water distribution, transportation, wireless communication, and the Internet.

Why are these networks vulnerable to such huge disruptions? We suggest four reasons. First, the goal in designing networks is most often efficiency and cost minimization. Evidence of this is the fact that the classic network design and facility location models (see Daskin (1995) for examples) most often have cost minimization objectives. Typically, this means designing networks to be used at or near their maximum capacity, making them inherently vulnerable to disruptions. Second, these networks are often not designed by a single decision maker. Rather, they are frequently designed by many decentralized decision makers who each have their own objectives and perception of risk. In this chapter we define risk as function of both the likelihood and severity of disruptions to a network. Third, rather than being designed at a single point in time, many physical networks gradually evolve over time, often in reaction to changes in demand. Again, this approach often results in a network that is sub-optimal in terms of vulnerability. An interesting example of this is found in the studies by Barabasi and Albert (1999) and Albert et al. (2000). Barabasi and Albert (1999) show that many real world networks can be described by a particular growth model that involves an increase in the number of nodes over time, and a particular type of attachment called preferential attachment, where new nodes are more likely to attach to existing nodes that

are highly connected. Albert et al. (2000) demonstrate that these networks usually maintain connectivity in the presence of random single-element failures but are vulnerable to strategic attacks. Finally, even if risk is considered in the design of a network, often only the risks present at the time of the design are considered. This is a problem because risks change over time. Thus, a network that can function well in the presence of risk at the time it was designed may not be able to do so many years into the future.

In light of these vulnerabilities, there is a growing body of literature studying networks subject to disruptions. In this chapter, we review the literature that deals with making networks efficient and able to perform well in the presence of disruptions. That is, we discuss the science of designing new networks as well as protecting and modifying existing networks while considering both efficiency and risk. There are a few other notable surveys related to this chapter. Snyder et al. (2010) survey the literature relating to disruptions in supply chains, covering a broader array of topics than just networks. Brown et al. (2005a) provide a tutorial on defending networks against attackers. Snyder et al. (2006) provide a survey and tutorial on disruptions to supply networks, covering both design and hardening models. We consider the survey presented in this chapter to be a complement to the paper by Snyder et al. (2006) because it discusses additional types of networks, and presents additional strategies to reduce the risk associated with an existing network besides hardening, such as redundancy and secrecy. It should be noted that most of the models presented in Snyder et al. (2006) use operations research/management science (OR/MS) techniques, primarily mathematical programming. However, the additional topics that we include in this chapter have been studied by researchers with a diverse set of backgrounds (e.g., physics, economics, and reliability), bringing different assumptions and problem solving techniques to the forefront. This review includes work by researchers from industrial engineering/operations research/management science, business/management, geography, computer science, civil engineering, physics, mathematics/statistics, political science, and economics. The diversity of backgrounds amongst researchers in the area of network disruptions can also be observed by the many different types of journals in which the papers in this review were published. We believe that the inclusion of the additional topics into this

paper will help expose researchers from many different disciplines working on network disruption problems to different ways of approaching these problems. We consider the main contributions of our review paper to be: 1) a comprehensive review of network disruption problems, covering many different application areas with a focus on how researchers have modeled these problems; 2) a helpful classification of this body of literature that includes work done by researchers from a diverse set of backgrounds; 3) a discussion of the gaps and imbalances in this body of literature; and 4) an identification of important areas for future research.

The remainder of this chapter is organized as follows. In Section 2.2, we define key terminology and introduce the scope of the review. In Section 2.3, we discuss models that are descriptive in nature. That is, their purpose is to make descriptive observations about systems that are subject to disruptions. Sections 2.4–2.6 focus on models that are prescriptive in nature. That is, they are typically used to recommend a particular course of action. Section 2.4 discusses general modeling techniques, which can be applied to variety of problems relating to network disruptions. Section 2.5 discusses design models, or those that can be used to explicitly consider disruption risk when designing a new network. Section 2.6 reviews various strategies for reducing the disruption risk of existing networks. We conclude in Section 2.7 with summary remarks and directions for future research.

2.2 Definitions, Classification Scheme, and Scope

2.2.1 Definitions

Before beginning this discussion it is important to clarify some terms that are often used in this research area. We base several of our definitions on the DHS Risk Lexicon (The Department of Homeland Security Risk Steering Committee, 2008). For clarity, whenever possible we use the terminology presented in this section rather than the terminology used in the particular papers cited.

In the introduction, we defined a *network* as a collection of nodes along with a collection of arcs representing physical or virtual relationships between nodes. Associated with the network are

one or more measures of *performance*, which measure how well the network performs its intended function. An *element* of the network is a node, arc or some special part of the network (e.g., a supplier or customer). In many cases nodes and arcs can be considered interchangeably; therefore we use the term ‘element’ as a generic term. When they cannot, we designate accordingly. An *element group* is a collection of elements. The *state* of an element or network may be operating, failed, or some level in between.

In this chapter we study networks under the possibility that *incidents*, such as natural disasters or terrorist attacks, may occur. When an incident is caused intentionally we call it an *attack* and when it occurs randomly we call it a *random* incident. Each incident has a *likelihood* of occurring. Related to incidents is the *failure* of an element. When failures are random, they are often assumed to be independent for the sake of tractability. Unless stated otherwise, the reader may assume that a paper considering random failures makes this independence assumption.

An *event* is an incident that degrades, causes the failure of, or destroys one or more elements. Not every incident is an event. The *degradation* of an element is a reduction in the performance of that element (e.g., capacity reduction, cost increase, quality decrease). In addition, events may be cascading, where the failure of one element causes flow redistribution and the overload of other elements, in turn causing them to fail. A *disruptive event* is an event that reduces the performance of the network. Again, not every event is a disruptive event, especially in the presence of redundancy.

The *vulnerability* of an element or element group is its susceptibility to degradation or failure given that an incident occurs. The vulnerability of a network is its susceptibility to a disruptive event, given that an incident occurs. In the remainder of the chapter we use the term vulnerability to refer to both element vulnerability and network vulnerability. In most cases, the intended use will be clear by the context, but when it is not we will specify. Some papers combine the likelihood and vulnerability of an element into a single number. When this is the case, we call this number the *failure probability* of the element. The *consequence* of an incident is the amount of damage and performance decrease it causes. A consequence may be in regard to an element, element group, or in regard to the entire network. In addition, consequence may also be *local*, such as the cost of

repairs or the number of lives lost. The *worst-case* consequence is the largest consequence possible given specified assumptions about the nature of the incident (e.g., at most two elements may fail at a time). We let the term *recourse* refer to the decisions and actions taken after a disruptive event to reduce its consequence. For example, when a bridge fails, the recourse in a transportation network may be to reroute trucks along an alternate route. The *risk* to a network is defined as the potential for a disruptive event and is usually given as a function (typically the product) of likelihood, vulnerability, and consequence.

The *robustness* of a network is its capability to perform well under the occurrence of incidents. Quantitatively, we define the robustness as the amount of consequence associated with a given failure strategy and magnitude (e.g., single element failures). For example, how much consequence does a single element failure cause? Network robustness may also be measured by the probability that an attack on the network causes the network to fail. We use the term *network reliability* as the probability that a network performs its intended function for a given amount of time under the occurrence of incidents. The difference between robustness and reliability is that robustness considers vulnerability and consequence only, while reliability also considers likelihood. The term *network survivability* relates to how many attacks a network can withstand before it cannot perform its intended function. In physics, resilience is the ability of a material to absorb and recovery energy. In this chapter we define the *resilience*² of a network as the ability of the network to be restored after an incident. A network is said to be resilient if it can ‘bounce-back’ after a disruption to the same state, or, in some cases, a better state.

When considering the possibility of terrorist attacks in the context of network risk, two agents or players work against each other in a competitive game. The *defender*, referred to in the literature as ‘she’, wishes to reduce the risk of the system via various actions such as design and risk reduction strategies. The *attacker*, or “he”, seeks to inflict damage on the network. If the attacker chooses the attack with the worst case consequence, we call him an *interdictor*. Unless otherwise noted, we assume that the interdictor’s actions are binary. That is, the interdictor either attacks an

²We credit Dr. Jose Emmanuel Ramirez-Marquez of Stevens Institute of Technology for this definition, described during a seminar at the University of Arkansas.

element or does not attack the element. If the attacker uses a non-optimal, or heuristic, strategy to choose his attack (e.g., attacking the element with the highest load), we call him a *strategic attacker*.

There are several ways to model disruptions. One way is via a joint probability distribution, derived to account for incidents implicitly. This approach is useful for tractability. In another common approach, incidents are explicitly modeled as a set of incident scenarios, each associated with a probability. This approach allows more detail to be included in the model at the expense of tractability. Some models consider all possible incident scenarios while others consider a limited set of scenarios. For example, for some problems it is reasonable to assume that incidents only affect a single element and therefore the model would only include scenarios where a single element fails.

A distinction can also be made between models that study *everyday networks* and those that do a *contingency* study. Models of everyday networks typically start with a classic logistics problem such as a facility location problem as the ‘underlying model’. The underlying model is then modified to account for disruptive events. These models usually provide a tradeoff between performance when the network is in a non-disrupted state (i.e., everyday operations) and the risk of disruptions. Separately, contingency studies either study contingency networks or only account for the post-disruption performance of everyday networks. *Contingency networks* are designed to operate only in response to a disruptive event although they are constructed prior to the event. An example of this is the prepositioning of inventory in preparation for a disaster.

To aid the reader, we also provide a mathematical framework for studying disruption problems. Consider a system characterized by a function h that measures its operational performance. Also, let ξ represent a random incident and z an incident due to an intentional attack. Before the disruption occurs, suppose the network is designed by a defender. Let x be the design decisions. These decisions are typically long-term and are said to be strategic decisions. Similarly, rather than designing a new system, the defender may wish to use various risk-reduction strategies, such as hardening parts of the network or adding redundancy. Risk reduction strategies, denoted by y ,

may be strategic, such as adding redundancy, or tactical, such as some counter-terrorism decisions. The last stage is the recourse stage, where decisions are made to minimize the consequence of the disruption. These decisions are constrained by the state of the network resulting from the disruptive event in the previous stage. Often, the problems that occur in this stage are classic logistics optimization problems such as shortest path and maximum flow problems (see Ahuja et al. (1993)). The expected recourse function $h(x, y, \xi, z)$ represents the expected post-disruption performance as a function of design and risk-reduction. The decisions made in this stage are operational, such as choosing how goods should flow through the network. Note that this function does not make any assumptions about whether a random disruption may occur at the same time as an attack. Figure 2.1 presents the sequential nature of these decisions.

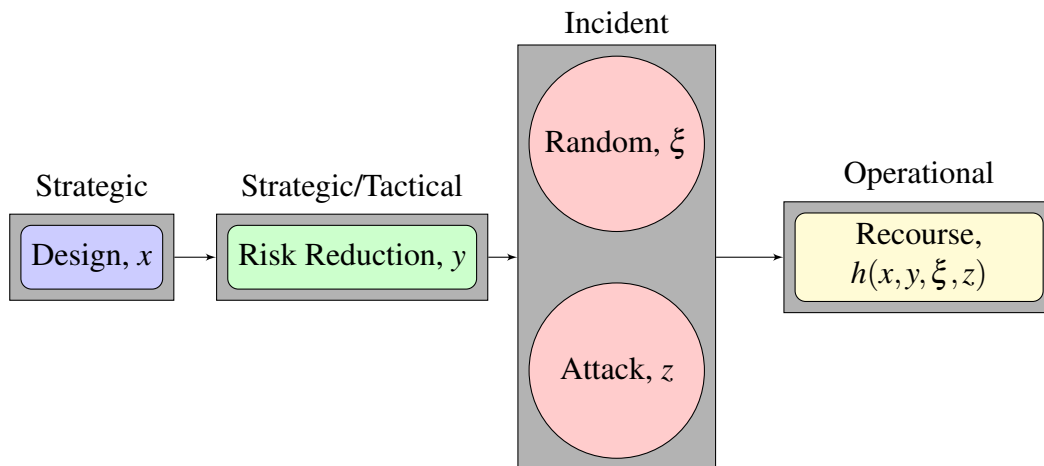


Figure 2.1: Network problems with disruptions: problem stages.

2.2.2 Classification Scheme

Papers in this survey are classified according to the following characteristics:

1. **Type of Network.** We characterize a network type by its topology and in some cases which parts of the network are vulnerable to events. In general, the suppliers, transshipment nodes, or arcs of a network may fail. The first type of network that we discuss is not really a network but a *set of elements*. These elements, also referred to as targets, are independent of each other such that the failure of one target does not affect the other targets. Also, most of the

time the spatial location of these networks is irrelevant to the problem. Rather, the important aspect of problems considering this type of network is the allocation of resources between targets. Many counter-terrorism-related papers consider this type of network. The next type of network that we consider is the *simple network*. These networks are called simple because of two characteristics: 1) elements can be in one of two states (operating or failed), and 2) every combination of element states results in one of two network states (operating or failed). An example of this is a series network, where the network is operating if and only if all of the elements are in the operating state. This allows the network risk to be expressed as a closed-form function of the individual element vulnerabilities and consequences, usually resulting in tractable equations. A significant amount of the work relating to risk reduction of simple networks has involved the following type: parallel, series, series-parallel, and parallel-series. In the literature it is more common to refer to these as ‘systems’; however, whenever possible we use the term ‘network’ for clarity. We also consider what we call *facility networks*, which are networks that consist of a set of facilities and a set of demand points. Facility networks are obtained by solving a facility location model. In section 2.5 we distinguish between two types of facility networks. In Section 2.5.1, we examine networks whose facilities are vulnerable to failure, such as in a supply chain network. These networks usually have a tree structure after the network is designed and are distinguished by the fact that direct connections exist from each demand point to its designated facility. In Section 2.5.2, networks whose arcs are vulnerable to failure are considered. These have a different topology from their facility-failure counterparts in that there no longer are direct connections between demand points and facilities. Because arcs may fail, the entire network, including transshipment or intermediate nodes, must be included. Thus, each demand point is connected to its designated facility via a path, or a set of arcs. It is worth noting that problems considering unreliable intermediate nodes can be modeled as a problem considering unreliable arcs, and vice versa (see Corley and Chang (1974)). These networks are often studied in the field of communications. Finally, we discuss models representing *complex networks*.

These networks have a general topology rather than a simpler topology like series or parallel. As a result, they are more difficult to represent analytically. These networks are often modeled as directed or undirected graphs and the arcs of the network may be weighted or unweighted. These networks often arise in problems such as computing the minimum cost flow or the maximum flow.

2. **Design or Risk Reduction?** The types of decisions made in the *design stage* depend on the type of network that is being designed. In facility location problems the decision is typically to decide how many facilities/sources to locate and where they should be located. In network design problems, the decision is typically to decide which arcs or transshipment arcs to build. In most cases, such as building warehouses, these are strategic decisions. However, in some cases, such as in combat operations, these may be operational decisions. In the *risk reduction stage*, rather than building new network elements, various strategies are employed to reduce the risk of an existing network. The strategy of adding elements or other forms of redundancy to a network could be considered as both design and risk reduction. When redundancy is considered in the design of a new network, we classify it as a design decision. However, when redundancy or new elements are added to an existing system we classify it as a risk reduction decision.
3. **Risk Measure.** Each model in this chapter assumes some risk measure. The risk measure defines how risk is captured in the model. When both the likelihood of incidents and the vulnerability of elements is known (i.e., the failure probability is known), a popular measure of risk is the *expected value* of the recourse function, or the expected consequence. When only the vulnerability of elements is known, then a common risk measure is the *conditional expected value*, which is the expected consequence given that an incident occurs. This measure does not require likelihood values. Other models measure risk as the *worst case consequence*, capturing the preferences of a risk averse decision maker. This approach is attractive because it avoids the problem of having to estimate likelihoods. Some mod-

els consider that disruptions occur due to attacks by an *attacker*. While an attacker can be modeled as a static threat like a natural disaster, several researchers have argued that this is not the correct approach (Bier et al., 2009; Hausken, 2002). They argue that the attacker should be modeled as being adaptive to the defender's decisions, using game theory. In the case of an interdictor, the attacker's disruption is the same as the worst case disruption for a given attack strategy and magnitude. We also consider the *survivability* and *robustness* risk measures, which were both defined in Section 2.2.1. Some papers measure risk using a *risk metric*, which is a proxy for the true risk measure. Each of the risk measures mentioned in this chapter may either appear in the objective function or as a constraint. Additionally, some models measure risk as a combination of the above approaches. Snyder and Daskin (2007) discuss other risk related modeling frameworks such as minimizing expected cost while bounding the cost for a scenario, minimizing absolute regret, and others.

2.2.3 Scope and Related Work

The strategy of this paper is to review different approaches for mitigating against the risks to networks. All of the papers reviewed are analytical in nature and primarily deal with operations research models.

We also limited our search to articles that considered the system-wide impact of disruptions. That is, we consider systems of elements where the overall performance depends on the state of all of its elements. Thus, we did not review work that studies local consequences such as lives lost, repair costs, etc. There has been a lot of work done in areas such as risk analysis studying the local impacts of disruptions.

Finally, in this paper we consider large-scale disruptions as opposed to disruptions due to wear and tear. However, we acknowledge that there is a fine line between these two sources of disruptions.

2.3 Descriptive Models

In this section, we discuss descriptive models, or those that describe or analyze the changes in system performance as a result of disruptions. Because of the existence of surveys and books on this topic, the purpose of this section is to raise key points that will be useful in the exposition of the remainder of this review and to refer the reader to useful references. Sullivan et al. (2009) provide a helpful categorization and discussion of the literature in network disruption analysis, focusing on analyzing network vulnerability and reliability. Murray et al. (2008) and Grubestic et al. (2008) categorize and survey approaches for assessing network vulnerability. They categorize vulnerability assessment into four approaches: scenario-specific, simulation, strategy-specific, and mathematical modeling. Understanding the last two items is important in reading the rest of this review so we briefly discuss them here.

Strategy-specific vulnerability assessment approaches seek to identify the vulnerability of a network to specific types of attacks, such as random failures and attacks on the nodes with the highest degree. In the last decade or two, there has been a lot of interest in developing theoretical models to describe the topology of real-world networks such as biological networks, social networks, the world-wide web, etc. Prevalent models include random networks (Erdos and Renyi, 1959), small-world networks (Watts and Strogatz, 1998), and scale-free networks (Barabasi and Albert, 1999). Scale-free networks exhibit a hub-and-spoke structure, where a small number of nodes have a high degree. Small-world networks are characterized by a small average shortest path length between nodes and a high amount of node clustering. In addition, there has been considerable interest in assessing the vulnerability of these models to specific attack strategies. The vulnerability of these networks is measured as the change in a network efficiency measure, such as the length of the average shortest path, resulting from an event. In particular, researchers have found that scale-free networks, which model networks such as the world-wide web, have a high survivability or robustness against random incidents (e.g., random failures) but a low survivability and robustness against intentional attacks (Albert et al., 2000). The metrics for survivability and robustness typically involve some measure of network connectivity. This body of literature is

discussed further in Grubestic et al. (2008).

Mathematical modeling approaches seek to identify the worst case consequence of a network. The worst case consequence is often measured using an interdiction model, where a strategic attacker seeks to inflict maximal damage on a network. Models have been presented for the interdiction of facilities (Church et al. (2004)), shortest path networks (Israeli and Wood (2002)), and maximum flow networks (Wood (1993)). Smith (2011) provides a basic introduction to interdiction models and Smith and Lim (2008) present a more extensive discussion. Church et al. (2004) categorize interdiction studies (see Table 1 in their paper).

Church and Scaparra (2006) present a novel approach for graphically displaying the reliability of a system subject to an interdictor, called a reliability envelope. The reliability envelope is a graph of system efficiency after a disruptive event versus the magnitude of the event. For each system disruption magnitude, the decision maker can see the best case consequence, the worst case consequence, and the difference between the two, giving the decision maker a broader description of the risk to the system. To determine the worst case consequence the model in Church et al. (2004) is used and the authors develop a model for determining the best case. However, this approach can be used to develop a reliability envelope for any type of system and interdiction scenario. Demonstrating this, the authors develop a model for the stochastic interdiction of facilities and use it to construct a probabilistic reliability envelope.

2.4 General Modeling Techniques

Before beginning the discussion on specific models for design and risk-reduction, this section discusses modeling techniques that are useful for any type of problem with the two-stage structure described in Section 2.1.

A popular technique for both design and risk-reduction problems is to formulate it as a mixed-integer program (MIP) and then solve it via the wealth of methods available for this formulations. However, there are many network design and risk reduction problems for which researchers have not been able to formulate them as MIPs. These include many stochastic problems and problems where disruptions are due to an optimizing attacker. It is important to note that the strategic nature

of the problems discussed in this review make them usually require integer variables. As a result, it is rare for these problems to be formulated as linear programming (LP) problems.

Stochastic mixed-integer programming, which is a special case of mixed-integer programming, is a very powerful technique whose scenario-based framework is well suited to problems involving random disruptions. One advantage of this technique is that the user is given a lot of flexibility in defining scenarios and therefore it is not difficult to add additional side constraints. Also, there exist well established methods for solving stochastic programs. A drawback of this method is that for some problems, the number of possible failure scenarios is quite huge, making it very difficult to solve. Bailey et al. (2006) introduce a modeling framework for defender-attacker problems called stochastic programming with adversarial recourse (SPAR). In the first stage of their model, the defender makes long-term strategic decisions such as design of a new network or reducing the risk of an existing network. After the first stage, design uncertainty is realized in the form of discrete scenarios. After the design decision is made and the design uncertainty is realized, the attacker attacks the system. The attacker's problem is a stochastic, multi-period interdiction problem, modeled as a Markov Decision Process (MDP). Uncertainty in the attacker's problem is a function of the design decision and the design uncertainty realization. Thus, the SPAR model is a stochastic program with Markov Decision Process (MDP) subproblems.

Finally, a modeling framework that is useful for the situation when disruptions are due to an attacker is game theory. This approach is used extensively in the papers discussed in Section 2.6.1. The benefit of using game theory is that useful analytical results can be obtained. The drawback is that these analytical results can usually only be obtained for very simple networks. A special type of game, called a Stackelberg game (Von Stackelberg and Peacock, 1952), occurs when a defender's actions are followed by an attacker. Stackelberg games can be modeled as a bilevel optimization problem (Bard, 1998), where the defender solves the leader's problem and the attacker solves the follower's problem. Bilevel optimization problems are in general quite difficult to solve.

2.5 Design Models: Reducing the Risk of New Networks

Section 2.3 discussed ways to assess the vulnerability and risk of an existing network. However, a decision maker may also wish to consider risk when designing new networks. In this section we discuss models that incorporate risk into the strategic design of networks. They demonstrate that, in general, increasing redundancy (increasing excess capacity) and utilizing elements with less risk are strategies that reduce overall system risk. However, adding extra capacity and choosing elements with less risk usually is costly. A general result of the models in this section is that a large reduction in risk can be obtained by a relatively small increase in the cost of the design.

Figure 2.2 shows the left side of a tree diagram that describes the organization of this review. The leaves of the tree represent the different categories into which the papers in this review are organized. The first row of the tree after the root node divides papers by the type of strategy they consider, whether design or risk reduction. The second row classifies papers by the type of network considered. The third row represents the risk measure considered in the paper; the risk measures listed are expected value (EV), worst case (WC), survivability (SURV), robustness (ROB), and risk metric (MET). The nodes in the third row, each representing a category, contain the number of papers in the category as well as the bibliographic numbers of each of the papers in the category. The child nodes of the 'Facility network: facility failures' node are also categorized by the recourse measure used, namely distance-related measures (Dist.) or coverage-related measures (Cov.) and by whether or not they are location-inventory models (Loc.-Inv.). Also the papers within the survivability child node under the 'Complex networks' node are classified by whether the model seeks to maximize survivability subject to a cost constraint (Max. Surv.) or minimize cost subject to a survivability constraint (Min. Cost.). For space reasons, we left out all of the categories that did not have any papers in them. These categories may not have any papers because either the category is not relevant or because the category is truly a gap in the literature. The categories without any papers are discussed in Section 2.5.4.

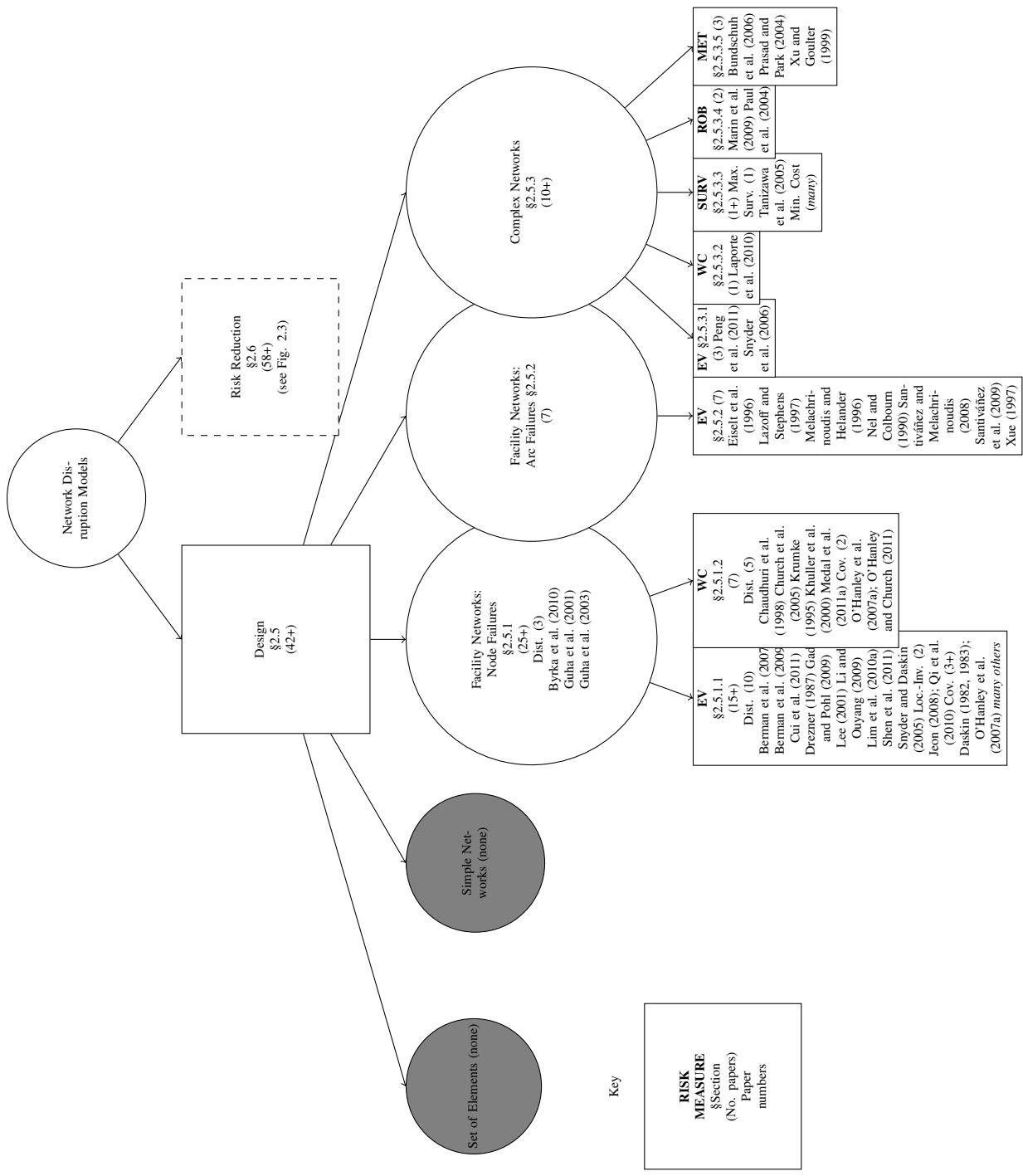


Figure 2.2: Classification diagram (left side)

2.5.1 Facility Networks: Facility Failures

In this subsection, we discuss models relating to facility location problems. It is assumed that the reader has a basic knowledge of the facility location literature. More information can be obtained from Daskin (1995) and Francis et al. (1992). Various recourse objectives exist among these models. The most common objectives are *distance-related* objectives, where network performance is related to distance or weighted distance between customers and their closest located and operating facilities, *coverage-related* objectives where customers are assigned to facilities to minimize the weighted number of customers that are left unserved, and *connectivity-related* objectives, where the objective is to try to connect as many demand points as possible. In this section papers are grouped by their recourse objectives.

We also classify facility network models according to the type of the set of possible facility locations. First, location problems in a *plane* allow facilities to be located at any point within a plane. Second, when the set of possible facility locations is a *tree network*, facilities may be placed anywhere on a network that does not contain cycles. Customers are usually located on the nodes of the tree network. These models are often more tractable because there is a single path between each pair of nodes in the tree network. Third, when the set of possible facility locations is a *cyclic network*, facilities may be placed on a network that contains cycles. The network is assumed to be undirected unless otherwise indicated. Again, customers are located at the nodes and facilities may be located anywhere on the network. Finally, we consider *discrete* location problems where facilities can only be located at a finite set of candidate locations.

One of the more general ways to model a system of unreliable facilities is as a fault tolerant location problem, which generalizes the classic location-assignment problem (Cooper, 1963). The generalization requires that each demand point be assigned to r_j facilities. Each demand-facility assignment carries a weight, with the closer facilities having a higher weight; that is $\lambda_{1j}\lambda_{2j} \leq \dots \leq \lambda_{r_j j}$, where λ_{kj} is the weight assigned to the term corresponding to the assignment of demand point j to the k^{th} closest facility to j . We refer to the version of this problem where exactly p facilities are to be located as the Fault Tolerant p -Median Problem (FTPMP) and

the version where each facility location is assessed a fixed charge is called the Fault Tolerant Facility Location Problem (FTFLP). All of the papers that we found in this area involved solving the FTFLP via approximation algorithms that perform various rounding techniques to the solution of the LP relaxation of this problem (Guha et al., 2001, 2003; Byrka et al., 2010).

2.5.1.1 Expected Value Risk Measure

Other authors have looked at special cases of the FTFLP and FTPMP. In particular, several authors have assumed that failures occur randomly and sought to minimize the expected total weighted distance. These types of problems have been referred to as reliability location problems. The p -facility version is referred to as the reliability p -median problem (RPMP) and the fixed charge version is referred to as the reliability facility location problem (RFLP). The models developed for these problems consider either 1) facilities have a uniform (same) failure probability, or 2) facilities have non-uniform (different) failure probabilities. We denote the uniform failure probability version of these problems as URPMMP and URFLP, and otherwise assume that the probabilities are non-uniform.

The FTPMP and RPMP can be used to model the URPMMP and URFLP with uniform failure probabilities q , by setting the weights $\lambda_k = q^{k-1}(1 - q)$. Snyder and Daskin (2005) present a model for the discrete version of the URPMMP. A multiobjective MIP formulation is given that trades off normal operating costs (not considering failures) with expected operating costs (considering failures). An efficient Lagrangian relaxation algorithm is presented to solve the multiobjective model.

Snyder and Daskin (2005) also present a multiobjective model and Lagrangian relaxation algorithm for the discrete URFLP. The authors demonstrate empirically via a tradeoff curve that a large decrease in risk (expected costs considering failures) can be obtained by a modest increase in day-to-day operating costs. Shen et al. (2011) present a model for the URFLP that is similar to Snyder and Daskin (2005) and develop a 2.674-approximation algorithm³.

³A α -approximation algorithm is a polynomial-time algorithm that always provides a solution with an objective value of at most α times the true optimal objective value.

The FTPMP and FTFLP can also be used to model the RPMP and RFLP with non-uniform failure probabilities q_i , by setting the weights as $\lambda_k = (1 - q_k) \prod_{i \in \bar{I}(k)} q_i$, where $\bar{I}(k)$ is the set of located facilities with a cost rank lower than k . Drezner (1987) presented a model for this problem in the plane without the common assumption of independent failure events. A neighborhood-search heuristic is given that decomposes the problem into p 1-median subproblems at each iteration. Lee (2001) presents a heuristic similar to Drezner's but using space-filling curves. Berman et al. (2007) develop a nonlinear formulation for the RPMP on a network. They solve the model using heuristics. They present structural aspects of their models and show that if co-location is allowed (multiple facilities can be located at the same site), then the Hakimi property (Hakimi, 1964, 1965) holds, which states that optimal facility locations are located at the nodes of a network even if they can be located anywhere on the network. Berman et al. (2009) study the URPMMP on a network where facilities are subject to failure and customers may not know if a facility has failed before visiting it. If a customer visits a failed facility he or she travels directly to the next closest facility. They seek to minimize the total expected cost of customer travel. They assume that facility failures are independent and equally likely. They analyze the structure of optimal solutions and provide heuristics to solve the problem.

Cui et al. (2011) present a continuum approximation (CA) model (see Daganzo (1984a,b); Daganzo and Newell (1986)) for the RFLP in the plane. In this model, the failure probability is a function of the facility location. They show that their model solves quickly and serves as a good approximation to the discrete RFLP. Interestingly, the CA model is able to predict total costs without details about facility locations and customer assignments. This type of model is particularly useful because it can be solved in closed form, enabling it to provide managerial insights and sensitivity analysis. Li and Ouyang (2009) present a CA model similar to that in Cui et al. (2011) but allow facility failures to be correlated. Shen et al. (2011) present a mixed-integer nonlinear program (MINLP) model for the RFLP. They solve their model using a greedy algorithm and a genetic algorithm. Cui et al. (2011) present a mathematical model for the RFLP that is very similar to that presented in Snyder and Daskin (2005) for the URFLP. Although the first model the

authors present is nonlinear, they use a standard linearization technique to convert it to a MIP. Cui et al. (2011) also use Lagrangian relaxation to solve their MIP formulation. Shen et al. (2011) also present a MINLP model for the RFLP but with the extension of including multiple facility types, where each type has its own failure probability. Lim et al. (2010a) take a slightly different approach in modeling the RFLP. They formulate the problem as a MIP and assume each demand point can have an unreliable primary facility as well as a perfectly reliable backup facility. This model is discussed in more detail in Section 2.7.2.

Another popular model in the facility location literature is the capacitated fixed-charge location problem (CFLP), which extends the FLP by assuming that each facility has a finite production capacity. While the classic CFLP has received lots of attention, there has been little work on the CFLP considering facility failures. Following our notation above, we refer to the CFLP with non-uniform failure probabilities as the reliability capacitated fixed-charge location problem (RCFLP) and the CFLP with uniform failure probabilities as the uniform reliability capacitated fixed-charge location problem (URCFLP). Snyder et al. (2006) give a scenario formulation for the RCFLP and discuss several side constraints that can be added to the model. (They do not present a solution procedure.) Gade and Pohl (2009) solve the RCFLP via a sampling technique called sample average approximation, which is often used to solve stochastic programming problems. All of the models for the RFLP and CFLP demonstrate that locating more facilities reduces disruption risk.

There have also been efforts to consider inventory costs in location models under disruptions. In a recent dissertation, Jeon (2008) considers both location and inventory in a supply chain subject to disruptions. The model presented in the first paper of the dissertation is an extension of both the URFLP and the location model with risk pooling (LMRP) (Daskin et al., 2002; Shen et al., 2003), which approximates inventory costs in a location model. The LMRP incorporates the cost savings resulting from the risk-pooling effect, which states that the pooling of inventory at a few distribution centers is cheaper than storing smaller quantities at many retailers. This effect drives LMRP solutions toward locating a smaller number of facilities. However, as discussed above, the presence of disruptions in the RFLP drives solutions to locate a larger number of facilities.

Thus, this model, which we term the uniform failure probability reliability location model with risk pooling (URLMRP), models this tradeoff. Because the inventory costs introduce a concave term in the objective function, two alternative approaches are used to account for the nonlinearity: a Lagrangian relaxation approach and a piecewise linear approximation of the objective function using special ordered sets of type 2 (SOS2). In the second paper of the dissertation, three more models are presented that relax some of the assumptions made in the URLMRP. The first model adds a distance requirement for a retailer to be served by a distribution center (DC). The second model relaxes the assumption that each retailer must be served by a DC. These two models are solved using the SOS2 approach. The third model considers the heterogeneous failure probability and capacity version of the URLMRP, which we denote as the CRLMRP. They model it using a scenario-based formulation and solve it via sample average approximation with SOS2 (SAA-SOS2). In the third chapter of the dissertation, a model is presented for the multi-echelon version of the CRLMRP, considering the presence of suppliers that serve DCs and that they may themselves fail. Suppliers are uncapacitated and for a supplier to serve a DC, it must first be activated for that DC, incurring a fixed cost. This model is also modeled as a scenario-based model. SAA-SOS2 is used to solve the problem as well as a Tabu search algorithm. Qi et al. (2010) consider disruptions in a location-inventory model with a different assumption. They assume that disruptions only affect inventory costs. Thus, they assume that when events occur, retailers wait until it is over, rather than sourcing from the next closest supplier as in the URFLP.

Another popular location problem is the maximum-covering location problem (MCLP), which seeks to maximize the weighted customer coverage. Each customer has a pre-specified cover distance. As in other location problems, various distance metrics can be used such as planar distance or shortest path distance within a graph. A facility covers a customer if the distance from the facility to the customer is less than the customer's cover distance. Daskin (1982, 1983) extends this model by relaxing the assumption that facilities are always available, assigning facilities an identical failure probability. A facility failure may leave some customers uncovered; thus, the objective is to maximize the expected coverage. Thus, the problem is termed the maximum ex-

pected covering location problem (MEXCLP). The MEXCLP, along with several variations, has been fairly well studied. Rather than attempting a comprehensive literature review, we refer the reader to Daskin et al. (1988) and Berman and Krass (2002), who provide reviews relating to this topic and others. In general, these problems assume that failures are due to congestion, which is due to demand uncertainty. This differs somewhat from the focus of this review, which is disruptions, which we assume to be due to external factors. However, many of the models relating to the MEXCLP can be used to model disruptions as well.

O’Hanley et al. (2007a) present a MEXCLP-type model for a species conservation problem. The problem is to locate reserve sites among a set of locations each of which contain a population of endangered species. Each reserve site may fail independently according to a given failure probability. If a species is present at the location of one or more non-disrupted reserve sites after a disruption, the species is said to be covered for that disruption. The objective of the model is to maximize the expected species coverage. The original model presented is nonlinear but a piecewise linear approximation model is proposed. This model differs from the MEXCLP in that it assumes site failures are due to natural- or human-caused disruptions, rather than due to congestion.

2.5.1.2 Worst Case Risk Measure

Rather than including the costs of all assignments made at all levels, as in the FTPMP and FTFLP, some researchers have taken another approach. In this context it is assumed the decision maker is interested in minimizing the worst case sum of assignment costs, assuming that at most r facilities can fail at a time. This approach has two applications. First, it is useful in the case where failures are caused by an intelligent antagonist, who attempts to cause the most damage possible. Second, it is useful in the case where the decision maker is risk averse, seeking to prepare for the worst case. Another benefit of this approach is that it alleviates the need to estimate the likelihoods of various failure scenarios. Church et al. (2005) takes this approach in introducing a version of the discrete p -median problem that accounts for an interdictor that attacks facilities. This problem seeks to locate a set of p facilities that have the lowest worst-case consequence due to a disruptive event. This problem is modeled as a bilevel MIP and uses the median interdiction model of Church

et al. (2004) (see Section 2.3) for the lower-level problem.

Other researchers have taken a slightly different approach in minimizing the worst case assignment cost. Rather than minimizing the worst case *total* assignment cost, the objective is to minimize the worst case *maximum* assignment cost over all demand points. This has been called the r -neighbor p -center problem, the fault tolerant p -center problem, and the (p, r) center problem. Drezner (1987) studies the Euclidean distance version of this problem, where facilities may be located anywhere in a plane. Drezner mentions that although the problem can be solved using a set-covering algorithm presented in an earlier paper, the algorithm may be too computationally expensive. As a remedy, he presents a neighborhood-search type heuristic, which involves decomposing the problem into several 1-center problems.

Other authors have studied the discrete version of the r -neighbor p -center problem, where facilities can only be located at a finite set of locations. Krumke (1995) developed a 4-approximation algorithm for the unweighted version. Khuller et al. (2000) presents 3-approximation algorithms for both the unweighted version and weighted versions of the problem. Finally, Chaudhuri et al. (1998) presented a 2-approximation algorithm, the best possible. All of these approximation algorithms utilize graph-theoretic methods.

Medal et al. (2011a) presents a MIP formulation for the discrete r -neighbor p -center problem. A set covering based algorithm was also investigated and found to perform well.

O'Hanley and Church (2011) study the maximum covering location-interdiction problem which seeks to locate a set of p facilities in order to maximize a weighted combination of the initial coverage without failures and the minimum coverage that results from a loss of r facilities. This problem also uses the covering interdiction model of Church et al. (2004) as the interdiction problem. A MIP as well as a bilevel MIP model is presented and the bilevel model is shown to perform better than the MIP model. O'Hanley et al. (2007a) presents a similar model but with a different type of interdiction budget. Rather than restricting the number of interdictions to be r , each location is assigned a failure probability and a constraint is added that restricts the probability of the disruption to be larger than a threshold value, supplied by the user. A bilevel MIP model is presented for the

problem.

2.5.2 Facility Networks: Arc Failures

A number of papers in the telecommunications and computer network literature consider reliable location problems. These papers consider the random failure of arcs or nodes and allow for the possibility that a failure can cause the network to become disconnected. In the models discussed in previous sections, the implicit assumption is made that network connectivity is always sustained when a failure occurs. These models focus more on connectivity than weighted distance. They typically seek to minimize the expected number of customers disconnected after a network failure. This can be thought of as unmet demand in a supply chain context. These problems have some relation to the MCLP discussed previously in this review. The main difference is that in this section, we assume that a customer's demand is met as long as there is a path from that customer to some supplier. In the MCLP problem, the weighted number of disconnected customers objective can be modeled by assigning an infinite coverage distance to each customer.

Santiv  nez and Melachrinoudis (2008) present the problem of locating a facility on a tree with unreliable edges that minimizes the expected number of unsuccessful responses to demand requests over all customers. They call this the reliable 1-center problem and present efficient solution algorithms. They model the operational probability of an edge as an exponential function of physical distance. Santiv  nez et al. (2009) study the same problem on a network. Nel and Colbourn (1990) study the problem of locating a single facility on a network that maximizes the expected number of nodes reachable by operational paths. Melachrinoudis and Helander (1996) study the same problem on a tree. They term this the *reliasum problem*. They present an $O(n^3)$ and an $O(n^2)$ algorithm to solve this problem. Xue (1997) presents an $O(n)$ algorithm for the problem studied in Melachrinoudis and Helander (1996).

Eiselt et al. (1996) introduce the problem of locating p facilities on a network where one node or one link can fail. They seek to minimize expected disconnected demand and term their problem the p -Unreliable Network Location Problem (p -UNLP). A low-order polynomial algorithm is presented to solve this problem optimally. Lazoff and Stephens (1997) investigate the problem

of locating data replicas in a network in order to maximize the availability of the data to demand nodes. They look at the read access problem and write access problem. For read access, demand nodes must be able to connect to at least one data replica. They mention that this is equivalent to the unreliable 1-median problem, which was studied in Eiselt et al. (1996). For write access, demand nodes must be able to access all data replicas. They assume that edges failures are asynchronous (happen one at a time), which reduces the probability space.

2.5.3 Complex Networks

Another line of research has sought to design complex networks under the threat of disruptions. Recall that in this review we define complex networks as those that do not have a series/parallel structure. We refer to this area of research as the design of unreliable complex networks. A rather distinct line can be seen in the literature between models for locating unreliable facilities (LUF), discussed in 2.5.1, and designing unreliable complex networks (DUCN), discussed here. The difference lies in the types of solution approaches. Thus far, a majority of the LUF models have been mixed integer problems (MIP) that are polynomial in size relative to the number of facilities and customers. This is an accomplishment, given that these problems can be thought of as two-stage problems (locate-disruption-allocate) and that they typically consider all possible failure scenarios, a set that increases exponentially in the number of facilities. However, most of the optimization models for the DUCN problem have been either stochastic programming or bilevel programming models, both of which usually are less tractable than polynomially-sized MIPs. This is likely a sign of the increased difficulty of designing unreliable complex networks. This difficulty may lie in the fact that in the LUF, the post-disruption optimization problem (a.k.a., the ‘recourse problem’ in the stochastic programming literature) is easier. Most of the facility location models assume uncapacitated facilities, allowing customers to be assigned to their closest operating facility following a disruption. However, the recourse problem for DUCN is often non-trivial, such as the shortest path or maximum flow problems.

2.5.3.1 Expected Value Risk Measure

Snyder et al. (2006) present a two-stage stochastic programming model for the fixed charge network design problem under disruption event risk. However, they do not give a solution approach. Peng et al. (2011) study the logistics network design problem (LNDP) under disruptions. The LNDP involves the location of capacitated suppliers and transshipment nodes, the assignment of suppliers to customers, and the selection of flows through the network. Both suppliers and transshipment nodes may be disrupted. A p -robust stochastic programming model is presented that minimizes construction and flow costs subject to the constraint that the relative regret in a scenario (cost for a scenario relative to the optimal cost for that scenario) is no greater than p . The model is solved using a heuristic approach.

2.5.3.2 Worst Case Risk Measure

Laporte et al. (2010) study a problem where a defender seeks to design a railway transit network in the presence of an attacker that wishes to inflict maximum damage to the network. The objective of the planner/defender is to maximize the minimum demand met over all single-arc-failure scenarios. They model the problem as both a maximin integer linear program (ILP) and via game theory. This is the only paper that we identified within this category; hence, this category is listed as a ‘gap’ in Section 2.5.4.

2.5.3.3 Survivability Risk Measure

Researchers have also tried to figure out how to design networks to improve survivability. There are two main approaches that researchers have taken related to survivability, which we discuss in this section.

First, researchers have studied how to optimize survivability subject to a cost constraint. Tanizawa et al. (2005) present a model for optimizing the survivability of a network subject to waves of failures. Each wave includes both random and intentional failures, which occur at pre-specified rates. They show that the most survivable network in this case is one whose distribution of node degree is bimodal, and derive the optimal distribution parameters. This study is related to the body of work

from the statistical physics community discussed in Section 2.3 because it considers theoretical network topologies.

Second, researchers have looked at how to optimize network design cost subject to a constraint on survivability. A body of research called survivable network design has emerged within the operations research community to address this problem. This body of research has typically modeled these problems as MIP models that minimize network construction cost subject to a requirement that the network maintains connectivity after all single-element failures. Rather than attempting to review all of this literature here, we refer readers to the survey papers by Grottschel et al. (1995) and Kerivin and Mahjoub (2005).

2.5.3.4 Robustness Risk Measure

Marin et al. (2009) present an integrated model for railway network design and line planning under the failure of arcs. The model is designed to produce solutions that are robust in regard to both total travel time and user costs. Robustness is defined as the maximum travel time or user cost increase resulting from a single-element failure. Paul et al. (2004) present a model for maximizing the robustness of a network to both random failures and intentional attacks subject to a cost constraint. Robustness is defined as the probability that an attack causes the network to become disconnected. They identify design rules for various network topologies. Again, because of the lack of papers in this section, it is listed as a ‘gap’ in Section 2.5.4.

2.5.3.5 Risk Metric

Bundschuh et al. (2006) present several models for considering disruption risk in the design of supply networks. In particular, they focus on reliability, robustness, and contingency as risk reduction measures. Reliability is defined as the probability that no elements in the network have failed. The drawback of this definition is that adding additional elements in the design phase, and thereby increasing redundancy, actually decreases reliability. The authors enforce robustness by constraining the amount of goods that can be obtained from one supplier. Finally, contingency is added to the supply chain via emergency safety stock reserved for disruptive events and purchase options

on additional supply in the event of a disruption. Models are developed for each of these measures as well as for combinations of the measures. They find the reliability-contingency model produces the best results. Besides the reliability-only model, all of the models indicate that large reductions in risk can be obtained by considering risk in the design phase. Xu and Goulter (1999) present a model for designing water distribution networks that minimizes cost subject to a constraint on a reliability measure. Prasad and Park (2004) present a multiobjective model for designing water distribution networks that optimizes both cost and excess capacity, a proxy for risk reduction.

2.5.4 Future Work

In this section we discuss areas of future work that relates specifically to the design of networks. In order to assess gaps in this body of research, we briefly discuss here the categories for which we did not list any papers. Recall that these categories were left out of Figure 2.2. To start, we mention the categories for which there is a good reason that we did not find any papers or at least did not list them in this review. The ‘Set of elements’ network did not have any papers but for good reason. In this network the elements are truly independent of each other and their performance is independent of their location. Therefore, the only design decision made here is to decide how many elements to include. Because this is a simple decision, it is usually included along with another decision, such as risk reduction. Therefore, we include these papers in Section 2.6.1. We also did not include any papers studying ‘Simple networks’. The reason for this is that this area of research has already been well studied in the field of reliability optimization. For more information, we refer the reader to a book by Kuo et al. (2001).

There are also a number of categories that we consider to truly be gaps in the literature. Considering facility networks with facility failures, we did not find any papers considering the following risk measures: conditional expected value, survivability, robustness, risk metric, and multiple risk measures. We consider all of these to be relevant but due to space limitations we leave the reader to think about them in more detail. The research considering facility networks with arcs failures has thus far focused on the expected value risk measure and only considered connectivity as a recourse objective. We think that other risk measures are important to study. Coverage-related

recourse objectives may also prove to be interesting. However, if one considers distance-related recourse objectives, then the network is essentially what we define as a complex network, which is covered in Section 2.5.3. Finally, neither the conditional expected value risk measure nor multiple risk measures were considered for complex networks. We think that these are both worthy of consideration.

To assess imbalances in the number of papers that fall into each category, Table 2.1 shows the number of papers studying various combinations of the network types and risk measures. The network types included in the table are those mentioned in this section and the risk measures included are expected value, worst case, risk metric, survivability, and robustness. In our calculation we only include the papers that we described in this review. We can see from these statistics that a majority of papers have focused on the expected value risk measure. The best risk measure to use depends on characteristics such as the network type, the type of disruption, and the risk preference of the decision maker. Thus, it is important for researchers to consider other measures besides expected value.

Table 2.1: Quantity of papers studying different network type and risk measure combinations.

	Risk measure						Total	%
	Expected value	Worst case	Risk metric	Surviv.	Robust.	Other		
Fac. net.: fac. fail	15	7	0	0	0	3	25	60%
Fac. net.: arcs fail	7	0	0	0	0	0	7	17%
Complex networks	3	1	3	1	2	0	10	24%
Total	25	8	3	1	2	3		
%	64%	21%	8%	3%	5%	8%		

2.6 Risk-Reduction Models: Reducing the Risk of Existing Networks

Despite all of the work discussed in the previous section, sometimes it is too expensive or otherwise not possible to design a new system from scratch (e.g., consider the world-wide web). To address this situation, a body of literature has emerged to address how to modify network systems with the purpose of reducing risk. In this section, we distinguish between a risk reduction strategy, which is an abstract action such as increasing security, and a particular risk reduction solution, which

would specify how the increased security is allocated. One might ask the question why we cannot just use the descriptive models discussed in Section 2.3 to find the most critical elements in the network and simply focus on them in a risk reduction strategy. Many authors have pointed out that this may result in less than desirable solutions. The reason is that when a risk is reduced for one element in the network (e.g., increasing the security at a port), the way risk is distributed among all of the network elements changes. Thus, several authors have argued that the risk reduction and risk assessment decisions should be integrated. However, descriptive models can still be used to find the best risk-reduction solution by evaluating different risk-reduction solution alternatives in a total enumeration scheme. However, for many networks the number of possible alternatives is prohibitively large. The models in this section are mostly optimization models, addressing this difficulty. In general, the papers in this section demonstrate that a substantial risk reduction can be obtained through a modest investment in a risk reduction strategy.

There exist many strategies to modify networks to reduce their risk to disruptions. We classify these strategies into 9 categories: vulnerability reduction, likelihood reduction, element consequence reduction, failure probability reduction, redundancy, rewiring, restoration, increasing attacker's cost, and informational measures. Vulnerability reduction strategies attempt to reduce the likelihood that an incident becomes an event or a disruptive event. One common way of doing this is by hardening elements. Another term for hardening which is used a lot in the literature is fortification, which we use when it helps our explanation. The hardening decision is typically represented as a binary variable and if a facility is hardened it cannot fail. Two useful properties about hardening problems have been proven in the literature.

Remark 1. The optimal set of facilities to harden must contain at least one element from the optimal set of interdicted elements (Church and Scaparra (2007)).

Remark 2. An interdicator will never (optimally) attack a hardened element (Scaparra and Church (2008a)).

Remark 1 is rather intuitive; if the defender does not thwart the attacker's optimal solution, then the attacker will not change his strategy. This demonstrates that the solutions generated by

an interdiction model are not sufficient in prescribing which elements to harden. Remark 2 is also intuitive. An attack on a hardened element is guaranteed to not increase the interdictor's objective function. However, an attack on an unhardened element may improve the attacker's objective function.

Some authors have modeled the vulnerability of an element as a function of the amount of defense resources allocated. Related to this approach is the *contest success function* (Skaperdas, 1996), which is often used in the economics literature to model conflicts between players. Using a contest success function, the vulnerability of an element is expressed as a function of both the defender's allocation of protection resources and the attacker's allocation of resources. Contest success functions usually have some sort of contest intensity parameter that determines the form of the function. Another way of affecting the likelihood of attacks is to increase the price that an intelligent attacker pays for an attack. Finally, a defender may use the separation of targets to make failures of elements less dependent, thus reducing the network vulnerability.

Likelihood reduction strategies attempt to reduce the likelihood of an incident. Since it is difficult, if not impossible, to prevent the occurrence of natural disasters, these approaches usually involve preventing terrorist attacks and unintentional man-made incidents. Examples include investment in border defense, counter-terrorist operations, and intelligence (Powell, 2007b). The reduction of the likelihood of man-made incidents, such as a fire in a factory, could be modeled as function of defensive resource allocation which represents preventive measures such as changes in processes. This function has been modeled as a continuous function of the resource investment, which we call *continuous likelihood reduction*, as well as a discrete function, where elements are protected at different levels, which we call *discrete likelihood reduction*. A common approach is to model the reduction of the failure probability which we call *continuous failure probability reduction* and *discrete failure probability reduction*.

Another method of risk reduction is *element consequence reduction*, or reducing the consequence of an incident. If failures are modeled as capacity degradation rather than complete failures, one approach is to invest resources to reduce the amount of degradation that occurs given

a disruptive event. Adding *redundancy* to the network is another way to reduce the risk of the network and make it more robust. This can be done by adding new elements, removing an existing component in order to add a new element in a better location, and increasing the capacity of existing components. *Rewiring* involves redesigning the network without adding components to reduce the risk of the network. *Restoration* measures attempt to reduce the consequence of a disruptive event by allocating resources to increase network restoration capacity. Another method of reducing the risk of a network involves taking measures to increase the cost or effort required by the attacker to attack the network. *Informational* measures involve the use of information, or lack of it, to thwart would-be attackers. Strategies include secrecy, deception, signaling, the use of false targets, as well as increasing the accuracy of the defender's information. Finally, some papers consider combinations of the above approaches and several papers consider tradeoffs between approaches.

As mentioned above, most of the models relating to reducing the risk in existing networks involve investing resources to reduce risk. Many of these models include some budget constraint on the amount of risk reduction resources available to the defender. These budgets are typically modeled as either a constraint on the total number of risk reduction activities or on the total cost of risk reduction.

The modeling construct for these problems depends largely on the complexity of the network. For relatively simple networks, the construct of choice is game theory. Indeed, most of the results obtained for these types of networks involve closed-form expressions analyzed from a game theory perspective. This allows the authors to find defender-attacker equilibriums and make many useful and often counter-intuitive insights into the problems that we discuss in the following sections. Roughly speaking, an equilibrium solution is an overall solution that both sides are satisfied with. In general it is difficult to model more complex networks using closed-form expressions because finding the network state for a combination of element states is often itself a nontrivial optimization problem. As a result, these problems have been studied using more sophisticated and computationally intensive techniques such as mathematical programming and meta-heuristics.

In addition to classifying by network type and risk measure, we also categorize papers in this section by the risk reduction strategy employed. When a paper considers more than one strategy, we classify the paper by its main purpose. For example, if a paper uses a hardening approach to demonstrate the efficacy of secrecy, we would classify the paper as using an informational strategy. However, if a paper's main goal is to model the tradeoff between multiple strategies, then we would classify it as using 'Multiple Strategies'. Finally, they are classified by the risk quantification approach, or whether they quantify risk using expected value, worse case, etc.

The right side of tree diagram describing the organization of this review is shown in Figure 2.3. The first two rows after the root node are organized in the same way as Figure 2.2. The third row again represents the risk measure considered in the paper. Two additional risk measure categories, conditional expected value (CEV) and multiple risk measures (MULT), are included in the third row of this diagram. We also add the consideration of both random failures and intentional attacks (R & A) as a risk measure. The papers within the nodes in the third row are also categorized by the risk-reduction strategy taken in the paper. These strategies include failure probability reduction (Fail. Prob.), vulnerability reduction (Vul.), redundancy (Red.), increasing the cost of attack (Incr. Att. Cost.), rewiring (Rew.), multiple strategies (Mult.), and a tradeoff between strategies (Trade.). As in Figure 2.2 the papers within the survivability child node under the 'Complex networks' node are classified by whether the model seeks to maximize survivability subject to a cost constraint (Max. Surv.) or minimize cost subject to a survivability constraint (Min. Cost.). When papers consider multiple strategies separately in the same paper, we consider each strategy considered to be a separate paper for the purpose of counting the number of papers for each risk reduction strategy. However, we count it as one paper when counting the number of papers for each risk reduction measure. Again, we left out all of the categories that did not have any papers in them. As in Section 2.5, a category may not have any papers because either the category is not relevant or because the category is truly a gap in the literature. The categories without any papers are discussed in Section 2.6.5.

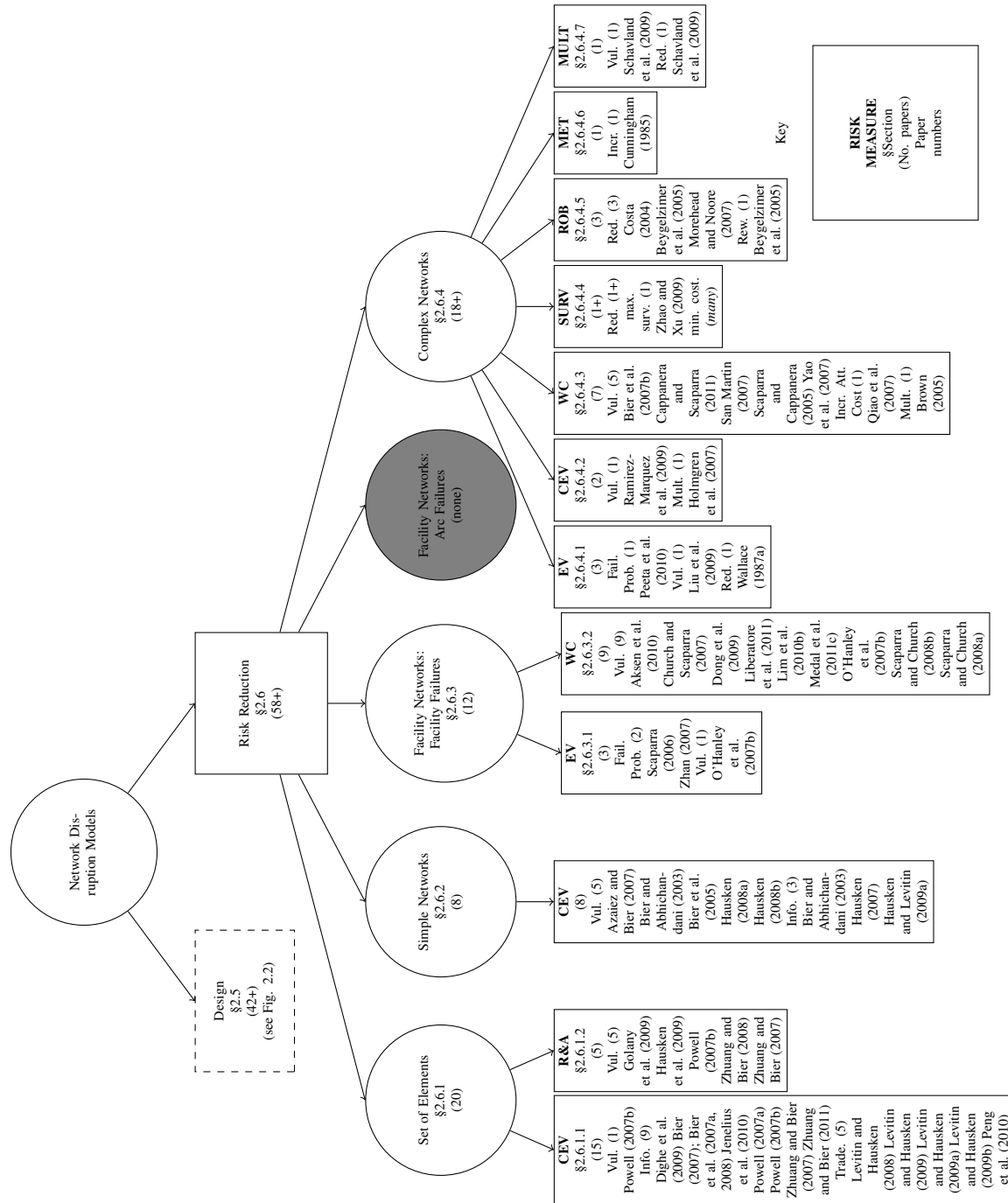


Figure 2.3: Classification diagram (right side)

2.6.1 Set of Elements

2.6.1.1 Conditional Expected Value Risk Measure

Most of the papers on protecting a set of elements have assumed that likelihood cannot be controlled and thus focus on the expected consequence given that an incident has occurred.

Vulnerability Reduction Vulnerability reduction has been modeled using contest success functions Hausken et al. (2009); Levitin and Hausken (2009,a,b, 2008); Peng et al. (2010); Zhuang and Bier (2008), a function of defense resources allocated (Bier et al., 2007a; Bier, 2007; Bier et al., 2008; Jenelius et al., 2010; Powell, 2007a; Zhuang and Bier, 2008), and hardening (Dighe et al., 2009). However, the purpose of most these papers is to examine the efficacy of other measures such as using informational measures, rather than to examine the efficacy of vulnerability reduction. Thus, we place these papers in other categories within this section.

Powell (2007b) takes an interesting look at vulnerability reduction. He presents a model that allows the defender to allocate protection resources between counter-terrorism, which reduces the vulnerability of all elements, and the vulnerability reduction of specific sites.

Informational Several articles have examined the effectiveness of informational strategies used by a defender such as secrecy and deception. Dighe et al. (2009) present a model where the attacker knows how many allocations are made but does not know the particular allocations. They find that partial secrecy is preferable to full disclosure for the defender. Zhuang and Bier (2010) examine three disclosure strategies: truthful disclosure, secrecy, and deception and find that all three strategies may be present at equilibrium. Other papers have modeled situations where the defender does not have perfect information about the attacker. Bier and others (Bier, 2007; Bier et al., 2007a, 2008) present a model where a defender protects a collection of elements against an attacker who wishes to attack a single element. The defender only knows the probability distribution of the attacker's preferences. In this model, the authors find that the defender prefers his allocation to be made public. Most of the work on protection networks assumes that the attacker has perfect information. Jenelius et al. (2010) relax this assumption and provide a model that assumes that

the attacker observes the utilities for attacking particular elements with random observation errors. They find that if the defender falsely assumes that the attacker has perfect information, the defender's allocations could yield significantly suboptimal results. Powell (2007b) also looks at this problem, assuming that the defender has uncertainty about which attacker she will face but knows that she can only face 2 types of attackers.

Another important problem characteristic relating to informational strategies is the sequence in which the two agents act. The agents may play a simultaneous game where neither agent has any information about the other agent's actions. Also, a two-period game may take place where the attacker makes decisions after the defender. In this situation, the attacker may or may not have information about the defender's actions. Zhuang and Bier (2007) and Bier et al. (2007a) show that under certain assumptions, when the defender is able to hide information from the attacker, she has a first-move advantage in a sequential game. Powell (2007a) studies this same situation, focusing on the fact that the defender's allocation sends a signal to the attacker about which elements the defender values. This type of game is called a signaling game. Using a game theory model, the tradeoff the defender makes between protecting her most valuable elements and avoiding the sending of signals.

Tradeoff Between Strategies Given the availability of two or more strategies, the decision maker may wish to know how to divide her resources between the strategies. Several papers have examined this tradeoff and share a number of common traits in how they model the problem (Peng et al., 2010; Levitin and Hausken, 2009a, 2008, 2009b). First, the objective of the network is to meet a demand so that the recourse function is the cost of unmet demand. Second, it is assumed that the defender distributes her resources evenly amongst all or some of the elements. Third, the attacker chooses a subset of elements to attack and distributes his resources evenly amongst them. Fourth, a contest success function is then used to model the element vulnerability. Each of these papers models a different tradeoff between two or more risk-reduction strategies. Peng et al. (2010) present a model where the defender allocates resources between protecting existing (genuine) elements and deploying false elements. They also consider that false elements can be detected to be

false by the attacker with a specified probability. Levitin and Hausken (2009) model the situation where a defender allocates resources between deploying new elements, essentially designing the system, and protecting the elements. Levitin and Hausken (Levitin and Hausken, 2008, 2009a) allow the defender to allocate resources between deploying genuine elements and deploying false elements. As a natural extension to the above tradeoffs, Levitin and Hausken (2009b) allow the defender to trade off between protection, redundancy, and deploying false elements.

2.6.1.2 Defending Against Random Incidents and Strategic Attacks

We deviate slightly from our classification in this section and consider multiple risk measures. The reason is that the papers of this section have a stronger commonality: they each consider the tradeoff between reducing the risk of both strategic and non-strategic (probabilistic) incidents.

Golany et al. (2009) compare the optimal policies for defending a network against probabilistic failures to the optimal policies for defending against a strategic attacker. For both models, the element vulnerabilities are a function of resources allocated by a defender. The objective in the probabilistic case is to minimize the expected value and the objective in the strategic attacker case is to minimize the worst case consequence. They find that the best protection solution against probabilistic attacks involves protecting the elements that received the greatest impact from protection. In contrast, in protecting against a strategic attacker, it is best to allocate protection resources to reduce the maximum vulnerability over all elements. Hausken et al. (2009) present a model that allows the defender to tradeoff between investing in resources for protecting against terrorism, protecting against random failures, and protecting against both (all hazards protection). The objective of their model is to minimize the conditional expected value (consequence) and they use a contest success function to model vulnerability. Zhuang and Bier (Zhuang and Bier, 2007, 2008) present a model where a single defender allocates two types of resources: 1) resources for defending against probabilistic failures, and 2) resources for defending against strategic attacks. The objective of their model is to minimize the conditional expected value (consequence). They use a function similar to a contest success function to model vulnerability to strategic attacks and use a function of defender resource allocation to model probabilistic incident vulnerability. Powell

(2007b) also looks at how to allocate resources to protect against a threat that has both a strategic and a non-strategic component.

2.6.2 Simple Networks

In this section we discuss simple networks, or networks whose topology can be described as a combination of series and parallel sub-networks. The key characteristics of these networks are 1) the elements are assumed to be identical and 2) because the networks have a series/parallel structure, the state of the entire network can be described analytically as a function of the states of the individual elements. These two characteristics make these networks amenable to closed-form, analytical analysis. The papers in this section all consider the conditional expected value risk measure.

Vulnerability Reduction Bier and Abhichandani (2003) consider the defense of both series and parallel networks where the defender allocates protection resources to network elements to protect against an attacker that has the objective of maximizing his success probability. Conversely, the defender has the objective of minimizing the attacker's success probability. Bier et al. (2005) consider the same problem as in Bier and Abhichandani (2003) except that the attacker now wishes to maximize the expected damage of an attack, rather than the probability. These papers model vulnerability as a function of the defense resource allocation. The attacker attacks the element that has the largest attack utility, typically the one with the largest vulnerability. Azaiez and Bier (2007) extend the work of Bier and Abhichandani (2003) by modeling the protection of a combined series/parallel network where the defender allocates resources to maximize the cost to the attacker of the defender's worst case attack.

Hausken (2008b) provides models for defense against a strategic attacker for both series and parallel networks using a contest success function to model vulnerability. Hausken (2008a) extends this work to an arbitrarily complex series/parallel network with the goal of determining whether the defender prefers a parallel-series network or a series-parallel network. He found that when everything else is equal, the defender prefers a series-parallel network.

Informational Bier and Abhichandani (2003) and Hausken (2007) provide results that indicate that secrecy and/or deception may be effective strategies for the defender. Hausken and Levitin (2009a) present a model where the defender allocates resources between protecting existing (genuine) elements and deploying false elements.

2.6.3 Facility Networks: Facilities Fail

Like in Section 2.5.1, in this section we consider facility networks where facilities are prone to failure. The difference between the models in this section and those in Section 2.5.1 is that in this section we are modifying existing facilities rather than building new ones.

2.6.3.1 Expected Value Risk Measure

Failure Probability Reduction Only a few authors have proposed models for failure probability reduction regarding facility location networks. Zhan (2007) presents two nonlinear models for the RMPF, which are fortification versions of the RFLP model presented in Shen et al. (2011). As in Shen et al. (2011), the objective is to minimize the sum of the expected service cost and the fail-to-serve penalty cost. Zhan (2007) presents a model for continuous failure probability reduction and shows that it is a special case of the generalized linear multiplicative programming problem (GLMP) (see Ryoo and Sahinidis (2003) for more details). To solve the model, the vertex enumeration method (Horst et al., 2000) is used, which is a method used for GLMP problems. Zhan also presents a MINLP model for discrete fortification. Because the objective function of this model is monotonically non-decreasing, it can be solved by a monotonic branch-reduce-bound algorithm developed in Zhan (2007). Scaparra (2006) presents models for continuous and discrete failure probability reduction. The straightforward formulations of these problems are nonlinear. To overcome this, network-flow type models, which are linear, are developed. The network models use balance flow constraints to account for the probability that customers are served by a given facility. Although these models can be solved by standard methods for mixed-integer programs, a greedy randomized adaptive search procedure (GRASP) is developed to solve large instances.

Vulnerability Reduction O’Hanley et al. (2007b) study the hardening version of the maximum expected covering location problem (MEXCLP) in the context of biological conservation, which we denote as the maximum expected covering location problem with hardening (MEXCLPH). The problem is to choose a set of sites to denote as reserve sites, which is equivalent to hardening the sites. Each site contains a population of various wildlife species and has a nonidentical probability of failure. A reserve site cannot fail. A species is left unprotected (uncovered) and becomes extinct if all of the sites that it inhabits are disrupted. The objective is to minimize the expected weighted loss of species, equivalent to the minimizing the expected number of uncovered customers. They refer to their problem as the minimum expected coverage loss problem (ECL). The authors model this problem like a maximum covering location problem (MCLP) but add an additional weight (the probability of species survival) to the objective function, resulting in a model that has the same structure as the classic MCLP. The multi-period version of this problem is also studied, where the probability that a species is exterminated is a function of the number of periods it is left unprotected. This problem is modeled as an expected value problem (see Birge and Louveaux (1997) for details).

2.6.3.2 Worst Case Risk Measure

Vulnerability Reduction A majority of the work relating to facilities has dealt with hardening. All of the papers in this section discuss a hardening extension of the r -interdiction median problem (RIM) (Church et al., 2004) developed (see Section 2.3 of this paper) called the r -interdiction median problem with fortification (RIMF). If exactly q facilities can be fortified then the problem is the r -interdiction median problem with q -fortification (RIMQF). This model involves a game against an interdictor subject to a budget constraint that wishes to maximize the total cost of satisfying customer demand. It is assumed that both the defender and attacker have perfect information.

Church and Scaparra (2007) present a MIP model for the RIMQF that minimizes the maximal cost over all possible interdiction scenarios, or all possible ways to interdict r out of p existing facilities. To reduce the size of their model, the authors utilize some properties of the problem to remove unnecessary variables and constraints. Additional variables are consolidated using ideas

from the Condensed Balinski Constraints with the Reduction of Assignment Variables (COBRA) formulation of the p median problem (Church, 2003).

Scaparra and Church (2008a) reformulate the RIMQF model presented in Church and Scaparra (2007) as a maximum covering problem, which enables them to overcome some of the computational challenges of the previous model. Their model essentially tries to cover (prevent) the set of interdiction scenarios that result in the biggest impact. Remark 2 provides a theoretical foundation for this formulation by limiting the possible interdictor scenarios. They then show how heuristics can be used to obtain bounds, which reduce the size of their model. The approach in this paper is flexible because it can handle any underlying model (e.g., covering problem) for which the evaluation of interdiction patterns can be done in polynomial time. This differs from the RIMQF model presented in Church and Scaparra (2007), which is tailored to the structure of the p -median problem. The approach is also valid for the RIMF.

Scaparra and Church (2008b) present a bilevel MIP formulation of the RIMQF. They provide an implicit enumeration (IE) algorithm to solve the problem. This algorithm utilizes Remark 1 to reduce the size of the enumeration tree. Since a RIM problem is solved at each node in the tree, the authors present a streamlined formulation of the RIM and utilize variable consolidation (see Church (2003)) and closest-assignment constraints. They demonstrate empirically that their new RIM problem with the other reductions solves faster the RIM model presented in Church et al. (2004). They also demonstrate computational improvements over the maximal covering approach in Scaparra and Church (2008a).

Lim et al. (2010b) develop a two-population genetic algorithm for the RIMQF, which exploits the defender-attacker competition in the problem. The first population contains the defender strategies and the second contains attacker strategies. As the algorithm progresses, the two populations compete against each other and evolve with this competition. The benefit of this approach is that it does not make many assumptions about the underlying problem so it can be used for any problem that involves hardening elements against an interdictor. It is shown empirically that the algorithm performs well at solving large-scale RIMQF instances.

Several extensions have been made to the basic RIMQF. One of the limitations of the RIMQF is that it assumes that r , the number of disrupted facilities, is known. Liberatore et al. (2011) present a stochastic version of the RIMQF (S-RIMQF), where only the probability distribution of r is known to the defender. They present a maximum covering type formulation that is similar to that in Scaparra and Church (2008a). Bounds are developed to reduce the size of the model and three heuristics are developed to solve the problem. Results show that when r is random it is important to model it as such. Aksen et al. (2010) study the RIMF with a budget constraint on the fortification resources. They also allow facilities to purchase extra capacity prior to an incident to accommodate customers who migrate from another failed facility. This is termed ‘flexible capacity’. They present a bilevel MIP model with added closest assignment constraints. The model is solved using an implicit enumeration (IE) algorithm. Dong et al. (2009) study a modified version of the RIMQF where the objective is to maximize the worst case minimal time satisfaction over all customers. The time satisfaction for a customer is assumed to be a linear, convex, or concave function of the distance to its assigned facility (Ma and Wu, 2006). They show that accounting for time satisfaction in the objective function results in significantly different solutions.

Medal et al. (2011c) addresses the problem of hardening facilities with the objective of minimizing the maximum worst case consequence over all demand points, the same objective used in Medal et al. (2011a). An MIP formulation is presented as well as a exact algorithm based on the location set covering algorithm. Findings indicate that this objective is not only realistic, but also is much more tractable than considering the minimization of the worst case total consequence, as in the RIMF models mentioned earlier in this section.

Vulnerability Reduction O’Hanley et al. (2007b) also consider a worst-case version of the maximum expected covering location problem with hardening (MEXCLPH). Rather than minimizing the expected species loss, the objective is to minimize the worst case species loss. Like the model in O’Hanley et al. (2007a), the interdiction budget is in the form of a constraint on the probability of the occurrence of the disruption. Again, a bilevel MIP model is presented.

2.6.4 Complex Networks

In this section we mention papers that have considered the risk reduction of complex networks. The increased difficulty observed when going from unreliable facility networks to unreliable complex networks, mentioned in Section 2.5.3, is also present in risk reduction problems, as will be observed in the rest of this section.

2.6.4.1 Expected Value Risk Measure

Failure Probability Reduction Peeta et al. (2010) present a two-stage stochastic programming model for reducing the risk of contingency transportation networks with bridges that are prone to failure. A discrete failure probability reduction approach is presented that reduces the failure probabilities for bridges in the network. The recourse problem is a capacitated minimum cost network flow problem. The Taylor series expansion of the objective function is used to reformulate it as a multi-linear function and Sample Average Approximation approach is used to solve the reformulated model.

Vulnerability Reduction Liu et al. (2009) present a two-stage stochastic programming model for the problem of hardening bridges within a contingency transportation network discussed in Peeta et al. (2010) with the objective of minimizing the expected travel time. Because of their assumption that the travel time for an arc depends on the flow through that arc, they model the second stage problem as a (nonlinear) convex multicommodity flow problem. To account for the nonlinear second stage, they use an extension of the L-shaped method that utilizes the concepts of Generalized Benders' Decomposition, which is well-suited for nonlinear problems.

Redundancy Wallace (1987a) considers the problem of increasing the capacity of arcs in a network with the objective of maximizing the expected maximum flow subject to random failures. It is demonstrated that this problem can be formulated as a two-stage stochastic program with network recourse, for which specialized solution approaches exist (see Birge and Louveaux (1997)).

2.6.4.2 Conditional Expected Value Risk Measure

Vulnerability Reduction Ramirez-Marquez et al. (2009) present a model for protecting a network against an attacker that distributes his resources evenly among all elements. The recourse objective is flow maximization and the overall objective is to maximize the expected max flow. The defender chooses a subset of arcs to defend and then distributes his resources evenly amongst them. The vulnerability of each arc is modeled using a contest success function. Since the attacker allocates a positive amount to each arc, any unprotected arc is completely failed. To solve their model, an evolutionary algorithm is used to identify protection allocation solutions and Monte Carlo simulation is used to evaluate candidate solutions.

Multiple Strategies Holmgren et al. (2007) present a model that includes protection as well as restoration as strategies to reduce the risk to an electrical power grid. The recourse problem is a time-dependent maximum flow problem that captures the time to restore the network after a disruption. Thus, the consequence of a disruptive event is a function of its duration. In this problem, the vulnerability of an element is a function of the defender's allocation of protection resources. The defender may also allocate resources to recovery, affecting the repair time. A tradeoff is made between these two options. Three different attacker strategies are examined: 1) maximize expected negative consequences, 2) maximize the probability that a negative consequence is above a threshold, and 3) choose targets randomly. The model is used to generate the best protection strategy for each attack scenario. However, the authors do not suggest a way to generate protection strategies that perform well against several attack scenarios. The approach is demonstrated on a Swedish power network.

2.6.4.3 Worst Case Risk Measure

Vulnerability Reduction San Martin (2007) provides a specialized formulation and algorithm for the shortest path r -interdiction problem with q -fortification (SPRIG). Computational results show nested and reformulation-based decomposition algorithms to be twice as fast as direct decomposition. Cappanera and Scaparra (2011) study the problem of defending a shortest path net-

work as a hardening problem considering a strategic attacker. They reformulate the hardening action as an attack cost increase action. An implicit enumeration procedure is suggested. Scaparra and Cappanera (2005) suggest a max covering formulation for this problem like that presented in Scaparra and Church (2008a) (see Section 2.6.3.2). They also propose the same procedures for a max-flow problem hardening-interdiction problem.

Bier et al. (2007b) study the problem of defending a power network against a strategic attacker. When choosing a new element to attack, the attacker chooses the arc with the largest load. The defender and attacker are subject to a constraint on the maximum number of hardened edges and attacked edges, respectively. The defender's recourse problem is to minimize the total cost of distribution (load generation) and unmet demand (load shedding). A greedy algorithm is presented where the recourse problem, the attacker's problem, and the defender's problem are solved sequentially in a loop for a pre-specified number of iterations. In the attacker phase, the element with the largest flow is interdicted. In the defense phase, the defender hardens the elements that are most desirable to the attacker. The algorithm is demonstrated on the IEEE reliability test system one and two area networks (Grigg et al., 1999). Yao et al. (2007) present an exact algorithm for a similar problem to that studied in Bier et al. (2007b). They use a delayed cut generation approach similar to Benders' decomposition. They also test their approach on the one area network used in Bier et al. (2007b).

Increase Attack Cost Qiao et al. (2007) study the problem of allocating resources to a water supply network that is subject to an adversarial attack. The resource allocation increases the cost an attacker incurs to attack an element. They develop a model that maximizes the minimal value of a risk metric over a set of element groups. The risk metric for an element group, which the authors define as resilience, is defined as the cost incurred by the attacker to attack the element group divided by the consequence of the disruptive event associated with that element group (this definition is different than the one in Section 2.2.1). The set of element groups considered is the set of all subsets less than a predetermined maximum cardinality, which is the maximum number of arcs that an attacker may attack simultaneously. Due to the hydraulic constraints inherent in a

water supply network, a simulation model is used to estimate the consequence of a component's failure. A genetic algorithm is used to solve the model.

Multiple Strategies Brown (2005) presents a time-indexed model for hardening and expanding the capacity of the links of an oil pipeline network against a strategic attack. The recourse problem for each time period is essentially a maximum flow problem. In addition, attacks are also time-indexed.

2.6.4.4 Survivability Risk Measure

Zhao and Xu (2009) study the effect that the adding of edges has on increasing the survivability of scale-free networks. Survivability is defined as the number of node removals that a network can endure before it becomes disconnected. Two types of node removals are analyzed: random removals and removals of the nodes with the highest degree.

Another line of research deals with allocating spare capacity resources to a network to ensure its survivability in the presence of failures (Ambs et al., 2000; Veerasamy et al., 1999; Balakrishnan et al., 2001, 2002). Problems in this area have been typically modeled as an MIP model with the objective of minimizing the cost of spare capacity allocation subject to a constraint requiring that enough spare capacity exists so that flow can be routed in single-edge failure scenarios.

2.6.4.5 Robustness Risk Measure

The studies described in this section related to robustness share a common organization. First, they usually study some common network topology model, such as the random networks first studied in Erdos and Renyi (1959) and the scale-free networks first studied in Barabasi and Albert (1999). Second, they usually define robustness as the effect that node removals have on the networks. Nodes are either removed randomly or according to a heuristic rule such as highest node degree. The effect of node removals is measured using some metric of connectivity (Costa, 2004; Beygelzimer et al., 2005; Morehead and Noore, 2007) or metric related to the shortest path distances between nodes (Beygelzimer et al., 2005). Third, these studies seek examine the benefit of various risk reduction strategies such as adding additional edges to the network (Beygelzimer et al., 2005;

Costa, 2004; Morehead and Noore, 2007), called augmentation, and rearranging the placement of existing edges, called rewiring (Beygelzimer et al., 2005). Augmentation and rewiring are done randomly or according to a heuristic rule.

2.6.4.6 Risk Metric

Cunningham (1985) considers a risk metric called the ‘strength’ of a network, which is a measure of the cost of edge removals over the number of disconnected subgraphs resulting from the edge removals. Two models are presented: one in which a defender maximizes the strength of the graph by increasing the edge attack costs subject to a budget constraint and another where the defender minimizes cost subject to a lower bound on the strength of the graph. In addition to edge removals, the problem of removing nodes is also considered.

2.6.4.7 Multiple Risk Measures

Schavland et al. (2009) consider both hardening and component capacity increases in protecting a network against an attacker using a multiobjective game theoretic model. The two objectives are to maximize the two-terminal reliability as well as the worst case expected maximum flow.

2.6.5 Future Work

In this section we discuss areas of future work that relate specifically to the risk reduction of networks. To start, we mention the gaps that we found in the literature. We did not find any categories that we immediately deemed not worthy of consideration. Hence, in our opinion all of the categories that did not have any papers in this review are worthy of some further thought.

For the ‘Set of elements’ and ‘Simple’ networks, the only risk measure used among the papers in our review is the conditional expected value risk measure. This is because the research on these types of networks has focused on vulnerability without trying to estimate the likelihood of an incident. It may be interesting to consider incident likelihoods for this type of network. Considering facility networks with facility failures, the only risk measures considered were expected value and worst case. We believe that others are worthy of further consideration although some of them (esp. conditional expected value) will probably add complexity to the problem. We did not find any

work on facility networks with arc failures yet we view this category to be a relevant one. Finally, although there are papers studying complex networks for each of the risk measures, a few of the categories have only one paper. These are conditional expected value, risk metric, and multiple measures. There is probably more room for further study in these areas.

Next, we mention imbalances that we found in the number of papers within each of our categories. Table 2.2 shows the number of papers studying various combinations of the network types and risk reduction strategies. The network types included in the table are those mentioned in this section and the strategies included are vulnerability reduction (VUL), likelihood reduction (LI), element consequence reduction (ECR), failure probability reduction (FP), redundancy (RED), restoration capacity (REST), rewiring (REW), increasing the attacker's cost (INCR), and informational measures (INFO). The last two columns represents papers that consider multiple approaches (MULT), or a tradeoff (TRADE) between multiple approaches, respectively. The statistics show that vulnerability reduction is the strategy of choice for a majority of papers. Many of the other strategies have received little attention. Because these strategies are viable for most problems, they are deserving of more study. The consideration of tradeoffs between multiple strategies has received little attention outside of the study of a set of elements. Because most decision makers have several strategies to consider when trying to reduce the risk of a network, tradeoff studies are an important area of future work, especially for more complex types of networks.

Thus far, the fortification models have considered less types of networks than design models. Network types that have not received any attention include facility networks with unreliable facilities and multi-echelon supply chain networks.

2.7 Conclusions

In this review we discussed networks that are subject to disruptions. The focus was mainly on how to reduce their risk to disruptions via design and via risk reduction strategies such as hardening. We also briefly discussed descriptive models, which seek to assess the vulnerability and risk of networks with respect to disruptions. We observed that the study of networks under disruption risk is an important area of research, with many authors demonstrating that considering risk in the

Table 2.2: Qty. of papers studying different network types and risk reduction strategies.

	Primary risk reduction strategy											Sum	%
	VUL	LI	ECR	FP	RED	REST	REW	INCR	INFO	MULT	TRADE		
Set of elements	6	0	0	0	0	0	0	0	9	0	5	14	45%
Simple network	5	0	0	0	0	0	0	0	3	0	0	3	10%
Facility net.- fac. fail	10	0	0	2	0	0	0	0	0	0	0	2	6%
Facility net.- arcs fail	0	0	0	0	0	0	0	0	0	0	0	0	0%
Complex net.	8	0	0	1	5	0	1	2	0	3	0	12	39%
Total	29	0	0	3	5	0	1	2	12	3	5		
%	94%	0%	0%	10%	16%	0%	3%	6%	39%	10%	16%		

design process or implementing risk-reduction strategies can have a substantial impact on reducing disruption risk. This is also a rich area of research, adding the consideration of risk to problems that are already considered to be hard. We also observed that this research area has grown a lot in the last ten years, attracting many researchers. These problems have been addressed by researchers from the fields of economics, industrial engineering/operations research, civil engineering, physics, geography, computer science, and business, among others. During the course of our discussion we classified the literature and pointed out various areas of future work. In the next section, we suggest areas of future research that are broader in nature, focusing on extensions to the models in this survey.

2.7.1 Future Work: Imbalances

In this section we discuss imbalances in the number of papers studying each category. Considering facility networks with facility failures, Table 2.3 gives statistics on the number of papers studying various facility location measures and risk measures. The risk measures included are expected value (EV) and worst case (WC). Once again, we only counted the papers that were described in this paper, excluding from our count the topics that have been surveyed previously. The table shows that a majority of papers have considered distance-related recourse objectives. While distance is an important measure for commercial applications, other objectives such as coverage may be more applicable to public sector applications such as disaster relief. Additionally, one can observe from this table that the design papers typically consider the expected value risk measure and the risk reduction papers usually consider the worst case. In our opinion, expected value and worst case, as well as other risk measures are relevant in both design and risk reduction.

Also, we found that only one paper in his paper considered capacitated facilities, Gade and Pohl (2009). When capacity is not considered, when a disruption occurs, demand points can always be allocated to their closest non-disrupted facility. However, when capacity is considered, the problem of allocating demand points to facilities is more complicated. As a result, capacitated models may produce significantly different solutions than their uncapacitated counterparts.

Table 2.3: Qty. of papers for recourse objective and risk measure combinations for facility networks with facility failures.

Recourse Objective	Design			Risk Reduction			Total	%
	EV	WC	Total	EV	WC	Total		
Distance Related	13	5	18	2	8	10	28	76%
Distance with Inventory Coverage	2	0	2	0	0	0	2	5%
	3	2	5	1	1	2	7	19%
Total	18	7	25	3	9	12		
	49%	19%	68%	8%	24%	32%		

2.7.2 Future Work: Big Picture

Here we mention items of future work that either span both sections 2.5 and 2.6 or involve other topics. In addition to this section, we also recommend the future work sections in Snyder et al. (2006) and Snyder et al. (2010).

Many of the models in this review consider random incidents. The drawback of the random incident approach is that its results are dependent on likelihood and vulnerability information. As a result, it would be useful to know how sensitive these models are to the likelihood and vulnerability estimates. If these models are indeed sensitive to their inputs, it would be useful to have models that produce solutions that are robust to likelihood and vulnerability inputs.

There has been a considerable amount of work done in the risk analysis community in developing ways to assess the vulnerability and risk of infrastructures. However, these approaches are usually qualitative, as opposed to using the mathematical models mentioned in this survey. Thus, integrating the work done in risk analysis and quantitative mathematical modeling may prove to be fruitful. Also, in risk analysis and vulnerability assessment, the mitigation is done after the risk assessment, in sequence. However, since mitigation activities change the risk assessment, these two steps should be integrated.

The papers that study risk-reduction of an independent set of elements in Section 2.6.1 relax many of the assumptions made in most of the models for more complex networks. Some of the relaxed assumptions include: 1) the attacker's resources are known with certainty to the defender, and 2) attacks are successful 100% of the time, and 3) elements are either completely protected or

not protected at all. While Liberatore et al. (2011) addresses (1) and several interdiction models address (2), it would be useful if these assumptions were relaxed in more design and risk-reduction models. The papers in Section 2.6.1 also simultaneously consider multiple measures, another aspect that would be useful to consider in more complex networks. Of particular importance is the simultaneous consideration of random failures and strategic attacks within a design or risk reduction model. This is because it is often the case that networked infrastructures are vulnerable to multiple types of hazards.

Only a few papers have considered the case where design and risk reduction decisions are made simultaneously. In the context of locating unreliable facilities, Snyder and Daskin (2005) and Cui et al. (2011) include perfectly reliable locations in their models, which can be thought of as ‘fortified’ facilities. However, it is determined exogenously, or prior to solving the model, which facilities are perfectly reliable and hence risk reduction is not an output of the model. Lim et al. (2010a) present a model where the decision maker chooses between locating unreliable facilities and reliable ‘backup’ facilities, at a higher cost. However, their model assumes that if a demand point’s primary facility fails, the demand point is then assigned to its perfectly reliable backup. There are two ways in which this assumption may not hold in reality. First, it is likely that if a demand point’s primary facility fails it will then be assigned to the next closest open facility, rather than going directly to a reliable backup. Second, this assumption allows for a facility to be assigned as one demand point’s primary facility and another demand point’s backup. In some situations this may not be a satisfactory assumption, especially if the demand points require the same commodity type. Thus, this paper takes an approach to risk reduction that is somewhat different to the papers mentioned in Section 2.6. Medal et al. (2011c) have developed a model that integrates facility location and facility fortification decisions. The objective is to minimize the maximum worst case disruption consequence over all demand points. An MIP model as well as an exact set covering based algorithm are presented. To our knowledge, there has been no other work on integrating design and risk reduction decisions. Due to the lack of work in this area, it is an important area of future work.

Also, most of the papers in this survey assume a single decision maker such as a private company. However, the management of most public infrastructures involves multiple stakeholders. It is possible that the decisions generated by single-decision-maker models are not satisfactory for all of the stakeholders involved in public critical infrastructure. One way of addressing this is by developing models that generate risk equitable solutions. Another way is to develop models that explicitly account for the multiple stakeholders.

Most of the studies included in this review paper assume risk to be static. In reality, risks change over time. Therefore, it would be useful to have models that helped decision makers make strategic decisions to mitigate against time-varying risks.

Further, distance-related and connectivity-related recourse objectives have thus far been studied by researchers from different backgrounds and with different applications. Distance-related recourse objectives are popular in supply chains and connectivity-related objectives are popular in communication networks. However, connectivity-related objectives are also appropriate for supply chains because they can be a proxy for customer service; i.e., when the network becomes disconnected, it usually is unable to serve some of its customers. As a result, it would be useful for models to integrate distance-related and connectivity-related objectives.

Acknowledgments

We would like to thank the United States Department of Homeland Security (DHS) who sponsored this work through the Mack-Blackwell Rural Transportation Center at the University of Arkansas, a National Transportation Security Center of Excellence, on grant number DHS-1101. However, the views expressed in this chapter do not represent those of DHS, but rather those of the authors. We would also like to thank Brittini King, for her assistance on this project.

3 Overview

The theme of this dissertation is to develop models for strategic decision-making in preparation for a network disruption. Two types of strategic decisions are studied: location, i.e., where to locate supply facilities and how many to locate, and facility protection, i.e., how to allocate protection resources among a set of supply facilities. The commonality between these two decisions is that they both consider how to make strategic decisions in preparation for a disruption. In particular, the goal of this dissertation is to understand the tradeoff between three objectives: (1) cost of strategic decisions, e.g., building and protecting facilities, (2) efficiency of the network, e.g., operating cost, and (3) vulnerability of the network, e.g., operating cost after a disruption.

3.1 Gaps in the Literature

As illustrated in Chapter 2, there are still important areas of research concerning designing and protecting networks subject to disruptions. The following are the important remaining areas that are addressed in this dissertation.

1. **Maximum distance objective with the worst-case risk measure.** Most of the literature on facilities has considered the objective of minimizing total distance. While the total distance objective is appropriate for situations where cost minimization is the goal (equivalently, profit maximization), it is not as useful for public-sector applications, in which other objectives such as equity are more important. Thus, it is important to study other objectives such as the maximum distance in order to help public sector decision-makers make better decisions.
2. **Integrated location and protection.** There are currently no exact solution methods for solving the integrated problem of simultaneously choosing a facility location and protection plan. Admittedly, many times in practice the facility location and protection decisions will naturally be made at different times or by different groups. However, if a decision-maker has the opportunity to make these decisions at the same time, it would be useful for them to have

a model that integrates the two decisions together. Also, it would be useful for a decision-maker to know the benefit of making these two decisions together, rather than separately.

3. **Imperfect protection.** A majority of the models in the literature assume binary, perfect facility protection. That is, a facility is either fully protected or fully unprotected. Further, a fully protected facility can never fail. While this may be a good approximation of reality in some cases, there may be some cases where protection is far from perfect and therefore should not be modeled as such. Models of imperfect protection should help decision-makers make better decisions about how to protect facilities.
4. **Multi-state capacity.** A majority of models in the literature assume that after a facility is exposed to a hazard (e.g., tornado, bomb, etc.) it is either completely inoperable or completely operable. While this may be a good approximation in some cases, there may be other cases where facilities can be in a partially degraded state. In these cases, multi-state capacity models may provide decision-makers with better solutions.

There are several reasons why these remaining areas are not straightforward to address:

1. **Maximum distance objective with the worst-case risk measure.** Typically, models for facility location or protection subject to facility disruptions with the worst-case risk measure are often modeled as two-level defender-attacker problems. Defender-attacker problems for the total distance objective have been shown to be computationally demanding for large problems (Scaparra and Church, 2008a,b). Without disruptions, the total distance objective is easier to solve than the maximum distance objective (Snyder, 2006). Thus, it appears that the maximum distance objective will be especially hard to solve if facility disruptions are included.
2. **Integrated location and protection.** It was already mentioned that defender-attacker problems are computationally demanding, even if only facility location is considered (Scaparra and Church, 2008a,b). Thus, one might expect that integrating protection decisions would make the problem even more difficult.

3. **Imperfect protection.** In most protection models, the likelihood of element failure is a function of the first-stage protection decisions. Under perfect protection, this probability function is simple: the probability of failure is 0 if the element is protected and q (the nominal probability of failure) otherwise. If the protection is imperfect this probability function needs to be more detailed, often making the model more complicated and computationally taxing.
4. **Multi-state capacity.** Uncapacitated facility location problems with disruptions often use multiple-assignment formulations that rely on the fact that facilities have infinite capacity and thus a demand point can be assigned to a facility as long as it has not failed. These multiple-assignment formulations allow the problem to be formulated as a single-stage problem (Snyder and Daskin, 2005; Cui et al., 2011), even though the problem can be thought of as a two-stage problem. However, their capacitated counterparts cannot be formulated using this multiple-assignment paradigm. Thus, capacitated facility location problems with disruptions have been solved using two-stage stochastic programming models (Gade and Pohl, 2009) in which the total number of scenarios is 2^n , where n is the number of facilities. Thus, when multiple capacity states are introduced, there are k^n , where k is the number of capacity states. Thus, including a large number of capacity states can result in a very large number of scenarios, making the stochastic program difficult to solve.

These challenges are overcome in this dissertation by the following strategies.

1. **Maximum distance objective with the worst-case risk measure.** This challenge is overcome by using problem structure to reformulate what would normally be a three-stage problem (defender-attacker-operator) as a single-stage problem.
2. **Integrated location and protection.** This challenge is also overcome because it can be modeled as a single-stage problem rather than as a three-stage problem.
3. **Imperfect protection.** This challenge is overcome by using a discrete approximation to simplify the model and improve tractability.

4. **Multi-state capacity.** Because of the structure of this problem, the first-stage variables affect only the scenario probabilities and do not affect the second-stage objective function. Thus, the second-stage problems for each scenario can be solved apriori, making the solution approach much simpler.

This dissertation describes the development of models and appropriate solution techniques needed to address these remaining areas of research. The following section is a justification of the modeling assumptions that were made.

3.2 Modeling Assumptions

As mentioned in Chapter 2, there are several characteristics of a network disruptions model. The following paragraphs describe how the models in this dissertation can be classified and why particular modeling assumptions were made.

First, the models in this dissertation are network-based. That is, they consist of nodes, arcs, supply points and demand points. This modeling paradigm was chosen because it is useful for modeling networks such as supply chains, transportation systems, and power grids.

Another important modeling decision in a study on a network subject to disruptions is what elements of the network fail. In the literature, models have included disruptions at intermediate nodes, arcs, and/or supply facilities. The models in this dissertation assume that only the supply facilities can be disrupted. This is because supply facility disruptions are harder to overcome than intermediate node or arc disruptions.

An additional modeling consideration is the post-disruption problem, i.e., the problem that must be solved after the disruption occurs. There are many options for the post-disruption problem: the shortest path problem, the maximum flow problem, assignment problems, etc. This dissertation utilizes two versions of the assignment problem: capacitated assignment and uncapacitated assignment. In particular, the problem is to assign the demand originating from a set of spatially-dispersed customer locations to a set of capacitated (uncapacitated) supply facilities. The assignment problem is a logical choice in the context of disruptions because after a disruption decision-makers often have to choose how to modify their procurement plan in light of the disrup-

tion of one or more of their suppliers. As the experiments in Chapter 6 indicate, it is important to consider capacity when modeling a network that is subject to disruptions.

A further characteristic of a network disruptions model is the cause or origin of disruptions. Generally, disruptions are assumed to occur randomly or to be selected intelligently. Under random disruptions, typically used to model natural disasters, the location and magnitude of disruptions is unknown to the decision-maker. Thus, these disruptions are considered to be exogenous to the model and are often modeled using a probability distribution. In contrast, under intelligent attacks, the location and magnitude of disruptions is selected by an intelligent adversary who usually wants to maximally degrade the performance of the network. Thus, these attacks are considered to be endogenous to the model. Both natural disasters and terrorist attacks are a real threat to networks such as supply chains and transportation networks. This dissertation considers both random disruptions and intelligent attacks.

Another important modeling consideration is the type of risk measure used. To model a risk-neutral decision-maker, the expected-value measure is used, which plans for the average-case impact of a disruption. To model a risk-averse decision-maker, the worst-case risk measure can be used, which generates decisions that plan for the worst-case impact of a disruption. Another class of risk measures are quantile-based measures, which maximizes the probability that the disruption impact is less than a threshold. This dissertation uses both the expected-value measure and the worst-case measure. However, the solution method used in Chapter 6 can be adapted for a quantile-based risk measure.

Finally, another classification is the objective function used to represent network performance. Common objectives for facility location problems include: (1) minimizing the cost of constructing facilities (i.e., the location set-cover problem; see Toregas et al. (1971); Toregas and ReVelle (1972)), (2) minimizing total distance from demand points to supply facilities (i.e., the p -median problem; see Hakimi (1964)), (3) minimizing the maximum distance from a demand point to its closest supply facility (i.e., the p -center problem; see Hakimi (1965)), and (4) maximizing the amount of demand serviced (i.e., the maximum-covering location problem; see Church and ReV-

elle (1974)).¹ This dissertation considers both (2) and (3). Objective (2) is relevant because it can represent the cost to distribute goods to all customers. Also, this objective can be converted to measure average distance by multiplying by one over the number of facilities. Objective (3) is useful for contexts in which equitable solutions are desired. By minimizing the maximum distance, i.e., the worst customer service, this measure generates solutions that are equitable with respect to the distance a customer has to travel to reach its closest facility.

3.3 Contributions

What follows is a discussion of the contributions of this dissertation.

Chapter 4 investigates the r -All-Neighbor p -Center Problem (RANPCP), which is to locate p facilities that are subject to attacks. The objective is to minimize the post-attack maximum distance from a demand point to its closest operating facility. The worst-case risk measure is used. The goal of this study is to gain further understanding of the tradeoff between (1) system design cost, (2) system performance (without disruptions), and (3) vulnerability. Several new mixed-integer programming (MIP) models are presented and compared to an existing model. In addition, a binary search algorithm is presented that outperforms the MIP models. This chapter makes the following contributions:

1. a new MIP formulation for the RANPCP along with a class of valid inequalities and upper and lower bounds;
2. an empirical study that compares the run time of the new MIP formulation to the run time of existing formulations;
3. an empirical study to estimate the scalability of binary search algorithm;
4. several managerial insights relating to the tradeoff between (1) system design cost, (2) system performance (without disruptions), and (3) vulnerability.

¹A model for Objective (2) can be used to optimize Objective (4) by the following transformation. First, choose a coverage distance δ and define new distances d'_{ij} such that $d'_{ij} = 1$ if $d_{ij} \leq \delta$ and zero otherwise. Second, use a maximization objective rather than minimization.

Chapter 5 studies a problem that integrates network design and network protection decisions. In particular, Chapter 5 extends Chapter 4 by integrating facility hardening decisions along with the facility location decision. Facility hardening is a special case of facility protection in which a hardened facility is immune to disruptions. A natural three-stage formulation is presented for the problem. Next, it is shown that problem structure can be used to reformulate the problem as a single-stage problem. A MIP model is presented for the single-stage problem and a binary search procedure, which outperforms the MIP model, is also presented. The contributions of this chapter are

1. the first attempt to integrate facility location and facility hardening decisions while considering the maximum distance objective;
2. an exact procedure for solving the integrated location-hardening model; and
3. managerial insights involving (1) the benefit of integrating location and hardening decisions and (2) the tradeoff between (a) system design cost, (b) system performance (without disruptions), and (c) vulnerability.

Chapter 6 is a preliminary investigation into a facility protection problem in which several common assumptions are relaxed. The assumptions are: 1) facility protection is perfect, 2) there is only one allocation level, and 3) facilities can only be in two capacity states. The problem is first modeled as a nonlinear integer two-stage stochastic program. Next, problem structure is used to reformulate the model as a linear integer two-stage stochastic program. In this formulation, the second stage network flow can be solved apriori. An implementation of the L-shaped method is then used to solve the stochastic program. In addition, a Local Search algorithm is used to solve a special case of the problem. The contributions of this chapter are

1. the first study of a problem that considers both imperfect protection and capacity;
2. a reformulation that removes the non-linearity from the initial formulation and allows the second-stage problems to be solved apriori;

3. an L-Shaped implementation for the reformulation;
4. preliminary insights regarding which features a model of facilities subject to disruptions should have; and
5. a Local Search algorithm that appears to perform well for a special case of the problem.

Thus, the major contributions of this dissertation are:

1. new mathematical models for several problems involving the location and protection of facilities subject to disruptions;
2. the discovery and identification of appropriate solution techniques for these models;
3. managerial insights relating to the tradeoff between (1) system design cost, (2) system performance (without disruptions), and (3) vulnerability;
4. insights regarding which features a model of facilities subject to disruptions should have.

4 On the R-All-Neighbor P-Center Problem

Abstract

In this paper we consider a generalization of the p -center problem called the r -all-neighbor p -center problem (RANPCP). The objective of the RANPCP is to minimize the maximum distance from a demand point to its r^{th} -closest located facility. The RANPCP is applicable to unreliable facility location and emergency response. The contributions of this paper are 1) the development of a new mixed-integer programming (MIP) formulation for the RANPCP, 2) a thorough empirical analysis of the tractability of the RANPCP, and 3) an empirical analysis that gives several useful insights into the RANPCP. For example, we found that the RANPCP produces solutions that not only minimize vulnerability but also perform reasonably well (43% from optimal, on average) when failures do not occur. In contrast, if disruptions are not considered when locating facilities, the consequence due to facility failures is about 630% higher, on average, than if disruptions had been considered. Thus, our results show the importance of optimizing for vulnerability.

4.1 Introduction

The facility location problem is a fundamental problem that has been studied for a long time by researchers from many different disciplines. Over time, as researchers began to develop location models for specific applications such as locating fire stations and ambulances, location models began to include the unavailability of facilities and vehicles in order to reflect reality. For example, ambulances in large metropolitan areas are very busy and not always available for service.

In response, researchers began developing deterministic facility location models that address facility unavailability by considering backup coverage. Other researchers followed by considering the use of backup coverage as a method for mitigating against facility unavailability caused by terrorist attacks, random failures of facilities, and congestion of servers. As a result, facility location research has developed further to include backup-coverage extensions of the p -median, p -center, set covering problem, and maximal set covering problem.

In this paper we study the r -all-neighbor p -center problem (RANPCP), a generalization the p -center problem. The RANPCP is in a class of problems called bottleneck, or minimax, problems (Hochbaum and Shmoys, 1986). Specifically, this problem seeks to minimize the maximum distance from a demand point to its r^{th} closest located facility. The RANPCP is applicable for facility location systems that require backup coverage because of facility unavailability. Drezner (1987) showed that the problem of locating facilities in order to minimize the damage caused by a strategic interdictor is equivalent to the RANPCP. Also, this problem is applicable for locating emergency vehicles that must respond to events requiring more than one vehicle. The RANPCP minimizes the arrival time of the r^{th} vehicle. While other researchers have developed approximation algorithms and heuristics for this problem, our work is the first study devoted to exact solution methods for the RANPCP.

This paper makes the following contributions:

1. We develop a new mixed-integer programming (MIP) formulation for the RANPCP along with a class of valid inequalities and upper and lower bounds.
2. We perform an empirical study to estimate the scalability of the solution methods used.
3. We give several managerial insights by performing a tradeoff analysis.

4.2 Literature Review

Most of the literature on facility location with backup coverage has focused on the degradation in overall service incurred when some facilities become unavailable and unable to serve customers. Specifically, most of the models involve locating backup facilities to minimize this potential degradation.

Some research has considered that facilities become unavailable because of random causes: natural or man-made disasters, congestion of servers, etc. Drezner (1987) was the first to consider random facility failures in the p -median model and his research was extended by others (Lee, 2001; Snyder and Daskin, 2005; Berman et al., 2007). Snyder and Daskin (2005) and Cui et al. (2011) have modeled facility failures in the fixed-charge location problem. Daskin (1982, 1983)

was among the first to consider random facility unavailability in the maximal covering location problem. His work was subsequently extended by Batta et al. (1989), who explicitly included queuing in their model.

Rather than considering random failures, other research has sought to minimize the worst case degradation in service; in other words, they measure the risk of facility failures by the worst case degradation. We are not aware of any published work on the p -median problem with the worst case risk measure. Several papers have included backup coverage in the set covering model. Including backup coverage in the set covering model ensures adequate coverage in the event of facility unavailability. Van Slyke (1982) was the first to include backup coverage in a set covering model and Church and Gerrard (2003) worked on the location set covering problem with facility failures, in which only one vehicle can be located at a potential location. One of the earliest models involving the maximal covering problem with backups was by Daskin and Stern (1981), who minimized two objectives: the total amount of demand coverage and the number of facilities located. This research has since been followed by others (see Brotcorne et al. (2003) for a survey).

Several authors have described approximation algorithms for a problem called the r -neighbor p -center problem (RNPCP). The RNPCP is an extension of the p -center problem in which the problem is to locate p facilities amongst a set of nodes in order to minimize the maximum distance from a client node, defined as a node that does not have a facility located on it, to its r^{th} closest located facility. Krumke (1995) developed a 4-approximation algorithm¹ for this problem. Chaudhuri et al. (1998) and Khuller et al. (2000) independently developed different 2-approximation algorithms for the r -neighbor p -center problem and show that a better approximation cannot be obtained in polynomial time.

Other authors have studied a version of the r -neighbor p -center problem in which all nodes are client nodes, rather than defining a client node as a node that does not have a facility located on it. Drezner (1987) called this problem the (p, r) -center problem and described a heuristic algorithm for the version where facilities can be located anywhere in a plane. Khuller et al. (2000) named this

¹An α -approximation algorithm is an algorithm that is guaranteed to find a solution with an objective function value of no more than α times the optimal objective value.

problem the r -all-neighbor p -center problem (RANPCP) and provided approximation algorithms that guarantee an approximation factor of 3 and if $r < 4$, an approximation factor of 2.

Elloumi et al. (2004) presented a new model and an exact solution method for the p -center problem (PCP). They mention that their model and solution method can also be used to solve the RANPCP. They find that the LP relaxation bound of their model is at least as good as that of the standard p -center MIP model (see Daskin (1995)). They also show that in many cases their bound is strictly better. They relax a subset of the integer variables in their model to find a tight lower bound for the PCP. They also show that their tight lower bound can be computed by solving a polynomial number of linear programs within a binary search algorithm. They show that their lower bound is at least $1/3$ of the optimal objective when the distances obey the triangle inequality and at least $1/2$ of the optimal objective when distances are symmetric. However, they do not prove that the approximation factors for their bounds are valid for the RANPCP. Their computational results show when they incorporate their lower bound into the standard binary search algorithm for the PCP, the binary search algorithm is able to solve PCP instances of up to 1817 nodes. Because Elloumi et al. (2004) focused on the PCP, they leave out some of the details needed to extend their lower bound to the RANPCP and only present empirical results for the PCP.

Our work is the first to do in-depth analysis of the tractability of the RANPCP. Further, our work contains the first analysis of managerial insights related to the RANPCP such as how the RANPCP solutions perform under other objectives and how different objectives trade off against each other. Our solution methods and analysis make the following contributions. First, we introduce MIP formulations for the RANPCP that are different than the MIP formulation presented by Elloumi et al. (2004). We found that our MIP models are competitive with the MIP model from Elloumi et al. (2004) in terms of runtime. Second, we study the version of the RANPCP in which nodes have weights. The approximation factor guarantees for the approximation algorithms described above (Krumke, 1995; Chaudhuri et al., 1998; Khuller et al., 2000) are only valid for the unweighted RANPCP. Further, Elloumi et al. (2004) do not analyze the weighted version of the RANPCP, although their solution method could be used for the weighted version also.

4.3 Problem Description and Model

The r -all-neighbor p -center problem (RANPCP) can be defined as

locate p facilities amongst a set of candidate locations in order to minimize the maximum distance from a demand point to its r^{th} closest facility.

The RANPCP can be mathematically stated as follows. Let N be a set of points, $\mathcal{F} \subseteq N$ be a set of facilities and $\mathcal{D} \subseteq N$ be a set of demand points. Let d_{ij} be the cost incurred when facility $i \in \mathcal{F}$ serves demand point $j \in \mathcal{D}$. Because our solution methods are still valid if they are applied to a problem instance whose distances do not obey the triangle inequality, we could let $d_{ij} = h_j d'_{ij}$, where d'_{ij} is the distance from i to j and h_j is the weight of demand point j . For simplicity, in this paper we refer to d_{ij} as the distance from i to j . Let $D_j^r(X)$ be the distance from demand point j to its r^{th} closest located facility when a set of facilities $X \subseteq \mathcal{F}$ are located. The RANPCP requires that $|X| \leq p$, the number of facilities that may be located. The RANPCP can be stated as:

$$\begin{aligned} \min_{X \subseteq \mathcal{F}} \quad & \max_{j \in \mathcal{D}} D_j^r(X) \\ & |X| \leq p \end{aligned} \tag{4.1}$$

Drezner (1987) modeled the situation in which an interdictor seeks to destroy r facilities in order to maximize the maximum post-interdiction distance from a demand point to its closest available facility. He called this model the (p, r) -center problem. He noticed that the interdictor's optimal strategy is to choose a demand point and interdict the r closest located facilities to that demand point. Thus, the (p, r) -center problem is equivalent to the $(r + 1)$ -all-neighbor p -center problem.

In an optimal solution to the RANPCP, each demand point is covered by at least r facilities, meaning that each demand point is within U^* distance units of r facilities, where U^* is the optimal maximum distance. Thus, the parameter r can represent either the number of covers or the number of neighbors required by each demand point.

The RANPCP model has some relationship to several concepts in risk assessment. First, the consequence modeled in the RANPCP is the increase in the maximum distance from a demand point to its closest facility when $(r - 1)$ facility disruptions have occurred. The RANPCP does not consider the likelihood of a facility disruption event; rather, it models the situation in which a facility disruption event has occurred. Further, in the RANPCP model the vulnerability of facilities is complete. That is, if a facility is affected by an event such as a natural disaster or attack, the facility is completely inoperable. Thus, the objective of the RANPCP is to minimize the worst case consequence.

In the rest of this section we describe our MIP model.

4.3.1 Three Index Formulation

First, we present a straightforward formulation of the RANPCP. In this formulation we keep track of the corresponding level for each demand-facility pair. The ‘level’ at which a facility is assigned to a demand point is simply the distance rank of that facility in relation to the other located facilities.

Variables

- Y_i is 1 if a facility is located at i and 0 otherwise.
- $X_{ij\ell}$ is 1 if the facility located at i is assigned to demand point j and i is the ℓ^{th} closest located facility to j .

$$(M1) \quad \min U \quad (4.2a)$$

$$\text{s.t.} \quad \sum_{i \in \mathcal{I}} d_{ij} X_{ijr} \leq U \quad \forall j \in \mathcal{J} \quad (4.2b)$$

$$\sum_{i \in \mathcal{I}} X_{ij\ell} = 1 \quad \forall j \in \mathcal{J}, \ell = 1, \dots, r \quad (4.2c)$$

$$\sum_{\ell=1}^r X_{ij\ell} \leq 1 \quad \forall i \in \mathcal{I}, j \in \mathcal{J} \quad (4.2d)$$

$$d_{ij} X_{ij\ell} \leq d_{i'j} + M_j(1 - X_{i',j,\ell+1}) \quad \forall j \in \mathcal{J}; \ell = 1, \dots, r-1; \quad (4.2e)$$

$$i \neq i' \in \mathcal{I}$$

$$X_{ij\ell} \leq Y_i \quad \forall i \in \mathcal{I}, j \in \mathcal{J}, \ell = 1, \dots, r \quad (4.2f)$$

$$\sum_{i \in \mathcal{I}} Y_i \leq p \quad (4.2g)$$

$$X_{ij\ell} \in \{0, 1\} \quad \forall i \in \mathcal{I}, j \in \mathcal{J}, \ell = 1, \dots, r \quad (4.2h)$$

$$Y_i \in \{0, 1\} \quad \forall i \in \mathcal{I} \quad (4.2i)$$

Constraints (4.2b), in conjunction with the minimization objective in (4.2a), ensure that the objective value is equal to the maximum value of the weighted distance between demand points and their r^{th} closest located facility, over all demand points. Constraints (4.2c) require that a demand point be assigned to one facility at each level. Constraints (4.2d) prevent a facility from being assigned to more than one level for a demand point. Constraints (4.2e) enforce an ordering of the levels for each demand point. That is, the facility assigned to demand point j at level ℓ must have a smaller value of d_{ij} than the facility assigned at level $(\ell + 1)$. The constant M_j is assigned a large value such as $\max_{i \neq i' \in \mathcal{I}} |d_{ij} - d_{i'j}|$. Constraints (4.2f) specify that a demand point may only be assigned to a facility i at a level if the facility has been located at i . Constraints (4.2g) place a restriction on the number of facilities that are located. Constraints (4.2h) define binary assignment variables for only the r most desirable levels for each facility and demand point combination. Finally, constraints (4.2i) require the location variables to be binary.

4.3.2 Reformulation of the RANPCP

Unfortunately, Model $M1$ has a large number of assignment variables, $|\mathcal{I}| \times |\mathcal{J}| \times p$ to be exact. In addition, it has a disjunctive constraint (4.2e) for each pair of consecutive pair of levels $(\ell, \ell + 1)$. However, some of the variables in model $M1$ are unnecessary. In finding the optimal solution to the RANPCP it doesn't matter if demand points are assigned to the correct level for levels $\ell < r + 1$, because these assignments are not included in the objective function. The only requirement for the objective function to be computed correctly is that each demand point is assigned to its correct $(r + 1)^{\text{th}}$ level. Thus, it is enough to require that if $X_{ijr} = 1$ and $X_{i'j\ell} = 1$ (with $\ell < r$) then i must be further to j than i' . Thus, many of the disjunctive constraints (4.2e) are unnecessary. We take advantage of this fact in formulating a more compact model.

Variables

- X_{ij} is equal to 1 if the facility located at i is assigned to demand point j as its $(r - 1)^{\text{th}}$ or closer located facility and 0 otherwise.
- Z_{ij} is equal to 1 if the facility located at i is assigned to demand point j as its r^{th} closest located facility and 0 otherwise. Since this variable represents the assignment from a demand point to one of its backup facilities, we call it the 'backup variable'.

Indices

- i_j^ℓ is the ℓ^{th} closest facility to demand point j .

$$(M2) \quad \min \quad U \quad (4.3a)$$

$$\text{s.t.} \quad \sum_{i \in \mathcal{I}} Z_{ij} \leq U \quad \forall j \in \mathcal{J} \quad (4.3b)$$

$$\sum_{i \in \mathcal{I}} Z_{ij} = 1 \quad \forall j \in \mathcal{J} \quad (4.3c)$$

$$\sum_{i \in \mathcal{I}} X_{ij} = r - 1 \quad \forall j \in \mathcal{J} \quad (4.3d)$$

$$d_{ij}X_{ij} \leq d_{i'j} + M_j(1 - Z_{i'j}) \quad \forall j \in \mathcal{J}, i', i \neq i' \in \mathcal{I} \quad (4.3e)$$

$$X_{ij} + Z_{ij} \leq Y_i \quad \forall i \in \mathcal{I}, j \in \mathcal{J} \quad (4.3f)$$

$$\sum_{i \in \mathcal{I}} Y_i \leq p \quad (4.3g)$$

$$X_{ij}, Z_{ij} \in \{0, 1\} \quad \forall i \in \mathcal{I}, j \in \mathcal{J} \quad (4.3h)$$

$$Y_i \in \{0, 1\} \quad \forall i \in \mathcal{I} \quad (4.3i)$$

The objective (4.3a) and constraints (4.3b) serve the same purpose as in model *M1*. Constraints (4.3c) require that a demand point be assigned to exactly one facility at level r . Constraints (4.3d) ensure that $r - 1$ facilities are assigned to levels $r - 1$ or lower. Constraints (4.3e) enforce an ordering of the levels for each demand point. That is, the facilities assigned to demand point j at levels 1 through $(r - 1)$ must have a smaller value of d_{ij} than the facility assigned at level r . The constant M_j is the same as in *M1*. When two facilities have the same distance to a demand point, the following constraints should be used:

$$d_{ij}X_{ij} < d_{i'j} + \varepsilon + M_j(1 - Z_{i'j}) \quad \forall j \in \mathcal{J}, i', i \neq i' \in \mathcal{I} \quad (4.4)$$

The quantity ε should take a value less than the minimum absolute difference between two values of d_{ij} . Constraints (4.3f) specify that a demand point may only be assigned to a facility i at a level if the facility has been located at i . Constraints (4.3g) place a restriction on the number of facilities located. Constraints (4.3h) define binary assignment variables for each facility and demand point combination. Finally, constraints (4.3i) require the location variables to be binary.

One may notice that in $M2$, some of the Z_{ij} variables will be 0 in an optimal solution. In particular, for a given j , Z_{ij} will be zero for all facilities closer than the r^{th} facility. To explain this formally we first need to introduce further notation. Let i_j^ℓ be the ℓ^{th} closest facility to demand point j .

Now we state our observation in the form of a remark:

Remark 3. There exists an optimal solution to model $M2$ with $Z_{i_j^\ell j} = 0$ for all $1 \leq \ell \leq r-1$ and for all $j \in \mathcal{J}$.

Proof. (By contradiction.) Suppose there exists an ℓ ($1 \leq \ell \leq r-1$) such that in the optimal solution to model $M2$, $Z_{i_j^\ell j} = 1$ for some $j \in \mathcal{J}$. As a result, $\sum_{1 \leq \ell' \leq r-1} Z_{i_j^{\ell'} j} < r-1$ and by constraints (4.3e), $\sum_{r-1 < \ell' \leq |\mathcal{I}|} Z_{i_j^{\ell'} j} = 1$. Hence, there exists an ℓ' ($r-1 < \ell' \leq |\mathcal{I}|$) such that $Z_{i_j^{\ell'} j} = 1$.

Case 1: All of the values of d_{ij} are different.

By our choice of ℓ and ℓ' , $d_{i_j^\ell j} < d_{i_j^{\ell'} j}$. As a result, constraint (4.3e) is violated for j if $Z_{i_j^\ell j} = 1$ and $Z_{i_j^{\ell'} j} = 1$.

Case 2: There exists $i, i' \in I$ such that for some $j \in \mathcal{J}$, $d_{i'j} = d_{ij}$.

By our choice of ℓ and ℓ' , $d_{i_j^\ell j} \leq d_{i_j^{\ell'} j}$. As a result, constraint is (4.4) violated for j if $Z_{i_j^\ell j} = 1$ and $Z_{i_j^{\ell'} j} = 1$. \square

Because of Remark 3, all variables $Z_{i_j^\ell j}$ for all $1 \leq \ell \leq r-1$ and all $j \in \mathcal{J}$ can be removed from Model $M2$. We denote the new model that is formed by removing variables from model $M2$ as model $M2 - C$.

The linear programming (LP) relaxation of $M2 - C$ can be tightened by adding the following constraints:

$$rZ_{i_j^\ell j} \leq \sum_{\ell'=1}^{\ell-1} Y_{i_j^{\ell'} j} \quad \forall j \in \mathcal{J}; r \leq \ell \leq |\mathcal{I}|. \quad (4.5)$$

These constraints require that for a given demand point j , if its ℓ^{th} closest facility, i_j^ℓ , is chosen as its safe facility (i.e., $Z_{i_j^\ell j} = 1$), then r facilities must be located that are closer to j than i_j^ℓ (i.e.,

$\sum_{\ell=1}^{\ell-1} Y_{i_j^\ell} = r$). We include these tightening constraints in all of our experimentation using Model $M2 - C$.

4.4 Binary Search Algorithm

As an alternative to solving the above MIP models using branch and bound, we can also solve the RANPCP using a binary search algorithm similar to the one used to solve the p -center problem (Daskin, 1995). The binary search algorithm for the p -center problem solves a series of location set cover problems to find the optimal maximum distance. Our binary search algorithm for the RANPCP uses the multi-set-cover location problem (MSCLP) (Church and Gerrard, 2003) in place of the location set cover problems. The multi-set-cover location problem modifies the LSCP because it requires that each demand point be covered by at least ℓ facilities, rather than 1. The MSCLP is modeled as follows:

$$\text{(MSCLP}(\delta)\text{)} \quad \min \sum_{i \in \mathcal{I}} Y_i \quad (4.6a)$$

$$\text{s.t.} \quad \sum_{i \in \{i: d_{ij} \leq \delta\}} Y_i \geq r \quad \forall j \in \mathcal{J} \quad (4.6b)$$

$$Y_i \in \{0, 1\} \quad \forall i \in \mathcal{I} \quad (4.6c)$$

The RANPCP can be solved using the following binary search algorithm:

Notation Let $D = \{D_1, \dots, D_{|\mathcal{I}| \times |\mathcal{J}|}\}$ be the set of all inter-node distances, $\{d_{ij}\}_{i \in \mathcal{I}, j \in \mathcal{J}}$, arranged in increasing order.

Algorithm

1. Set $lbIndex = 0$ and $ubIndex = |D| - 1$.
2. Set $index = lbIndex + \lceil \frac{ubIndex - lbIndex}{2} \rceil$.
3. Obtain a heuristic solution to $\text{MSCLP}(D_{index})$, \bar{Y} . Let $\hat{Y} = \{i \in \mathcal{I} | \bar{Y}_i = 1\}$ be the set of located facilities and $|\hat{Y}|$ be the number of located facilities.

- (a) If $|\hat{Y}| \leq p$, set $ubIndex = index$ and return to step 2.
4. Build a heuristic solution to the RANPCP using the set cover solution \bar{Y} and let \hat{V} be the corresponding post-disruption radius.
- (a) If $D_{ubIndex} > \hat{V}$, set $ubIndex$ to the index corresponding to the value \hat{V} in the set D . Return to step 2.
5. Solve MSCLP(D_{index}) to optimality, obtaining solution Y^* .
- (a) If $|Y^*| < p$, set $lbIndex = index + 1$. Otherwise, set $ubIndex = index$. Return to step 2.

One way to find a heuristic solution to MSCLP in Step 3 is by using a heuristic for the set cover location problem described by Balas and Ho (1980) and modified here for the MSCLP. First, a demand point is said to be single covered if there is at least one facility within D_{index} . A demand point is multi-covered if there are at least r facilities within distance D_{index} . Let n_i be the number of facilities that can single-cover demand point i within distance D_{index} . Proceeding through the list of demand points by increasing order of n_i , cover a demand point i by locating the facility that single-covers the maximum number of un-multi-covered demand points. Continue until all of the demand points are multi-covered. Then remove all facilities that are redundant. A facility is redundant if all demand points are multi-covered after that facility is removed.

The method for building a heuristic solution to the RANPCP in step 4 is as follows.

1. If $|\hat{Y}| > p$, remove the facility whose removal minimizes the increase in the RANPCP objective.
2. Repeat step 1 until $|\hat{Y}| \leq p$
3. Return the modified set \hat{Y} as the RANPCP heuristic solution.

4.4.1 Bounds

We computed lower and upper bounds before using the binary search algorithm in order to reduce the set of distances over which the binary search algorithm searches. These lower and upper bounds can be added to the binary search algorithm by removing all values in D that are less than the lower bound or greater than the upper bound.

A simple lower bound can be obtained by locating the closest r facilities to every demand point:

$$LB_0 = \max_{j \in \mathcal{J}} \{d_{i_j^r, j}\}$$

Note that when $r = 1$, this lower bound is zero.

Eloumi et al. (2004) described another lower bound for the PCP, but did not describe how this lower bound can be modified for the RANPCP. Thus, here we describe how to modify their lower bound for the RANPCP. First let $i_j^r(r, i)$ be the r^{th} closest location to demand point j , not including location i . Let $\gamma_i = \max_{j \in \mathcal{J}} d_{i_j^r(r, i), j}$. Sort the γ_i values in increasing order $\gamma_1 \leq \gamma_2 \leq \dots \leq \gamma_{|\mathcal{J}|}$. Then, $LB_1 = \gamma_{|\mathcal{J}|-p}$. Note that LB_1 is not always zero for $r = 1$. In our experimentation, we used $\min\{LB_0, LB_1\}$ as our lower bound.

Eloumi et al. (2004) also described two upper bounds for the PCP but these bounds cannot be directly extended to the RANPCP because the RANPCP requires each demand point to be covered $(r + 1)$ times while the PCP only requires each demand point to be covered once. A simple upper bound for the RANPCP can be obtained by assuming that the p furthest facilities to every demand point are located. In this case, the r^{th} closest located facility to a demand point j will be located at the $(|\mathcal{J}| - p + r)^{\text{th}}$ closest location to j . Thus, the upper bound is:

$$UB_0 = \max_{j \in \mathcal{J}} \{d_{i_j^{|\mathcal{J}|-p+r}, j}\}$$

These bounds can also be used to improve the tractability of the MIP models. A lower bound lb may be added to models 1, 2, and 2C by adding the constraint $lb \leq V$. An upper bound ub may be added to model 1 by removing any variable $x_{ij\ell}$ for which $d_{ij} > ub$. An upper bound ub can be

added to models 2 and 2C by removing variables X_{ij} and Z_{ij} if $d_{ij} > ub$.

4.4.2 Greedy Heuristic

The following greedy heuristic, which is a modification of a heuristic by Mladenović et al. (2003) for the p -center problem, can be used to find an initial solution to the RANPCP. This objective value of this initial solution is an upper bound, which we call UB_1 .

1. Solve the 1-center problem ($\arg \min_{i \in \mathcal{J}} \max_{j \in \mathcal{J}} d_{ij}$) and place r facilities at the 1-center.
2. Remove a facility from the 1-center and place it at the node that minimizes the resulting objective increase; repeat until only 1 facility is located at the 1-center.
3. Let the set \mathcal{J}' be the set of locations that do not have a facility and let $\Delta(i)$ be the objective function decrease associated with locating a facility at i . Locate a facility at $i' \in \arg \min_{i \in \mathcal{J}'} \Delta(i)$. Repeat until all p facilities have been located.

4.5 Computational Experimentation

In this section we report the results of our computational experimentation with the goal of providing an empirical analysis of the scalability of our solution techniques. We also compare our MIP models to a MIP model from Elloumi et al. (2004), which we call Model E.

All experiments were run on a 64-bit 2.66GHz AMD processor running the Linux operating system with 16GB of memory. All MIP models, including the multi-set-cover location problem, were solved with CPLEX v12.1 using Java Concert Technology.

Before solving an instance, we first found a lower bound $LB = \max\{LB_0, LB_1\}$, an upper bound $UB = \min\{UB_0, UB_1\}$, and a feasible solution produced by the greedy heuristic in Section 4.4.2. For models 1, 2, and 2-C, we used the upper bound to eliminate variables (see Section 4.4.1) and seeded the branch and bound algorithm with an initial feasible solution. For the binary search algorithm and Model E, we used the upper and lower bounds as the initial upper and lower bounds for the algorithm.

We tested our solution methods on the 18 datasets from the facility location literature (see Appendix). We obtained datasets from various sources, including the traveling salesman problem library (TSPLIB) (Reinelt, 1991). These datasets come from several different sources including population centers in a city, cities within the United States, and manufacturing drilling problems. The size of problem that a practitioner would need to solve depends on the application and the level of aggregation. For example, if the practitioner judged that it was sufficient to consider cities as nodes then a realistic problem may consist of 300 nodes or less. However, if it is necessary to model districts in a city, then the number of nodes could be several thousand, depending on how a district is defined and how many districts are in the city.

For each instance we reported the time to optimality. When the optimal solution was not found after 5 hours, we reported the optimality gap defined by $(UB - LB)/UB$ and reported the run time as 18000 seconds. For the binary search algorithm, we also report the number of iterations run until optimality was reached.

4.5.1 Experimental Design

In our experimentation we varied the values of the number of facilities, p , and the number of neighbors, r . Each dataset with a value of p and r defines an instance of the RANPCP. We tested a set of instances for each dataset. Each set of instances for a dataset includes the instances $(p = 5, r = 1)$, $(p = 5, r = 2)$, $(p = 10, r = 1)$, and $(p = 10, r = 2)$ and the ratios $p/|\mathcal{S}| = 0.1, 0.2, 0.3$ and $r/|\mathcal{S}| = 0.1, 0.2, 0.3$. For each value of p in the set of instances, there is an instance with $r = 1$ and an instance with $r = 2$.

4.5.2 Comparison of MIP Models

The goal of this subsection is to compare the computation time of the MIP models discussed in this paper as well as Model E. Table 4.1 shows the computational results for using CPLEX branch and bound to solve several instances of the RANPCP. Each row contains the results from solving an instance of the RANPCP using Models 1, 2, 2-C, and E. For each model, the root bound ratio, run time, and gap are reported. The root bound ratio (RBR) is the ratio of the lower bound CPLEX

found after the root node to the optimal objective. When an optimal solution is not found after 5 hours, the root ratio is the ratio of the lower bound after the root node to the final upper bound value. Thus, when a feasible solution isn't found after 5 hours, both the RBR and the Gap are reported as ∞ .

The table shows that none of the models has the lowest run time in all instances. However, Models 1, 2, and 2-C mostly outperform Model E for the lor200 instance.

The table also shows that instances with $r = 2$ take longer than instances with $r = 1$ due to the larger size problem. This indicates that the RANPCP takes longer to solve than the p -center problem, a special case of the RANPCP with $r = 1$.

4.5.3 MIP Models vs. Binary Search

In this section we compare the computation time of our MIP models to that of the binary search algorithm. Table 4.2 displays the computational results for our MIP models and the binary search algorithm applied to several instances. Each cell in the table contains the time required to solve the problem to optimality. The table shows that the binary search algorithm requires much less computation time than the MIP models. These results are consistent with the results reported by Elloumi et al. (2004), who compared a binary search algorithm with a MIP model for solving the p -center problem.

Table 4.2: Computation time for binary search vs. MIP Models for various problem instances

No.	Dataset	p	r	Model 1	Model 2	Model 2-C	Model E	BS
1	lor100	5	1	2.4	6.1	4.2	1390	0.12
2	lor100	5	2	2039	5411	1754	317	0.08
3	lor100	10	1	0.4	2.5	2.5	4996	0.16
4	lor100	10	2	34	24	15	60	0.05
5	lon150	30	2	14610	58	38	3514	0.13
6	lon150	30	9	18000	18000	18000	18000	1.7
7	lon150	45	5	18000	18000	18000	3597	0.25

Table 4.1: MIP results for various datasets

No.	Dataset	p	r	Model 1			Model 2			Model 2-C			Model E		
				RBR	time (s)	Gap	RBR	time (s)	Gap	RBR	time (s)	Gap	RBR	time (s)	Gap
1	lor100	5	1	0.74	2.4	0	0.74	6.1	0	0.74	4.2	0	0.94	1390	0
2	lor100	5	2	0.74	2039	0	0.77	5411	0	0.77	1754	0	0.82	317	0
3	lor100	10	1	0.76	0.4	0	0.76	2.5	0	0.76	2.5	0	0.89	4996	0
4	lor100	10	2	1	34	0	1	24	0	1	15	0	1	60	0
5	lor200	5	1	0.66	435	0	0.66	55	0	0.66	36	0	∞	18000	∞
6	lor200	5	2	0.32	18000	0.57	0.32	18000	0.58	∞	18000	∞	∞	18000	∞
7	lor200	10	1	0.74	55	0	0.74	54	0	0.74	441	0	∞	18000	∞
8	lor200	10	2	0.86	18000	0.14	1	1803	0	0.74	18000	0.26	∞	18000	∞

4.5.4 Scalability of Binary Search Algorithm

In this section we examine how much the computational performance of the binary search algorithm is affected by changing problem parameters such as the number of locations, the number of facilities, and the number of neighbors.

First, consider how different datasets affect runtime. Table 4.3 shows summary statistics for the computational results of several datasets.

Table 4.3: Summary statistics for runtime and number of iterations for all instances of each dataset

Dataset	Total computation time (s)				Number of iterations			
	min	max	avg.	var.	min	max	avg.	var.
sw55	0	0.1	0.02	<0.01	0	10	4.9	19.52
lor100	0	0.26	0.06	<0.01	0	14	6.4	45.49
lon150	0.02	4.4	0.91	1.1	0	15	13	11.44
lor200	0.02	2.9	0.56	0.7	0	16	8.8	43.59
lor300a	0.04	17	1.8	16.2	0	17	6.5	51.15
lor300b	0.04	17	1.8	16.2	0	17	6.5	51.15
lor400a	0.1	70	6.3	253	0	18	8.1	53.05
lor400b	0.1	69	6.3	246	0	18	8.1	53.05
beas500	0.65	55	8.2	246	3	5	4.2	0.58
beas600	0.6	235	17	2825	1	5	3.4	1.15
beas700	1.1	201	22	3045	1	5	2.9	2.05
beas800	1.2	272	26	4808	1	5	2.5	2.04
lor818	7.6	5454	861	1690336	11	19	16	5.49
beas900	0.41	1275	80	84894	0	5	1.3	3.54
u1060	12	13793	2272	15785497	10	17	14	4.77

Table 4.3 shows that datasets with more nodes usually require more computation time to solve. On the other hand, the number of nodes does not significantly influence the number of iterations. The binary search algorithm takes $\log_2(|\mathcal{L}|^2 + 1)$ iterations in the worst case (Elloumi et al., 2004). However, the upper and lower bounds for the RANPCP and the heuristic for the MSCLP often reduce the number of iterations.

Because the number of iterations does not change significantly with an increase the the number of nodes, we conclude that the time per iteration increases as the number of nodes increase. Therefore, the increased run time is due to the time required to solve a larger set cover problem at

each iteration.

We also examined the affect of $|\mathcal{S}|$, p , and r on the runtime and found that the runtime increased with $|\mathcal{S}|$. However, we did not find that p or r had much affect.

4.6 Tradeoff Analysis

In this section we use the RANPCP to analyze the tradeoffs between cost, the regular performance of the system, and potential consequence. Cost is measured as the number of facilities that can be located and the regular performance of the system is measured as the maximum closest distance from a demand point to its closest facility when all facilities are available (i.e., the objective to the p -center problem). The potential consequence is measured as the maximum closest distance from a demand point to its r^{th} closest facility, i.e., the max distance from a demand point to its closest facility when r facilities are unavailable. The experiments in this section were run on the computer described in Section 5.6.4.

4.6.1 Penalties for Only Considering a Single Objective

A weakness of the RANPCP model is that it only models potential consequence without modeling regular system performance. This is a problem because regular system performance is usually a primary objective and potential consequence a secondary objective. Thus, using the RANPCP has a benefit and a drawback. The benefit is that it minimizes the potential consequence that occurs when r facilities are unavailable. The drawback is that its solution may have worse system performance than the optimal solution obtained when considering only regular system performance. In this section we quantify this benefit and drawback empirically.

We use the following notation in our measurements. Define the *max closest distance objective* as the objective of minimizing the distance to from a demand point to its closest located facility. Let the RANPCP objective with r neighbors required be called the *max r^{th} closest distance objective*. For a given instance, let $Y_{(1)}^*$ be the optimal facility configuration for the max closest distance objective and let $Y_{(r)}^*$ be the optimal facility configuration for the max r^{th} closest distance objective. The functions $f_{(1)}(Y)$ and $f_{(r)}(Y)$ are the max closest distance and max r^{th} closest objective values

for a location configuration Y . Let the *relative objective function increase for not considering regular system performance* be $\gamma_{1,r} = \frac{f_{(1)}(Y_{(r)}^*) - f_{(1)}(Y_{(1)}^*)}{f_{(1)}(Y_{(1)}^*)}$ and the *relative objective function increase for not considering facility unavailability* be $\gamma_{r,1} = \frac{f_{(r)}(Y_{(1)}^*) - f_{(r)}(Y_{(r)}^*)}{f_{(r)}(Y_{(r)}^*)}$.

Table 4.4 shows summary statistics for $\gamma_{1,r}$ and $\gamma_{r,1}$ over all of our datasets and instances. The average relative objective function increase for not considering regular system performance is 0.43 while the average relative objective function increase for not considering facility unavailability is 6.3. Hence, if only one objective is used, it should be the post-disruption radius. However, since the objectives are conflicting, a bi-objective model is more appropriate.

Table 4.4: Summary statistics for $\gamma_{1,r}$ and $\gamma_{r,1}$ for all instances of each dataset

	$\gamma_{1,r}$				$\gamma_{r,1}$				$\gamma_{r,1}/\gamma_{1,r}$			
	min	max	avg.	var.	min	max	avg.	var.	min	max	avg.	var.
sw55	0.18	0.96	0.51	0.08	0.32	1.70	1.10	0.25	0.21	1.00	0.53	0.05
lor100	0.00	0.92	0.48	0.08	1.40	9.70	5.90	11.00	0.00	0.37	0.15	0.02
lon150	0.00	0.69	0.39	0.06	0.59	5.80	2.00	2.17	0.00	0.94	0.33	0.09
lor200	0.00	0.84	0.47	0.07	0.90	21.00	10.00	49.13	0.00	0.58	0.12	0.03
lor300a	0.00	1.50	0.64	0.22	0.81	26.00	13.00	75.18	0.00	1.00	0.16	0.08
lor300b	0.00	1.50	0.64	0.22	0.81	26.00	13.00	75.18	0.00	1.00	0.16	0.08
lor400a	0.00	1.30	0.52	0.19	0.85	37.00	16.00	142.91	0.00	1.30	0.15	0.12
lor400b	0.00	1.30	0.52	0.19	0.85	37.00	16.00	142.91	0.00	1.30	0.15	0.12
beas500	0.15	0.39	0.30	0.01	0.18	2.10	1.00	0.43	0.10	1.60	0.52	0.22
beas600	0.16	0.37	0.25	0.00	0.08	2.70	1.30	0.71	0.07	2.10	0.48	0.40
beas700	0.03	0.49	0.25	0.02	0.11	3.30	1.80	1.07	0.01	3.90	0.66	1.73
beas800	0.03	0.31	0.20	0.01	0.15	3.40	1.90	1.19	0.02	1.90	0.27	0.24
lor818	0.28	1.30	0.61	0.15	0.46	11.00	3.90	10.55	0.03	1.30	0.37	0.15
beas900	0.08	0.38	0.17	0.01	0.38	4.90	3.00	2.31	0.02	0.99	0.15	0.08
u1060	0.00	0.86	0.53	0.07	0.50	8.00	4.60	6.12	0.00	0.91	0.22	0.08
ALL	0.00	1.50	0.43	0.01	0.08	37.00	6.30	2625	0.00	3.90	0.29	0.18

4.6.2 Multi-objective Analysis: Max Closest Distance Vs. Max r^{th} Closest Distance

Given that the max closest distance and the max r^{th} closest distance objective may often be conflicting objectives, a Pareto efficient frontier for these two objectives may be useful to a decision maker. In this section we examine the shape of Pareto efficient frontiers for these two objectives.

A point on the curve is represented by (δ_1, δ_r) , where δ_1 is the max closest distance and δ_r is

the max r^{th} closest distance. Point A dominates point B if point A is better than point B in one objective and point A is no worse than point B in the other objective. A point that is not dominated by any other point is called a Pareto optimal point. A Pareto efficient frontier is composed of Pareto optimal points.

In our analysis we generate points on a tradeoff curve and then determine the Pareto points by where the curve jumps. A point (δ_1, δ_r) on the curve for a given value of p can be obtained by solving the RANPCP subject to the constraint that the max closest distance must be greater than δ_1 and returning δ_r as the optimal objective value. This problem can be solved with the binary search algorithm; simply substitute the MSCLP with the constrained MSCLP (C-MSCLP) and use binary search to find the optimal δ_r subject to the constraint that the max closest distance is greater or equal to δ_1 . The C-MSCLP is formulated as:

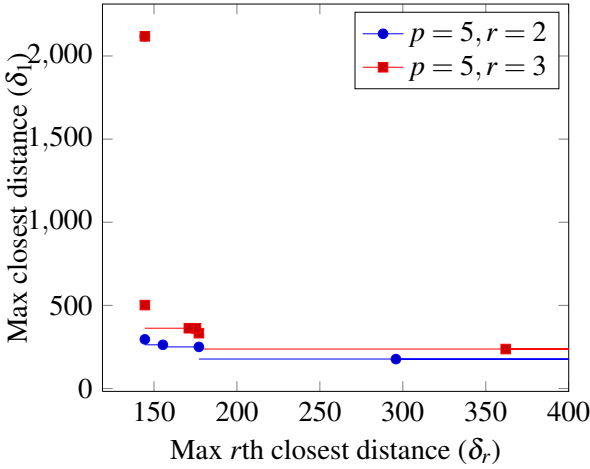
$$\text{(C-MSCLP}(\delta_1, \delta_r)) \quad \min \sum_{i \in \mathcal{I}} Y_i \quad (4.7)$$

$$\text{s.t.} \quad \sum_{i \in \{i: d_{ij} \leq \delta_1\}} Y_i \geq 1 \quad \forall j \in \mathcal{J}$$

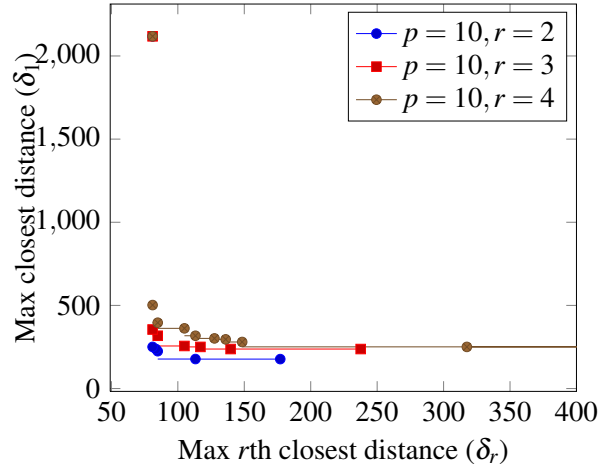
$$\sum_{i \in \{i: d_{ij} \leq \delta_r\}} Y_i \geq r \quad \forall j \in \mathcal{J} \quad (4.8)$$

$$Y_i \in \{0, 1\} \quad \forall i \in \mathcal{I} \quad (4.9)$$

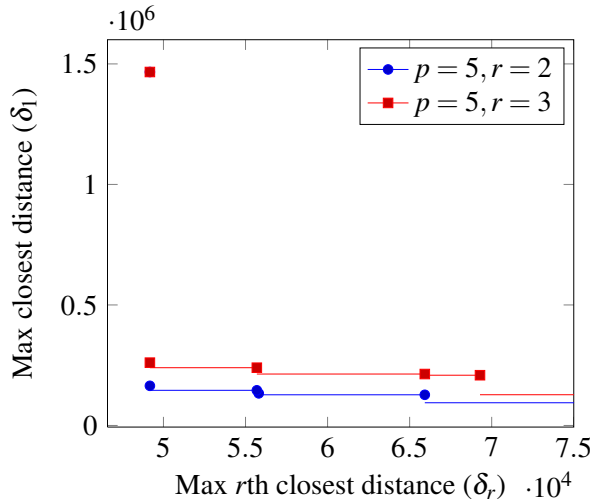
Figure 4.1 shows tradeoff curves between the max closest distance and max r^{th} closest distance for several instances of the sw55 and lor100 datasets. The ranges of the plots are adjusted to show the areas in which the plots are not flat. The lines that stop at the right edge of the plot actually continue to the maximum distances value of d_{ij} . The points on the plots represent Pareto optimal solutions. The lines that extend to the left of the points represent how far the max closest r^{th} distance can decrease if the max closest distance remains constant.



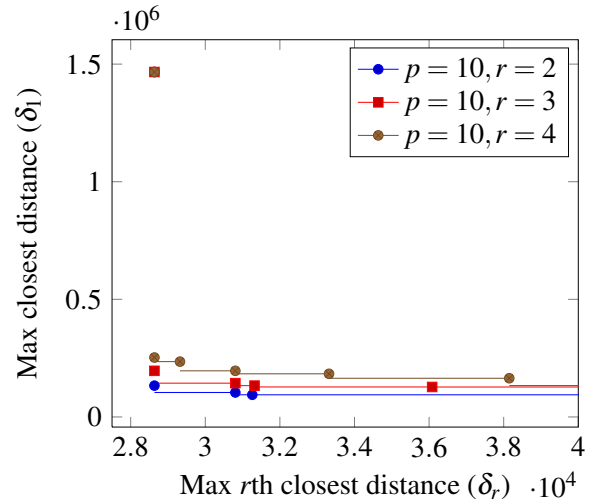
(a) Swain 55-node dataset (sw55), $p = 5$



(b) Swain 55-node dataset (sw55), $p = 10$



(c) Lorena 100-node dataset (lor100), $p = 5$



(d) Lorena 100-node dataset (lor100), $p = 10$

Figure 4.1: Regular max closest distance vs. max r^{th} closest distance for several datasets

As expected, as p increases the tradeoff curves shift to the left and as r increases the curves shift upward. An important feature of these plots is the length of the lines extending left of the Pareto optimal points. The length of these lines indicate that there are a lot of non-Pareto optimal solutions. Thus, if only one of the objectives is considered it is likely to find a point that is not Pareto optimal.

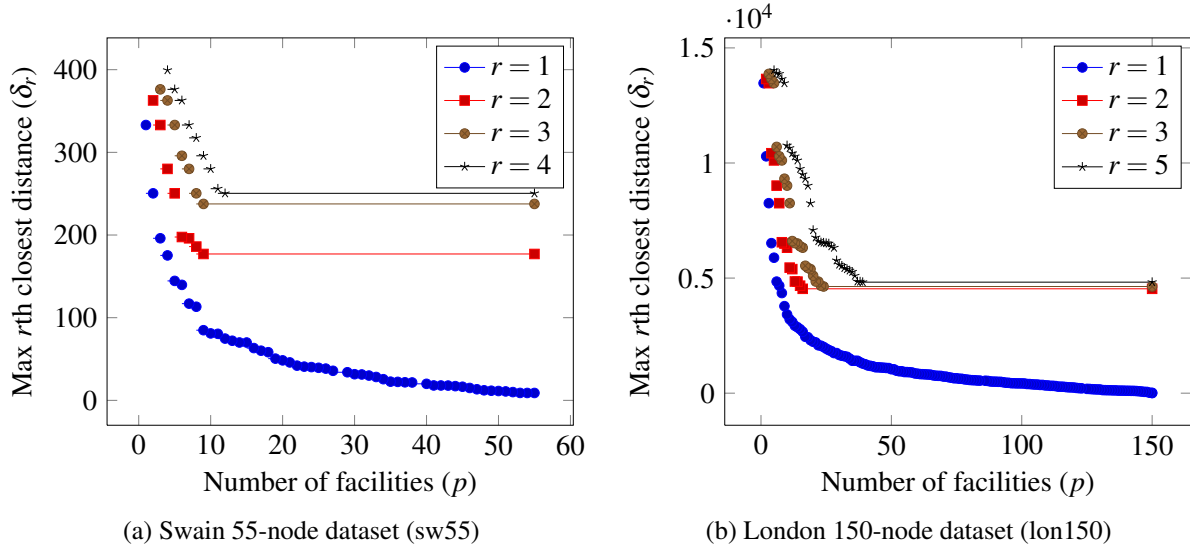


Figure 4.2: Number of facilities vs. max r^{th} closest distance for several datasets

4.6.3 Multi-objective Analysis: Cost Vs. Maximum r^{th} Closest Distance

The value of p , the number of facilities, is also likely to influence the optimal objective value of the RANPCP. This can represent the cost of building the system. Therefore, a decision maker may benefit from a tradeoff curve for the number of facility locations p and the max r^{th} closest distance. This curve can be generated by solving MSCLP(δ) for all values of δ in the distance matrix $\{d_{ij} : i \in \mathcal{I}, j \in \mathcal{J}\}$.

Figure 4.2 shows tradeoff curves for p vs. max r^{th} closest distance for several instances of datasets sw55 and lon150. The curves are similar in shape. This indicates that the shape of the curve is somewhat independent of the dataset and of r .

4.6.4 Saturation Point

In this section we investigate a property of the RANPCP called *saturation*. An instance of the RANPCP is saturated if the r closest facilities to a given demand point are located and the distance from that demand point and its r^{th} closest located facility is equal to the optimal objective value. When an instance is saturated for a given value of p and r , locating additional facilities does not improve the objective. Hence, the instance is saturated.

The specific analysis that we present in this section is an analysis of the point at which datasets become saturated for a value of r . Let $p^*(r)$ be the *saturation point* for a dataset with r neighbors. In other words, $p^*(r)$ is the smallest value of p such that the instance of the RANPCP with a given number of facilities p and number of neighbors r is saturated.

Figure 4.3 shows the saturation point vs. r for several datasets. For an instance with a given value of r , the saturation point can be found by setting $p = |\mathcal{S}|$ and using the binary search algorithm in Section 4.4. When the algorithm terminates, record the number of facilities located in the optimal solution. This is the saturation point. In the Figure, the saturation point is always greater than r . This implies that $p^*(r) \geq r$, which must be true because a saturated solution has at least r located facilities. The figure also shows that the saturation point is not monotonic with respect to r .

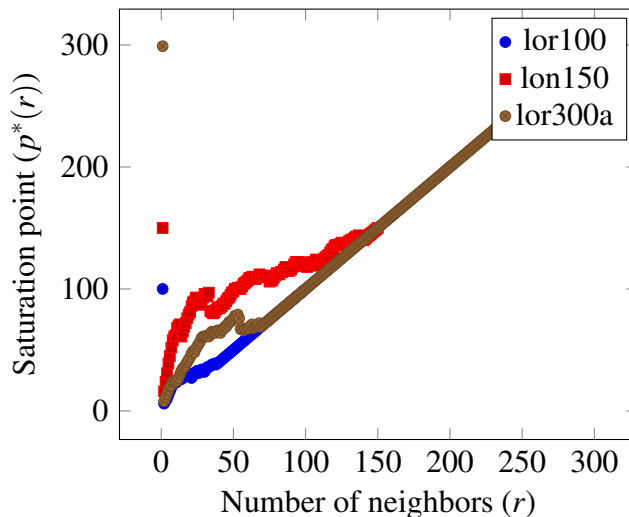


Figure 4.3: Saturation point vs. r for several datasets

Figure 4.3 also shows that the saturation curves eventually approach a point where the saturation point equals r . Let the *super-saturation point*, r_0 , be a value of r such that $p^*(r) = r$ and $p^*(r_0 + n) = r_0 + n$ for $n = 1, \dots, |\mathcal{S}| - r_0$. We observed the following about the super-saturation point:

1. Once the super-saturation point is reached, the solutions often become nested. In other words, the solution for $p^*(r_0 + n)$ is a subset of the solution for $p^*(r_0 + n + 1)$ for $n =$

$0, \dots, |\mathcal{S}| - r_0 - 1$.

- Objective values are not always the same for $p^*(r_0 + n)$ for $n = 1, \dots, |\mathcal{S}| - r_0$. Consequently, the bottleneck pair (the facility i and demand point j for which d_{ij} equals the optimal objective) are also different for different values of r .

Table 4.5 contains the super-saturation point for several instances. The table shows that the ratio $\frac{r^*}{|\mathcal{S}|}$ varies across datasets. Also, all of unweighted datasets had a super saturation point equal to $|\mathcal{S}|$. Thus, the saturation is clearly influenced by the demand point weights.

Table 4.5: Super-saturation point for various datasets

Dataset	Wtd.		Unwtd.		Wtd.-Unwtd.	
	r_0	$\frac{r_0}{ \mathcal{S} }$	r_0	$\frac{r_0}{ \mathcal{S} }$	r_0	$\frac{r_0}{ \mathcal{S} }$
sw55	22	0.45	49	1	-27	-0.55
lor100	38	0.38	100	1	-62	-0.62
lon150	140	0.93	150	1	-10	-0.07
lor200	67	0.34	200	1	-133	-0.66
lor300a	69	0.23	300	1	-231	-0.77
lor300b	69	0.23	300	1	-231	-0.77
Min	38	0.23	55	1	-231	-0.77
Max	140	0.93	300	1	-9	-0.07
Average	72	0.49	184	1	-113	-0.57

Appendix Section B.3 contains theoretical results related to saturation.

4.7 Example with Managerial Insights

In this section we discuss the implications of the empirical results described in Section 4.6. We explain these insights through a detailed analysis of the classic 55-node dataset from Swain (1971) (abbreviated sw55). The nodes in this dataset represent districts in the city of Washington, D.C. The nodes each have a weight that is proportion to the population at that node.

In this case study a decision maker wishes to locate ambulances within the districts of the city. The decision maker is especially interested in the response time for emergencies requiring more than one ambulance. Each demand point represents a district and has a weight corresponding to population. The decision is where to locate p ambulances within the 55 city districts.

First, consider the solution to the RANPCP with $p = 10$ and $r = 3$, shown in Figure 4.4a. The maximum distance from a demand point to its 3rd closest located facility for this solution is 237. The maximum closest distance for this facility configuration is 160.

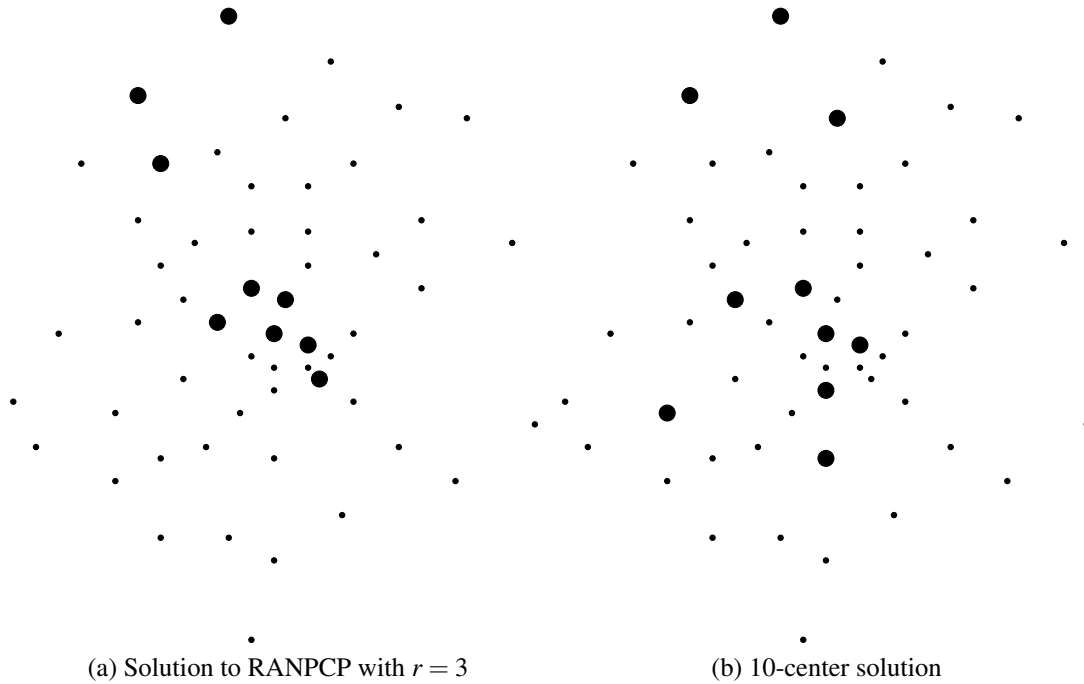


Figure 4.4: Ambulance locations for case study

Next, consider the problem of locating p facilities to minimize the maximum time required for the first vehicle to arrive at a scene. This problem can be solved using the classic p -center model. The solution for $p = 10$ is shown in Figure 4.4b. For this solution, the maximum distance from a demand point to its closest located facility is 81 and the maximum distance from a demand point to its 3rd closest located facility is 355.

The relative objective function increase for not considering backups, $\delta_{3,1}$, is $\frac{355-237}{237} = 0.49$. This means that if the p -center solution is chosen, the response time of the 3rd vehicle is 49% higher than if the RANPCP solution had been used. The relative objective function increase for not optimizing first vehicle response time is, $\delta_{1,3}$, is $\frac{160-81}{81} = 0.97$. This means that if the RANPCP solution is chosen, the response time of the first vehicle is 97% higher than if the p -center solution had been used. We also find that $\delta_{1,3}$ is almost 2 times greater than $\delta_{3,1}$.

Table 4.6 shows the relative objective function increase for various values of p and r for the sw55 dataset (for complete results, see Table 4.4).

Table 4.6: Relative objective function increase for considering and for not considering facility unavailability for dataset s5w5

No.	p	r	$\frac{p}{ \mathcal{S} }$	$\frac{r}{p}$	$\delta_{1,r}$	$\delta_{r,1}$	$\delta_{r,1}/\delta_{1,r}$
1	5	2	0.091	0.400	0.32	0.18	0.57
2	5	3	0.091	0.600	0.32	0.51	1.61
3	10	2	0.182	0.200	0.98	0.41	0.42
4	10	3	0.182	0.300	0.97	0.49	0.51
5	10	4	0.182	0.400	1.30	0.79	0.61
6	15	2	0.273	0.133	1.29	0.34	0.27
7	15	3	0.273	0.200	1.29	0.05	0.04
8	15	4	0.273	0.267	1.21	0.58	0.48
Average					0.96	0.42	0.56

The relative objective function increase metrics, $\delta_{r,1}$ and $\delta_{1,r}$, have several implications for decision makers. First, we found that both $\delta_{r,1}$ and $\delta_{1,r}$ have an average value of more than 0.4. This indicates that these increases are significant, showing that both the RANPCP model or the p -center model have weaknesses. Second, we found that the average of $\delta_{1,r}$ is greater than the average of $\delta_{r,1}$. We also found that $\delta_{1,r} > \delta_{r,1}$ in all but one of the instances tested for the Swain dataset. This indicates that the relative objective function increase for not considering regular system performance is greater than the relative objective function increase for not considering backups. Third, our empirical analysis in Section 4.6.1 showed that in some instances $\delta_{r,1} > \delta_{1,r}$ and in other instances $\delta_{r,1} < \delta_{1,r}$. Thus, neither model dominates the other. As a result, the best model to use depends on the preferences of the decision maker. Because neither the RANPCP model or the p -center model dominates, a tradeoff curve of both objectives is more useful to a decision maker than a solution for only one objective.

Figure 4.5 shows the saturation curve for the s55 dataset. There are two implications of saturation that we wish to explain. On one hand, if the maximum of the time required for the r^{th} vehicle to arrive at a scene is truly the only performance measure that the decision maker is interested in, then the saturation point shows the point at which locating more facilities is wasteful. However,

we have demonstrated in this section that it is sometimes not a good idea to only consider one objective. When the RANPCP model is used, the saturation point is the point at which locating additional facilities cannot improve the performance measure. However, it is likely that locating additional facilities may improve another objective. For example, adding an additional facility will always reduce the maximum time required for the first vehicle to arrive at a scene.

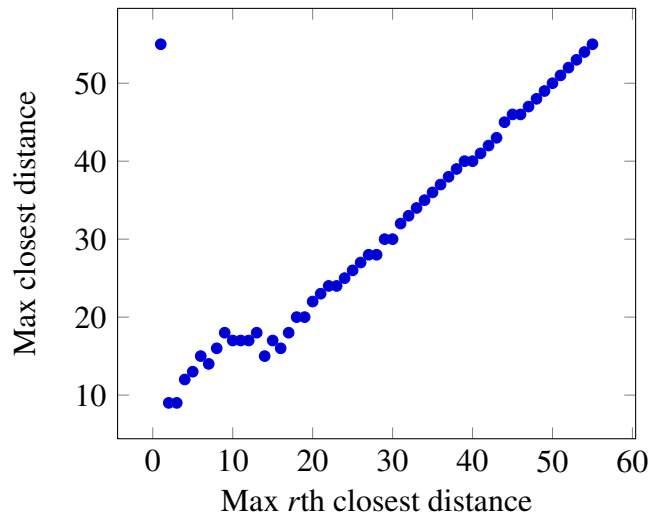


Figure 4.5: Saturation points for s55 dataset

4.8 Conclusions and Future Work

This paper described a study of the r -all-neighbor p -center problem (RANPCP) and makes the following contributions to the literature:

1. We developed a new mixed-integer programming model that is an alternative to an existing MIP formulation of the RANPCP. We found that the solution times of our model are competitive with an existing MIP model.
2. We discovered a structural property of the RANPCP called saturation, the point at which locating additional facilities does not improve the objective function. As we discussed, this property shows a drawback of considering potential consequences in isolation.
3. We described lower and upper bounds that can be used to improve the performance of the MIP models and binary search algorithm.

4. We performed an empirical study of the computational capability of our MIP models and a binary search algorithm.
5. We performed further experimentation to gain insights into the RANPCP.

We gained several insights from our computational study. First, the average computation time of the reformulated model (Model 2) was less than the average computational time of the straightforward model (Model 1). Adding valid inequalities to Model 2 (resulting in Model 2-C) improved the average computation time. However, the results were mixed. Overall, we found that Model 2, Model 2-C and an MIP model by Elloumi et al. (2004) were all competitive with regard to tractability. Second, the binary search algorithm took much less time to solve the RANPCP than solving the MIP models with branch and bound. Third, the computation time of the binary search algorithm was driven by the time required to solve multi-set-cover location problems. In general, the computation time increased as the number of potential facility locations increased.

Further experimentation revealed several insights into the RANPCP. First, we found that optimizing either regular system performance in isolation or potential consequence in isolation produces solutions that perform poorly in the other objective. We found that the solutions that are optimal for potential consequence have a regular system performance that is, on average, 43% more than the optimal. However, solutions that are optimal for regular system performance have a potential consequence value that is, on average, 6.3 times more than the optimal. Thus, if only one objective is modeled, it should be the potential consequence objective. We then generated solutions that are Pareto-optimal for the regular system performance and potential consequence objectives. The solutions that we generated showed that a significant reduction in potential consequence can be obtained by allowing a small reduction in regular system performance, and vice versa. We also used our model to analyze the tradeoff between the number of facilities built and the potential consequence. We found that for several instances, significant reductions in potential consequence can be obtained by building a small number of additional facilities.

In our opinion, the most important area of future work is to develop more efficient ways of doing a multiobjective analysis. In Section 4.6 we displayed Pareto-efficient frontiers for several

objectives including cost, regular system performance, and worst-case consequence. Our results show that this type of analysis can help a decision maker. However, we could only do this analysis for small datasets because we are not aware of any efficient methods to perform this analysis. Thus, a method that can efficiently identify Pareto optimal solutions with regard to two or more of these objectives would be useful.

There are several other objectives that may be of interest to people deciding to incorporate potential consequence when locating facilities. For example, the total distance objective could be used. In the situation where facilities are located, this could represent operating cost or efficiency. In the situation where vehicles are located, this could represent the average response time or the average efficiency. The total weighted distance after facility unavailability could also be considered.

Improvements in the computation time for the RANPCP can be made in several ways. First, if better heuristics are found, they can be integrated into the binary search algorithm. Other algorithms could be used to solve this problem besides binary search.

The models in this paper assumed that the number of neighbors, r , is known with certainty. It may be useful for decision makers to have a model that allows them to place a probability distribution on r in order to minimize the expected loss. This could be used to model the situation where a decision maker is unsure about the amount of resources that an interdictor has. It could also be used when a decision maker is interested in emergency response to different types of incidents, each of which require a different number of vehicles.

5 Integrated Facility Location and Hardening

Abstract

Two methods of reducing the risk of distribution systems are locating facilities in less vulnerable locations and hardening facilities. These two activities have been treated separately in most of the academic literature. This paper integrates facility location and facility hardening decisions by studying the minimax facility location and hardening problem (MFLHP), which has the objective of minimizing the maximum distance from a demand point to its closest located facility after facility disruptions. The formulation assumes that the decision maker is risk averse and thus interested in mitigating against the facility disruption scenario with the largest consequence, an objective that is appropriate for modeling facility interdiction. Using problem structure, the integrated version of the MFLHP can be formulated as a single-stage mixed-integer program (MIP). Rather than solving the MIP directly, the MFLHP can be decomposed into sub problems and solved using a binary search algorithm. Experimental results showed that the binary search algorithm can solve problems with up to 1323 nodes. The results also showed that integrating location and hardening is beneficial: a non-integrated method produces solutions that are, on average, up to 65% worse than the solutions produced by the integrated model.

5.1 Introduction

There are two main categories of methods to reduce the vulnerability of a network. One category, which we refer to as the *design* category, involves designing the network differently. The network can be designed with redundancy or excess capacity which allows it to withstand the shock of a disruption. Another category, which we denote as the *risk-reduction* category, involves reducing the risk of existing networks. Methods in this category include adding redundancy to an existing network and hardening elements of the network. A survey by Medal et al. (2011b) mentions several other methods for reducing the risk of an existing network.

In this paper, we describe a model that integrates the design and risk reduction categories. In

particular, the decisions that our model considers are how many facilities to locate (design), where to locate facilities (design), how many facilities to harden (risk-reduction), and which facilities to harden (risk-reduction). There are several questions that this paper attempts to answer. Would this integrated method result in a better solution? Does the improvement in solution quality from using an integrated method outweigh the potential increase in computational burden? This paper addresses these questions based on an empirical study by exercising our new developed mathematical models for this problem.

There are several other questions that this research addresses. First, how much less vulnerable are facility systems designed using our model than facility systems designed without considering the consequence of facility disruptions? Second, how well do facility systems designed using our model perform when there are no disruptions? Third, how much does it cost to design a system that is resilient to facility disruptions? Fourth, how much more computational resources does the integrated model require over the non-integrated model? Fifth, can the integrated model solve realistically-sized networks in a reasonable amount of time?

In this paper we describe a model for integrating facility location and facility hardening objectives. Facility hardening is a special case of facility protection in which a hardened facility cannot fail. Since our model has a min-max structure we use binary search by decomposing the problem into sub problems that are similar in structure to the set-cover problem. To improve the tractability of our binary search procedure, we augment it with initial bounds and a greedy heuristic. We also describe a special case of our model, which allocates hardening resources amongst a set of existing facilities. This model is very similar to the classic p -center model, which is formulated in Daskin (1995). We then report the results of computational experiments to give an idea of the scalability of our solution procedure. We also discuss the results of a set of experiments that were performed to assess the quality of the solutions produced by our model and the tradeoffs involved in locating and hardening facilities subject to disruptions.

This paper contributes to the literature by (1) describing a new model for integrating facility location and hardening decisions, (2) describing a binary search solution procedure for the integrated

problem along with a new lower bound, (3) describing a new model for hardening facilities, (4) presenting empirical results that give several insights into locating and hardening facilities subject to disruptions as well as the computational feasibility of our solution procedure.

5.2 Literature Review

There is an extensive amount of work on measuring the vulnerability of networks. A field called network science tries to model the vulnerability of real world networks (see Alderson (2008) for a discussion on how this area relates to operations research). Grubestic and Matisziw (2008) and Grubestic et al. (2008) have provided reviews of network vulnerability analysis. Finally, there has been a lot of work on modeling the reliability of networks (see Ball et al. (1995)).

Another way to study network vulnerability is to identify the most critical elements of the network. Researchers have used interdiction models to identify the critical elements of max-flow networks (Wood, 1993), shortest-path networks (Israeli and Wood, 2002), and drug-smuggling networks (Morton et al., 2007). Church et al. (2004) developed one of the first models for interdicting facilities.

Other researchers have developed models to design networks that are vulnerable to disruptions. Snyder et al. (2006) and Peng et al. (2011) have presented network design models that consider random element disruptions.

Researchers have also addressed how to reduce the risk of existing networks. Several researchers have described models for fortifying networks that are vulnerable to random element failure (Peeta et al., 2010; Liu et al., 2009; Wallace, 1987b; Holmgren et al., 2007). Others have examined how to fortify networks that are subject to interdiction Bier et al. (2007b); Yao et al. (2007); Qiao et al. (2007); Cappanera and Scaparra (2011). For more information on both of these topics, see the review by Medal et al. (2011b).

There has also been some research on locating facilities subject to disruptions. Drezner (1987); Snyder and Daskin (2005); Cui et al. (2011) have all presented models for locating facilities subject to random failures. Church et al. (2004) presented models for choosing the optimal way to interdict a set of facilities in order to maximize the post-interdiction total weighted distance. They

also present a model that maximizes the post-interdiction total weighted covered demand, or demand that is within a distance standard. Others have extended the work of Church et al. (2004) and studied how to locate facilities subject to interdiction. Church and Gerrard (2003) studied the multi-cover problem, which extends the set-cover problem by requiring that each demand point be covered by r facilities. The multi-cover problem is equivalent to minimizing the cost of locating enough facilities to ensure that after interdiction all of the demand points are still covered. O'Hanley and Church (2011) developed a bi-level model for the problem of locating facilities to minimize the post-interdiction total weighted covered demand. Drezner (1987) studied how to optimally locate facilities in order to minimize the post-interdiction radius; this problem is equivalent to the r -all-neighbor p -center problem (Krumke, 1995; Khuller et al., 2000; Elloumi et al., 2004; Medal et al., 2011a). Snyder et al. (2010) have reviewed the problem of locating facilities subject to both random failures and interdiction.

Rather than locating facilities, several studies have examined how to optimally allocate fortification resources to a set of existing facilities. O'Hanley et al. (2007b) and Zhan (2007) have presented models for fortifying facilities subject to random failures. Several authors have analyzed how to fortify facilities against interdiction. Church and Scaparra (Church and Scaparra, 2007; Scaparra and Church, 2008b,a) have studied the problem of how to harden facilities in order to minimize the post-interdiction total weighted distance. O'Hanley et al. (2007a) presented a bi-level model to optimally harden facilities in order to minimize the post-interdiction total weighted covered demand. We are not aware of any models for hardening facilities in order to minimize the post-interdiction radius; therefore, we present a model for this objective in the current paper. For more information on hardening facilities subject to interdiction, see the survey by Medal et al. (2011b).

Researchers have begun to explore making facility location and facility fortification decisions simultaneously. Snyder and Daskin (2005) and Lim et al. (2010a) extend existing location models to include random facility failures. Snyder and Daskin (2005) present extensions of the p -median and warehouse location models and include perfectly reliable and unreliable facility locations in

their model. Specifically, a facility is perfectly reliable if and only if it is located at a perfectly reliable location. Their computational results showed that adding reliable facilities significantly increases system resilience. Lim et al. (2010a) present an extension of the warehouse location problem in which the decision maker chooses between locating unreliable facilities and perfectly reliable backup facilities, at a higher cost. Each demand point is required to have a reliable backup. Thus, if a demand point's primary facility fails, the demand point is then assigned to its reliable backup. This assumption simplifies the model and allows the authors to provide several useful analytical results. Aksen et al. (2011) study an extension of the p -median problem in which facilities are susceptible to interdiction. They present a bilevel version of the budget-constrained median location model in which a defender locates and hardens facilities and then an attacker destroys a fixed number of unhardened facilities. Their model builds on the model of Snyder and Daskin (2005) by allowing any facility to be hardened, not just facilities at perfectly reliable locations. Their model builds on the model of Lim et al. (2010a) by modeling the assignment of demand points after failures in a different way: when a demand point's primary facility fails it is assigned to the next closest open facility, rather than going directly to a reliable backup. Aksen et al. (2011) study three methods of solving their model: an enumeration procedure, a two-phase tabu search algorithm, and a two-phase heuristic. In both the tabu search algorithm and the two-phase heuristic, the location and hardening decisions are made sequentially, rather than together.

Our work builds upon the literature on facility location and facility hardening in the following ways. First, our work represents the first attempt to integrate facility location and facility hardening decisions while considering post-disruption maximum distance. The maximum distance is a popular objective for locating facilities in the public sector Daskin (1995), because optimizing this objective produces equitable solutions. Second, we build on the work of Aksen et al. (2011) by providing an exact procedure for solving our integrated location-hardening model, rather than using heuristics that decouple the two decisions. Because the solutions produced by our procedure are optimal, we are able to measure the benefit of integrating the location and hardening decisions in a single model.

5.3 Problem Description and Models

To facilitate our analysis, we developed a mathematical model for the minimax facility location-hardening problem (MFLHP). In particular, our model prescribes how to optimally locate and harden a set of facilities. The purpose of this model is to

locate a set of facilities and harden a subset of the located facilities in order to minimize the maximum consequence over all possible disruption scenarios consisting of the disruption of r facilities. The consequence of a disruption scenario is the maximum distance from a demand point to its closest located and operating facility.

The MFLHP has several applications. First, it can be used to *locate unreliable response vehicles*. An unreliable vehicle can be made perfectly reliable by hardening it. The model would then prescribe the location and hardening of response vehicles in order to minimize the maximum response time after the disruption of r vehicles. Second, this model can be used to locate and *harden facilities that are subject to attack by a strategic attacker*. In this case, the strategic attacker attacks up to r facilities that maximally degrade the performance of the system. In this case, the performance of the system is the post-interdiction radius.

To understand our model, it may help to divide it into three stages: 1) the mitigation stage, 2) the disruption stage, and 3) the response stage. To explain our model, we use the generic term facility to refer to what we are locating and hardening. We could also use the term vehicle or the more specific term warehouse, depending on the application. The mitigation stage happens before the disruption occurs. In this stage, actions can be taken to mitigate against the disruption. The mitigation decisions in our model are where to locate facilities and which facilities to harden. The location and hardening decisions can be made together or separately. If a facility is hardened in our model, it is always available to serve a demand point. In the disruption stage, the disruption causes exactly r facilities to fail. Thus, if p facilities are located, there are $\binom{p}{r}$ combinations of facility disruptions. (Later, we show that we do not have to consider all combinations.) In the response stage, demand points are served by their closest located facility. Since the decision in this stage is so simple—find the closest facility to each demand point—this stage will be implicit in our model.

To understand the three-stage model it is helpful to think of it as consisting of three players acting in sequence: a defender, an attacker, and an operator. In the first stage, the defender mitigates against the actions of the attacker by strategically locating and hardening facilities. The defender's objective is to minimize the attacker's objective. The attacker, knowing the location and hardening actions taken by the defender, then destroys r facilities. The objective of the attacker is to maximize the operator's objective, i.e., maximize the post-interdiction radius. The operator, observing the actions of the attacker, pairs each demand point with its closest available facility in order to minimize the post-interdiction radius.

The following notation will be used in our model. Let \mathcal{I} be a set of potential facility locations. Let \mathcal{L} be a set of located facilities and \mathcal{H} be a set of hardened facilities. Only located facilities can be hardened; thus, $\mathcal{H} \subseteq \mathcal{L}$. Denote $\chi(\mathcal{L}, \mathcal{H})$ as the cost of all location and hardening activities, which is subject to a budget b . The set \mathcal{O} is the set of located facilities that fail due to interdiction or random causes. At most r facilities can fail in a disruption. A set of demand points is represented by the set \mathcal{J} . We measure the effectiveness of a facility located at i serving the demand point located at j using a measure ϕ_{ij} . The measure ϕ_{ij} could be the distance between i and j or a function of the distance between i and j . It could also represent the distance multiplied by the demand weight w_j . Let $D_j(\mathcal{L}) = \min_{i \in \mathcal{L}} \phi_{ij}$ be the distance from demand point j to the closest facility in the set \mathcal{L} .

Our three-level model is

$$\delta^* = \min_{\substack{\mathcal{L} \subseteq \mathcal{I} \\ \mathcal{H} \subseteq \mathcal{L} \chi(\mathcal{L}, \mathcal{H}) \leq b}} \max_{\substack{\mathcal{O} \subseteq \mathcal{L} \setminus \mathcal{H} \\ |\mathcal{O}| \leq r}} \max_{j \in \mathcal{J}} D_j(\mathcal{L} \setminus \mathcal{O}). \quad (5.1)$$

This is the *integrated MFLHP model*. The outer minimization problem, the defender's problem, is to locate a set of facilities \mathcal{L} from a set of candidate locations \mathcal{I} and harden a subset, \mathcal{H} , of the located facilities. The cost of location and hardening must be within a budget, b . The outermost maximization problem, the interdictor's problem, is to destroy a subset, \mathcal{O} , of the located, unhard-

ened facilities in order to maximize the post-interdiction radius. The interdicator can only destroy r facilities. The inner maximization problem is the operator's problem, which is to assign demand points to their closest operating facilities. Since this problem is trivial, it can be represented by taking the maximum closest distance over the set of demand points.

5.3.1 Describing the Most Disruptive Facility Disruption Scenario

Rather than trying to solve the three-stage formulation of the integrated MFLHP model, problem structure can be exploited to formulate the integrated MFLHP model as a single-stage problem. In particular, because of the minimax distance objective of the integrated MFLHP model, the interdicator's optimal solution can be described in closed form.

Because of the max-max structure of the interdicator's problem, we can determine exactly which facilities the interdicator would want to destroy to optimize his objective without having to solve an optimization problem. After the sets \mathcal{L} and \mathcal{H} have been chosen, the interdicator's problem is to solve

$$\delta^* = \max_{\substack{\mathcal{O} \subseteq \mathcal{L} \setminus \mathcal{H} \\ |\mathcal{O}| \leq r}} \max_{j \in \mathcal{J}} D_j(\mathcal{L} \setminus \mathcal{O}). \quad (5.2)$$

Let \mathcal{L}_j^i be the facilities that are closer to demand point j than facility i . Let $\mathcal{O}^*(\mathcal{L}, \mathcal{H})$ be the set of facilities that optimizes (5.2); that is, the set of facilities whose disruption maximizes the post-interdiction radius. The following proposition, which is an extension of an idea by Drezner (1987), characterizes the set $\mathcal{O}^*(\mathcal{L}, \mathcal{H})$ and shows that the interdicator will always interdict the closest unhardened facilities to a single demand point:

Definition 1. Let δ^* be the optimal post-interdiction radius. Demand point j' and facility i' are a *post-interdiction bottleneck pair* if $\delta^* = \phi_{i'j'}$. In this case, j' is called a *post-interdiction bottleneck demand point* and i' is called a *post-interdiction bottleneck facility*.

Proposition 1. *If i' and j' form a bottleneck pair, then the set $\mathcal{O}^*(\mathcal{L}, \mathcal{H})$ is composed of all of the*

facilities that are closer to j' than i' .

Proof. (See appendix.) □

The following corollary formalizes the conditions for a demand point and facility to form a bottleneck pair.

Corollary 1. *If i' and j' form a post-interdiction bottleneck pair, then either 1) i' is hardened, or 2) there are at least r located facilities that are closer to j' than i' .*

Proof. (See appendix.) □

5.3.2 Single-Level Model

This special structure, described by Proposition 1 and Corollary 1, allows the integrated MFLHP model to be modeled as a single-level model. By Corollary 1, we can remove the inner maximization problem; which reduces the integrated MFLHP model to the following problem:

$$\begin{aligned}
 U^* = \min \quad & \max_{j \in \mathcal{J}} D(\mathcal{L}_j^r(\mathcal{H})), \\
 & \mathcal{L} \subseteq \mathcal{I} \\
 & \mathcal{H} \subseteq \mathcal{L} \\
 & \chi(\mathcal{L}, \mathcal{H}) \leq b
 \end{aligned} \tag{5.3}$$

where $\mathcal{L}_j^r(\mathcal{H})$ is either the set of r closest facilities to j or the set of facilities closer than the closest hardened facility, depending on which set has a smaller cardinality. The min-max structure of (5.3) can be formulated as a single-level mixed-integer program (MIP), as shown in Section 5.3.3.

The following is an explanation of why Model 5.3 is equivalent to Model 5.1: Model (5.3) assumes that each demand point j is a post-interdiction bottleneck demand point and is assigned to its *post-interdiction bottleneck facility*. In reality, not every demand point will be a post-interdiction bottleneck demand point. After disruptions, demand points that are not bottleneck demand points

will not have to be assigned to their bottleneck facility. Thus, a demand point that is not a post-interdiction bottleneck can be assigned to a facility that is closer to its post-interdiction bottleneck facility. However, because of the min-max structure of the integrated MFLHP model, the only demand points that affect the calculation of U^* are the bottleneck demand points. Thus, Model (5.3) is equivalent to Model 5.1.

5.3.3 MIP Model

Model (5.3) has the min-max structure of a bottleneck problem (Hochbaum and Shmoys, 1986) and can thus be formulated as a single-level MIP. This single-level MIP is based on Corollary 1, which specifies the requirements for a post-interdiction bottleneck pair. Recall that if i and j form a post-interdiction bottleneck pair, then $\delta^* = \phi_{ij}$. In the single-level MIP, each demand point is treated as a post-interdiction bottleneck demand point and is assigned to its post-interdiction bottleneck facility, as in Model 5.3. Let W_{ij} be the post-interdiction bottleneck pair assignment variable that is 1 if i and j form a post-interdiction bottleneck pair and 0 otherwise. Let X_i be a binary variable that is 1 if a facility is located but not hardened at i and 0 otherwise and Z_i be a variable that is 1 if a facility at i is located and hardened and 0 otherwise. The cost of locating a facility at i is f_i and the cost of locating and hardening a facility at i is $(f_i + g_i)$. (Note that because of the way the location and hardening costs are defined, in an optimal solution $X_i + Z_i \leq 1$.)

A MIP formulation of the integrated MFLHP model is:

$$\min U \quad (5.4a)$$

$$\text{s.t. } U \geq \phi_{ij}W_{ij} \quad \forall i \in \mathcal{I}, j \in \mathcal{J} \quad (5.4b)$$

$$(r+1)W_{ij} \leq (r+1)Z_i + \sum_{i': \phi_{i'j} \leq \phi_{ij}} X_{i'} \quad \forall j \in \mathcal{J}, i \in \mathcal{I} \quad (5.4c)$$

$$\sum_{i \in \mathcal{I}} W_{ij} = 1 \quad \forall j \in \mathcal{J} \quad (5.4d)$$

$$W_{ij} \leq X_i + Z_i \quad \forall i \in \mathcal{I}, j \in \mathcal{J} \quad (5.4e)$$

$$\sum_{i \in \mathcal{I}} f_i X_i + \sum_{i \in \mathcal{I}} (f_i + g_i) Z_i \leq b \quad (5.4f)$$

$$X_i, Z_i \in \{0, 1\} \quad \forall i \in \mathcal{I} \quad (5.4g)$$

$$W_{ij} \in \{0, 1\} \quad \forall i \in \mathcal{I}, j \in \mathcal{J} \quad (5.4h)$$

The objective (5.4a), in conjunction with Constraints (5.4b), is to minimize the post-interdiction radius. Constraints (5.4c) model the requirement that i and j can only form a post-interdiction bottleneck pair if facility i is hardened or if r unhardened facilities are located closer to j than i . Constraints (5.4d) require that every demand point form a post-interdiction bottleneck pair with one facility. Constraints (5.4e), although not necessary because of the presence of Constraints (5.4c), tighten the LP relaxation. Constraint (5.4f) requires that the amount spent on location and hardening must be within a budget. Constraints (5.4g)–(5.4h) specify bounds on the variables.

An alternate MIP formulation of the integrated MFLHP model was also tested but the run times of the alternate formulation were not significantly better (see Appendix C.2).

5.4 Solution Methodology

Model (5.4) can be solved using an off-the-shelf MIP optimizer such as CPLEX. However, because of the bottleneck structure of the integrated MFLHP model, we chose to use a binary search algorithm. Hochbaum and Shmoys (1986) showed that all bottleneck problems can be solved by solving a series of auxiliary problems within a binary search algorithm that searches over values in the set of all possible radii. These auxiliary problems can be thought of as inverses of their

corresponding bottleneck problem. Specifically, this auxiliary problem takes a radius value as an input and outputs the cost of covering all objects within that radius.

Empirical evidence has shown that a binary search algorithm works well for the p -center problem, which is also a bottleneck (Elloumi et al., 2004). In the p -center problem, the objective is to locate p facilities to minimize the radius δ^* . The auxiliary problem for the p -center problem is the set-cover problem with unitary costs. If some radius δ is given as an input to the set-cover problem, the set-problem outputs how many facilities must be located, i.e., the cost, so that all demand points are covered within δ . Let $p^*(\delta)$ be the optimal number of facilities needed to cover all demand points within δ . If $p^*(\delta) \geq p$, then $U \leq \delta^*$, and δ is a new lower bound. If $p^*(\delta) < p$, then $U \geq \delta^*$, and δ is a new upper bound. Thus, a binary search can be performed over all values of δ to find δ^* . Binary search has been shown to be an effective solution method for the p -center problem because the set-cover problem with unitary costs is easier to solve than the p -center problem. The set-cover problem is easier to solve because it has less variables and has a tighter LP relaxation.

The binary search algorithm for the p -center problem can be modified for the integrated MFLHP. The main extension is that the auxiliary problem is different. In this paper, we solved the integrated MFLHP using a binary search algorithm with a modified auxiliary problem along with upper and lower bounds and a heuristic.

5.4.1 Auxiliary Problem

To use a binary search algorithm for the integrated MFLHP, the auxiliary problem must first be described. Define $\delta^{(r)}$ as the radius for the auxiliary problem. (Note that $\delta^{(r)}$ is now a parameter and not a variable, as it was in Model (5.4).) To evaluate whether a particular $\delta^{(r)}$ is above or below the optimal post-interdiction radius, the set-cover problem with location and hardening (SCP-LH) is used:

$$\text{SCP-LH}(\delta^{(r)}) \quad \min \quad \sum_{i \in \mathcal{J}} f_i X_i + \sum_{i \in \mathcal{J}} (f_i + g_i) Z_i \quad (5.5a)$$

$$\text{s.t.} \quad (r+1) \sum_{i: \phi_{ij} \leq \delta^{(r)}} Z_i + \sum_{i: \phi_{ij} \leq \delta^{(r)}} X_i \geq r+1 \quad \forall j \in \mathcal{J} \quad (5.5b)$$

The SCP-LH minimizes the cost required for every demand point to have a post-interdiction assignment distance less than or equal to $\delta^{(r)}$. The objective (5.5a) is to minimize the total cost of location and hardening. Constraints (5.5b) require that for each demand point j , either $r+1$ facilities within $\delta^{(r)}$ of j must be located or at least one facility within $\delta^{(r)}$ of j must be hardened.

5.4.2 Binary Search Algorithm

The binary search algorithm for the integrated MFLHP is similar to the binary search algorithm for the p -center problem. In addition to the modified auxiliary problem, we also add a heuristic for the auxiliary problem and use a polynomial algorithm to obtain bounds for the integrated MFLHP. The binary search algorithm is described in Appendix C.4.

Before starting the binary search, we attempt to find good upper and lower bounds for the integrated MFLHP to reduce the search space of the algorithm. In particular, we apply the binary search algorithm to the linear-programming relaxation of the auxiliary problem (5.5). Let the binary search algorithm that uses the linear-programming relaxation of the auxiliary problem be called the *relaxed binary search algorithm*. This idea was first employed by Elloumi et al. (2004) for the p -center problem.

We compute a lower bound for the MFHLP, by applying the binary search algorithm in Section 5.4.2, using the linear-programming relaxation of Model (5.5). The optimal radius returned by the binary search is a lower bound to the integrated MFHLP. We also added an additional step to the binary search algorithm so that it also returns an upper bound to the integrated MFHLP: 5a) Compute an integer-feasible solution to the integrated MFHLP from solution (X^*, Z^*) by applying the auxiliary problem heuristic in Appendix C.5. However, let $r_i = X_i^*$, $s_i = Z_i^*$, and $t_i = X_i^* + Z_i^*$.

There are other ways to obtain bounds for the integrated MFLHP. The linear-programming

lower bound (LP) can be obtained by solving the linear-programming relaxation of the MIP model (5.4). The partial relaxation lower bound (PR) can be obtained by solving the MIP model (5.4) with only X_i and Z_i relaxed for all i . In our experimentation, we found that BS lower bound required much less run time than the PR lower bound and yet the BS lower bounds were reasonably close to the PR lower bounds.

5.5 Special Cases

The MFLHP model has several special cases. In certain contexts, these special cases may be useful to decision makers. For example, when an analyst must recommend which facilities to harden among a set of existing facilities, a model that only involves hardening would be most appropriate. These special cases will also be used to compare against the integrated model. These special cases of the integrated model are: 1) location only, 2) hardening only, 3) location-then-hardening, and 4) location-then-hardening with ideal proportion.

5.5.1 Location and Hardening

One special case of the MFLHP is when there is no hardening. In this case it is not hard to show that, as in Proposition 1, the interdicator always attacks the closest facilities to the bottleneck demand point. Medal et al. (2011a) provided the most recent study of this problem for the case where $f_i = 1$ for all $i \in \mathcal{I}$. We call this method the *location-only-with-disruptions* (LOWD) method.

Another special case of the MFLHP is the problem of allocating hardening resources amongst a set of existing facilities. A MIP model for this problem is described in Appendix C.6.

5.5.2 Non-Integrated Methods

There are also several non-integrated methods that can be used to solve the MFLHP. In these methods, the location decisions are made and then the hardening decisions are made; thus, these methods are sequential. Although these methods do not guarantee an optimal solution to MFLHP, we use them to compare against the integrated MFLHP model. We examine two sequential methods: location-then-hardening and location-then-hardening with the ideal proportion of the budget allocated to location.

5.5.2.1 Location Then Hardening (LTH)

The first step in the *location-then-hardening* (LTH) method is to locate facilities using a model that minimizes the post-interdiction radius after r facility disruptions. A set of locations can be found using the r -all-neighbor p -center model (Medal et al., 2011a), which can be modified for the case where the cost of locating facilities is not the same for every facility. Let L be the proportion of the total budget allocated to locating facilities. That is, if b is the total budget, bL can be spent on locating facilities and $b(1 - L)$ can be spent on hardening facilities. Let X^* be the optimal vector of location variables prescribed by the r -all-neighbor p -center model with a budget of bL . Let $\mathcal{L}(X^*) = \{i : i \in \mathcal{I}, X_i^* = 1\}$. Then, choose the optimal set of facilities to harden using the hardening model in Appendix C.6, with $\mathcal{L} = \mathcal{L}(X^*)$ and budget $b(1 - L)$.

The LTH method provides flexibility. The decision maker can use different values of r for the location and hardening stages. The standard method is to use the same value of r for both location and hardening. However, the *location-without-disruptions-then-harden method* (LWDTH) is where $r_{LOC} = 0$ is used for the location stage and $r_{HARD} > 0$ is used for the hardening stage.

A drawback of the sequential LTH and LWDTH methods is that it can give an infeasible solution for a problem instance even if that problem instance is feasible to the integrated MFLHP. This happens when $B \times L$ is so small that it is not sufficient to locate at least $(r + 1)$ facilities. Since the interdictor can destroy all of the facilities, the location stage is infeasible.

5.5.2.2 Location Then Hardening: Ideal Proportion (LTHI)

One of the drawbacks of using the LTH method is that it is difficult to find the right value of L . Thus, the solution prescribed by the LTH method in Section 5.5.2.1 can be improved by using the value of L that minimizes the post-interdiction radius, namely \hat{L} . We call \hat{L} the *ideal proportion*. The *location-then-hardening method with ideal proportion* (LTHI) uses the sequential method with \hat{L} . The *location-without-disruptions-then-harden method with ideal proportion* (LWDTHI) is where $r_{LOC} = 0$ is used for the location stage and $r_{HARD} > 0$ is used for the hardening stage.

To find the ideal proportion, we simply enumerate $L = 0.0, 0.1, \dots, 1.0$ and choose the value

for which the sequential method returns the smallest post-interdiction radius. Although this is a crude method, we chose it because pilot experiments showed that the post-interdiction radius is not a smooth function of L . Thus, it is not straightforward to find \hat{L} . However, experimental results showed that the integrated method produces significantly better solutions than the LTHI method. Thus, we decided it was not worthwhile to try to develop a more efficient method to obtain \hat{L} .

5.6 Experimental Results

In this section we present empirical results that demonstrate that 1) the integrated MFLHP produces better solutions than non-integrated methods and that 2) the integrated method is able to solve reasonably-sized problems. We define a reasonably-sized problem as one that has at least enough nodes to represent all of the major cities in a country such as the United States. In Section 5.6.1 we demonstrate that the integrated model produces significantly better solutions than non-integrated models. In Section 5.6.2 we demonstrate that there is substantial benefit in considering hardening as an option to reduce the vulnerability of a system. In Section 5.6.3, we analyze the tradeoff between post-interdiction performance and performance without disruptions using the integrated model. To assess the computational feasibility of the integrated method, we ran a series of experiments. The results of these experiments, described in Section 5.6.4, show that the integrated model is comparable with non-integrated methods and that the integrated model is able to solve instances with up to 1323 nodes.

All experiments were run on a compute node on the Arkansas High Performance Computing Cluster. The node has 2 Xeon X5670 Intel processors, which each have 6 cores and a clock speed of 2.93GHz. The total of 12 cores share 24GB of memory. Computations were done on a 64-bit Linux operating system. All of the MIP and LP models, including the set-cover problems, were solved using CPLEX v12.1 Parallel MIP Optimizer with 12 parallel threads and default settings. The binary search algorithm was programmed in Java using CPLEX Concert technology.

The following notation will be used in the rest of this section. Let $X_{(r)}^*$ be the optimal location solution generated from a model that models the disruption of r facilities (see Section 5.5.1). Let Z_I^* be the optimal hardening solution generated from the integrated MFLHP model. Let $Z_{S(L)}^*$ be

the optimal hardening solution generated from the LTH method described in Section 5.5.2.1. Let $Z_{S(\hat{L})}^*$ be the optimal hardening solution generated from the LTH method with optimal proportion described in Section 5.5.2.2. Let $f_{(r)}(X, Z)$ be the post-interdiction radius for the location decision variable vector X and hardening decision variable vector Z . The function $f_{(r)}(X)$ is the post-interdiction radius for location decision variable vector X without any hardening. In this case all of the budget is allocated to location.

5.6.1 Benefit of Integration

In this section we report empirical results that indicate that there is a measurable benefit in integrating the location and hardening decisions. Let

$$\gamma(L) = \frac{f_{(r)}(X_{(r)}^*, Z_{S(L)}^*) - f_{(r)}(XZ_{(r)}^*)}{f_{(r)}(XZ_{(r)}^*)} \quad (5.6)$$

be the relative objective function increase for not integrating the location and hardening decisions. In other words, this is the relative objective function increase incurred when the LTH method is used in place of the integrated MFLHP model.

Table 5.1 shows the value of γ for various instances of the sw55 dataset. Each row shows the value of $\gamma(L)$ for different values of L . The 9th column gives the optimal proportion prescribed by the integrated MFLHP model,

$$\lambda^* = \frac{\sum_{i \in \mathcal{I}} f_i X_i^*}{\sum_{i \in \mathcal{I}} f_i X_i^* + \sum_{i \in \mathcal{I}} g_i Z_i^*}, \quad (5.7)$$

in which $X_i^* = 1$ if a facility is located at i in the optimal solution and 0 otherwise. The 10th column, labeled $\hat{L} - \lambda^*$, gives the *deviation in proportion*, which is the difference between the optimal proportion prescribed by the integrated MFLHP model and the ideal proportion prescribed by the LTH method. The ideal proportion, \hat{L} , is the value of L that minimizes $\gamma(L)$. The last column gives the value of γ when \hat{L} is used as the proportion of the budget allocated to location. Columns (5)–(7) of the table, displaying the results for $\gamma(L)$, show that γ is neither an increasing nor a decreasing function of L . Column (9), labeled $\hat{L} - \lambda^*$, shows that the ideal proportion was

always smaller than the optimal proportion. Column (10), labeled $\gamma(\hat{L})$, shows that the relative objective function increase for not hardening is significant even when the ideal proportion is used. The average relative objective function increase was 65%. The results show that $\gamma(\hat{L})$ increases as r increases and decreases as H increases, which indicates that the integrated method is most valuable when the system is subject to large disruptions and the cost of hardening is relatively low.

Table 5.1: Relative objective function increases for not integrating location and hardening decisions for the sw55 dataset

No. (1)	B (2)	H (3)	r (4)	$\gamma(L)$			λ^* (8)	$\hat{L} - \lambda^*$ (9)	$\gamma(\hat{L})$ (10)
				$L = 0.3$ (5)	$L = 0.6$ (6)	$L = 0.9$ (7)			
1	0.3	0.5	1	2.11	0.99	0.99	0.67	0.12	0.99
2	0.3	0.5	2	2.68	0.99	1.31	0.67	0.12	0.99
3	0.3	0.5	3	3.14	1.31	1.44	0.67	0.12	0.92
4	0.3	0.5	6	0.00	2.68	1.44	0.67	0.00	1.18
5	0.3	0.5	9	0.00	0.00	2.44	0.67	0.00	1.91
6	0.3	1	1	1.21	0.41	0.98	0.50	0.00	0.41
7	0.3	1	2	1.61	0.41	1.21	0.50	0.00	0.41
8	0.3	1	3	1.94	0.64	1.21	0.50	0.00	0.41
9	0.3	1	6	0.00	1.61	2.20	0.50	0.00	1.02
10	0.3	1	9	0.00	0.00	2.20	0.50	0.00	1.07
11	0.3	2	1	0.95	0.43	0.85	0.63	0.10	0.24
12	0.3	2	2	1.05	0.33	0.77	0.33	0.16	0.29
13	0.3	2	3	1.30	0.36	2.47	0.33	0.00	0.23
14	0.3	2	6	0.00	1.05	4.27	0.33	0.00	0.68
15	0.3	2	9	0.00	0.00	9.13	0.33	0.00	0.73
16	0.3	3	1	0.85	0.38	0.67	0.63	0.07	0.20
17	0.3	3	2	0.69	0.12	0.46	0.57	0.28	0.12
18	0.3	3	3	0.90	0.43	1.86	0.25	0.00	0.17
19	0.3	3	6	0.00	0.69	3.34	0.25	0.00	0.43
20	0.3	3	9	0.00	0.00	7.35	0.25	0.00	0.69
Average							0.49	0.05	0.65
Std. dev.							0.16	0.08	0.45

5.6.2 Benefit of Hardening

In this section we analyze the benefit of hardening by reporting the objective increase incurred when hardening is not included in a model. Let

$$\varepsilon^I = \frac{f_{(r)}(X_{(r)}^*) - f_{(r)}(XZ_{(r)}^*)}{f_{(r)}(XZ_{(r)}^*)} \quad (5.8)$$

be the relative objective function increase for not including hardening as an option when designing a system. In other words, this is the proportional increase in the maximum post-interdiction radius incurred when using the LOWD model (see Section 5.5.1) in place of the integrated MFLHP model.

Table 5.2 presents values of the optimal proportion, λ^* , and ε^I for the sw55 dataset with different values of B , r , and H .

Columns (4)–(7) show the value of λ^* for different values of H . The optimal proportion, λ^* , often equals $\frac{f_i}{f_i + g_i} = \frac{f_i}{f_i + Hf_i} = \frac{1}{1+H}$, especially when H is small. Thus, the proportion of the budget allocated to location is the same as the ratio between the cost of locating a facility and the cost of locating and hardening a facility. This ratio occurs when every located facility is hardened. Note that $\lambda^* \geq \frac{1}{1+H}$, which may be due to the fact that if a value $\lambda^0 < \frac{1}{1+H}$ of the budget is used to locate facilities, then the total cost of locating and hardening facilities will be less than the total budget (proof in Appendix C.1). In some cases, the optimal proportion is 1.0, meaning that none of the budget is allocated to hardening.

Columns (8)–(11) show the value of ε^I , the relative objective function increase for not considering hardening, for different values of H . The table shows that, for all of the instances, as H increases, the relative objective function increase for not considering hardening usually decreases. This result is intuitive because as H increases, hardening becomes too expensive and it is better to allocate more of the budget to location. The table also shows that as r increases, the relative objective function increase for not considering hardening is usually higher, indicating that it is more important to consider hardening when the system is subject to a large number of facility disruptions.

Table 5.2: Comparison of LOWD model and integrated MFLHP solutions for the sw55 dataset

No. (1)	B (2)	r (3)	λ^*				ε^l			
			$H = 0.5$ (4)	$H = 1$ (5)	$H = 2$ (6)	$H = 3$ (7)	$H = 0.5$ (8)	$H = 1$ (9)	$H = 2$ (10)	$H = 3$ (11)
1	0.1	1	0.67	1.00	1.00	1.00	0.28	0.00	0.00	0.00
2	0.1	2	0.67	0.50	0.60	1.00	0.70	0.33	0.33	0.33
3	0.1	3	0.67	0.50	0.33	0.57	0.92	0.50	0.13	0.50
4	0.1	6	0.67	0.50	0.33	0.25	1.63	1.06	0.55	0.55
5	0.1	9	0.67	0.50	0.33	0.25	1.86	1.24	0.68	0.68
6	0.2	1	0.67	0.70	0.67	0.73	0.51	0.27	0.01	0.01
7	0.2	2	0.67	0.50	0.82	0.77	1.03	0.64	0.21	0.21
8	0.2	3	0.67	0.50	0.33	0.73	1.19	0.77	0.31	0.14
9	0.2	6	0.67	0.50	0.33	0.25	2.22	1.60	0.92	0.50
10	0.2	9	0.67	0.50	0.33	0.25	3.41	2.57	1.63	1.06
11	0.3	1	0.67	0.50	0.65	0.63	1.20	0.56	0.38	0.25
12	0.3	2	0.67	0.50	0.53	0.65	1.95	1.10	0.64	0.36
13	0.3	3	0.67	0.50	0.33	0.31	2.11	1.21	0.73	0.43
14	0.3	6	0.67	0.50	0.33	0.25	3.17	1.97	1.33	0.92
15	0.3	9	0.67	0.50	0.33	0.25	3.96	2.53	1.76	1.28
Average			0.67	0.55	0.48	0.52	1.74	1.09	0.64	0.48
			Grand average				0.99			

5.6.3 Relative Objective Function Increases for Only Considering a Single Objective

In this section we demonstrate empirically that there is a significant penalty for only considering a single objective. In particular, the results show that the model that does not consider facility disruptions produces solutions with a high post-disruption radius. Conversely, since the integrated MFLHP model does not consider performance without disruptions, it produces solutions that perform sub-optimally when facilities do not fail. We measure two types of relative objective function increases. First, we measure the relative objective function increase—in terms of the increase in post-interdiction radius—incurred when the post-interdiction radius is not included as an objective. We also measure the relative objective function increase incurred when the non-disruption distance is not included as an objective. Table 5.3 lists all of the metrics that we use in our analysis.

The following notation is used in the Table 5.3. Let $f_{(0)}(X)$ be the non-disruption radius given the location decision variable vector X . Let $X_{(0)}^*$ be the optimal location solution generated by the

b -center problem, which minimizes the non-disruption radius.

Table 5.3: List of metrics

Metric	Equation	Method	Relative objective function increase incurred for...
$\eta_{r,0}^I$	$\frac{f_{(r)}(X_{(0)}^*) - f_{(r)}(XZ_{(r)}^*)}{f_{(r)}(XZ_{(r)}^*)}$	p -center	the maximum post-interdiction radius is not included as an objective
$\eta_{r,0}^{I,T(\hat{L})}$	$\frac{f_{(r)}(X_{(0)}^*, Z_{T(\hat{L})}^*) - f_{(r)}(XZ_{(r)}^*)}{f_{(r)}(XZ_{(r)}^*)}$	LWDTHI	the maximum post-interdiction radius is not included as an objective
$\eta_{0,r}^I$	$\frac{f_{(0)}(XZ_{(r)}^*) - f_{(0)}(X_{(0)}^*)}{f_{(0)}(X_{(0)}^*)}$	integrated MFLHP	the non-disruption radius is not included as an objective
$\eta_{0,r}^{S(\hat{L})}$	$\frac{f_{(0)}(X_{(0)}^*, Z_{S(\hat{L})}^*) - f_{(0)}(X_{(0)}^*)}{f_{(0)}(X_{(0)}^*)}$	LTHI	the non-disruption radius is not included as an objective
$\eta_{0,r}^{T(\hat{L})}$	$\frac{f_{(0)}(X_{(0)}^*, Z_{T(\hat{L})}^*) - f_{(0)}(X_{(0)}^*)}{f_{(0)}(X_{(0)}^*)}$	LWDTHI	the non-disruption radius is not included as an objective
$\eta_{0,r}^L$	$\frac{f_{(r)}(X_{(r)}^*) - f_{(0)}(X_{(0)}^*)}{f_{(0)}(X_{(0)}^*)}$	LOWD	the non-disruption radius is not included as an objective

Table 5.4 shows summary statistics of the metrics listed in Table 5.3 for the sw55 dataset across 20 instances, varying H and r but keeping $B = 0.3$. The first two columns compare the relative objective function increase for not considering the post-disruption radius for two methods: p -center and LWDTHI. Since the p -center does not consider the non-disruption radius at all, its optimal solutions have, on average, a 332% higher post-disruption radius than optimal solutions to the MFLHP. Since the LWDTHI model only considers the non-disruption radius in the location stage, its optimal solutions have, on average, a 139% higher post-disruption radius than optimal solutions to the MFLHP. The last four columns of the table compare the relative objective function increase for not considering the non-disruption radius for four methods: integrated MFLHP model, LTHI, LWDTHI, and LOWD. The LWDTHI model has the lowest relative objective function increase (0.40) because it optimizes for the non-disruption radius in the location stage and thus produces

solutions with a small non-disruption radius. Surprisingly, the MFLHP has the next lowest relative objective function increase (0.99). In comparison, this relative objective function increase is much lower than the relative objective function increase incurred by the p -center model for not considering post-disruption radius (3.32). This indicates that if only one objective is considered in the problem, the post-disruption objective should be included. However, since the two objectives are conflicting, a bi-objective model is most appropriate.

Table 5.4: Comparison of relative objective function increase results for sw55 dataset

	$\eta_{r,0}^I$	$\eta_{r,0}^{I,T(\hat{L})}$	$\eta_{0,r}^I$	$\eta_{0,r}^{S(\hat{L})}$	$\eta_{0,r}^{T(\hat{L})}$	$\eta_{0,r}^L$
Average	3.32	1.39	0.99	2.06	0.40	1.71
Minimum	0.43	0.25	0.27	1.44	0.22	1.44
Maximum	12.02	3.96	1.77	3.68	0.64	2.62
Std. dev.	3.01	0.99	0.54	0.66	0.17	0.47

5.6.4 Computational Experimentation

This section describes computational testing of the models presented in this paper. The computational testing includes several datasets from the facility location literature. Most of the datasets are motivated from real data and most of them have been used in other studies. The datasets vary in size (49–1323 nodes), distance metric used (e.g., Euclidean, great circle, etc.), and whether demand weights and facility location costs are homogeneous or nonhomogeneous. When a dataset has demand weights, then ϕ_{ij} is the weighted distance; otherwise it is the distance. All of the datasets are described in Appendix C.7.

The computational experimentation involved varying several parameters of the model in order to measure the effect of these parameters on the computation time. Table 5.5 gives a list of all of the parameters that were varied. This set of parameters was used throughout the rest of the experiments, unless otherwise mentioned. First, the budget b for a particular instance is a percentage of the cost of locating every facility. Therefore, letting B be the budget multiplier, $b = B \sum_{i \in \mathcal{J}} f_i$.

Second, the cost of hardening a facility located at i is a multiple of the cost of locating a facility at i . That is, $g_i = Hf_i$, where H is the hardening cost multiplier. The number of facility disruptions, r , was also varied. To facilitate comparison across datasets, $r = 1$ and $r = 2$ were considered for all datasets. We also considered the number of disruptions to be a percentage of the number of facility locations. When the location and hardening decisions are made in sequence, such as in Section 5.5.2.1, a certain amount of the total budget is allocated to location and the remainder is allocated to hardening existing facilities. Thus, the amount allocated to location is $L(B\sum_{i \in \mathcal{I}} f_i)$ and the amount allocated to hardening is $(1 - L)(B\sum_{i \in \mathcal{I}} f_i)$.

Table 5.5: Parameter values for experiments

Parameter	Description	Values used
$ \mathcal{I} $	number of facility locations	depends on dataset
$ J $	number of demand points	depends on dataset
B	budget multiplier	0.1, 0.2, 0.3
H	hardening cost multiplier	0.5, 1, 2, 3
r	number of facility disruptions	1, 2, 0.05 $ \mathcal{I} $, 0.1 $ \mathcal{I} $, 0.15 $ \mathcal{I} $
L	proportion of budget allocated to location	0.3, 0.6, 0.9

5.6.4.1 Integrated Method vs. Sequential Method

An important consideration in deciding whether the integrated model is the superior method for solving the MFLHP is how the computation times for the integrated method compare with the computation times for other methods. For instance, if the sequential method requires much less computation time than the integrated method, then it may be preferred in some contexts even if it provides inferior solutions. The experimental results in this section show that the integrated method is significantly better than the sequential method with the ideal proportion, and is comparable to the sequential method with an arbitrary proportion.

To obtain results for our comparison, we solved problem instances using integrated and sequential methods and compared the results. Specifically, we solved the instances using three methods:

the sequential method, $S(L)$ described in Section 5.5.2.1; the sequential method with the ideal proportion, $S(\hat{L})$, described in Section 5.5.2.2; and the integrated method, I , solved using the binary search algorithm described in Section 5.4.2. The integrated method binary search algorithm was used without the bounds and heuristics described in Section 5.4.2 to provide a fair comparison with the the sequential methods, which do not use bounds or heuristics.

Table 5.6 compares the run times of the LTH method and the integrated MFLHP model. Each row of the table shows summary statistics for the sequential method (taken over all values of B, H, r , and L) and the integrated method when applied to a particular dataset. The rows are ordered by average run time for the sequential method.

Table 5.6: Summary statistics for sequential and integrated run times (s)

Dataset	$S(L)$				I			
	Avg.	Min.	Max.	Std. dev.	Avg.	Min.	Max.	Std. dev.
sw55	0.09	0.07	0.11	0.01	0.12	0.09	0.17	0.02
lor402a	0.18	0.14	0.27	0.04	0.22	0.18	0.31	0.03
lor300a	0.19	0.14	0.26	0.04	0.23	0.17	0.38	0.05
lor200	0.20	0.14	0.28	0.05	0.23	0.17	0.38	0.05
lor100	0.20	0.14	0.26	0.04	0.24	0.17	0.38	0.06
b500	2.68	1.22	3.44	0.68	6.37	1.90	26.80	5.14
b700	3.54	2.37	4.42	0.69	7.42	2.86	25.79	5.92
b600	3.93	2.81	5.62	0.70	16.19	3.18	76.00	13.95
b800	9.97	7.48	12.51	1.47	21.12	7.32	71.02	15.35
b900	21.80	17.45	25.75	2.45	34.65	9.15	127.34	29.62
lor818	31.16	15.96	40.46	7.90	193.01	10.11	1998.52	395.38

The results in Table 5.6 show that for smaller datasets the run time for the integrated method is comparable to that of the sequential method; however, the table shows that the integrated method takes longer for datasets with a large number of nodes.

5.6.4.2 Computational Performance of Binary Search Algorithm

Another important consideration in deciding whether the integrated method is preferred over other methods is the computation time required to solve problem instances. Specifically, it is important

to consider these two questions: 1) is the integrated method able to solve problem instances that would be of interest to decision makers? and 2) is the integrated method able to solve these problem instances within an amount of time that is satisfactory to the decision makers? The experimental results in this section show that the answer to both of these two questions is yes.

To obtain the run times analyzed in this section, we solved the integrated MFLHP model for several problem instances of the r11323 dataset. The instances were obtained from the r11323 dataset by varying the parameters B , H , and r . The instances were solved using the binary search algorithm described in Section 5.4.2.

In a preliminary set of experiments, the run time for the binary search algorithm was several orders of magnitude smaller than the run time for the MIP model. Thus, we do not report additional results for the MIP model.

Experimental results for the binary search algorithm performance are shown in Table 5.7. The rows each contain the results for an instance of the r11323, which was the largest dataset for which the binary search was able to solve all instances to optimality. Columns (5)–(11) show the initial percentage gap between the lower bound and the final optimal solution (5), the initial percentage gap between the upper bound and the final optimal solution (6), the initial total percentage gap (7), the number of times the auxiliary problem is solved to optimality (8), the presolve run time in which the initial bounds are obtained (9), the run time of the binary search algorithm (10), and the total run time (11). In some of the rows, the initial upper bound is ∞ , which means that the relaxed binary search algorithm did not find an upper bound to the integrated MFLHP.

The results shown in Table 5.7 indicate that the binary search algorithm can solve large problems in a reasonable amount of time. Since deciding where to locate facilities is a strategic decision, we define a reasonable amount of time as no more than 24 hours, allowing a decision maker to do some “what-if” analysis. Although not all of the results are displayed, these results are representative for the remaining experiments on the r11323 dataset. The lower bound percentage gaps

Table 5.7: Run times (s) of binary search algorithm for r11323 dataset

No. (1)	B (2)	H (3)	r (4)	Initial gaps			# solved to opt. (8)	Run time (s)		
				LB (5)	UB (6)	LB+UB (7)		Pre-solve (9)	Solve (10)	Total (11)
1	0.1	0.5	1	0.20	0.36	0.56	11	267	99	367
2	0.1	0.5	67	0.00	0.36	0.36	9	132	76	208
3	0.1	0.5	199	0.00	∞	∞	10	159	100	259
4	0.1	3	1	0.01	1.20	1.20	13	173	116	289
5	0.1	3	67	0.03	∞	∞	12	62	184	247
6	0.1	3	199	0.02	∞	∞	11	58	80	137
7	0.3	0.5	1	0.28	0.06	0.34	8	1111	324	1435
8	0.3	0.5	67	0.01	0.06	0.07	6	598	234	832
9	0.3	0.5	199	0.00	0.00	0.00	1	517	39	556
10	0.3	3	1	0.02	1.14	1.16	11	616	201	817
12	0.3	3	67	0.03	0.04	0.06	7	158	62	295

are usually good, ranging from 0 to 28%. In fact, the relaxed binary search algorithm found the optimal solution to the integrated MFLHP in instances 2, 3, and 9. The heuristic and initial bounds are effective in reducing the number of times the auxiliary problem is solved to optimality. In the worst case performance of the binary search algorithm, 21 auxiliary problems must be solved to optimality; however, these results show that the auxiliary problem needed to be solved to optimality between 1 and 13 times. Most of the problems solve in a reasonable amount of time, from 137 to 1435 seconds.

5.7 Managerial Insights

In addition to determining the best method for the MFLHP, there are other important questions about location and hardening. For example, how do the solutions produced by the integrated MFLHP model perform when no facilities fail? Also, how does the post-interdiction radius objective change with the budget? Here we use the models described in this paper to provide answers to these questions through empirical results.

5.7.1 Relative Objective Function Increases for Only Considering a Single Objective

The integrated MFLHP model only optimizes one objective: the post-interdiction radius. However, in most situations facility disruptions are rare. Thus, a manager may be more interested in the non-disruption radius. Consequently, we raise an important question: How do integrated MFLHP solutions perform when no facility disruptions occur? Experimental results showed that integrated MFLHP solutions perform well compared to the post-interdiction performance of solutions produced by a model that doesn't consider disruptions.

For our analysis we used two models to solve problem instances: the integrated MFLHP model and the b -center model, which does not consider disruptions. For each model, we recorded two values: the post-interdiction radius and the non-disruption radius. We then used these values to compute two metrics from Table 5.3, the post-interdiction relative objective function increase,

$$\eta_{r,0}^I = \frac{f_{(r)}(X_{(0)}^*) - f_{(r)}(XZ_{(r)}^*)}{f_{(r)}(XZ_{(r)}^*)}, \quad (5.9)$$

and the non-disruption relative objective function increase,

$$\eta_{0,r}^I = \frac{f_{(0)}(XZ_{(r)}^*) - f_{(0)}(X_{(0)}^*)}{f_{(0)}(X_{(0)}^*)}. \quad (5.10)$$

Experiments were run for most of the datasets in Table B.1.

Table 5.8 shows summary statistics for the post-interdiction and non-disruption relative objective function increases. Each row shows summary statistics, taken across combinations of B , H , and r , for a particular dataset.

Table 5.8 shows that the average post-interdiction relative objective function increase (4.4) is 5.71 times higher than the average non-disruption relative objective function increase (0.77). Thus, if only one objective is considered it should be the post-disruption radius. However, since the objectives are conflicting, a bi-objective model is most appropriate.

Table 5.8: Summary statistics for comparison of LOWD method and integrated MFLHP model solutions

Dataset	$\eta'_{r,0}$				$\eta'_{0,r}$			
	Avg.	Min.	Max.	Std. dev.	Avg.	Min.	Max.	Std. dev.
sw55	3.6	0.4	15.2	3.5	1.0	0.3	2.1	0.5
lor100	11.9	2.6	41.2	9.9	1.5	0.5	2.9	0.8
lon150	3.4	0.6	9.4	2.4	1.0	0.3	1.9	0.4
b500	1.7	0.7	3.5	0.7	0.5	0.2	1.0	0.3
b600	1.9	0.8	3.4	0.7	0.6	0.2	1.1	0.3
b700	2.3	0.7	4.7	1.0	0.6	0.3	1.1	0.3
b800	2.4	1.0	4.2	0.8	0.6	0.2	1.1	0.3
lor818	8.8	0.6	24.6	7.2	0.6	0.2	1.2	0.3
b900	3.6	2.2	5.9	0.9	0.5	0.2	1.0	0.3
ALL	4.4	0.4	41.2	3.35	0.77	0.2	2.9	0.17

5.7.2 Budget Vs. Post-Disruption Radius

The budget, b , also influences the optimal objective value of the integrated MFLHP model. Therefore, a decision maker may benefit from a tradeoff curve between b and the post-disruption radius, δ_r . This curve can be generated by solving SCP-LH(δ_r) for all values δ_r in the set of distances $\{\phi_{ij} : i \in \mathcal{I}, j \in \mathcal{J}\}$.

Figure 5.1 shows the tradeoff curves for the budget b vs. the post-interdiction radius for both the integrated MFLHP model (Int.) and the location-only-with-disruptions method (LOWD). The curves for the integrated MFLHP model terminate at $b = 3,700,000$. However, since the curves are flat after $b = 1,000,000$, we only display values up to $b = 2,500,000$ on the x-axis. The curves for the location with disruptions method are the same for Figures 5.1a and 5.1b because the location with disruptions method does not use the parameter H . Figures 5.1a and 5.1b both show that the tradeoff curves for the integrated MFLHP model and LOWD method are similar when $r = 1$. However, the curves for MFLHP and LOWD are significantly different when $r = 8$; this difference is because hardening is more useful when r is large. Figures 5.1a and 5.1b also show that the tradeoff curve for the integrated MFLHP model with $r = 8$ and $H = 0.5$ dominates the tradeoff curve for $r = 8$ and $H = 3$. This is because hardening becomes less cost-effective as H increases. When $H = 0.5$, as in Figure 5.1a, the integrated MFLHP model tradeoff curves for $r = 1$

and $r = 8$ are identical because hardening is so cheap that every facility is hardened.

All of the integrated MFLHP model curves have a similar shape: the curve has a smooth decrease from $b = 0$ to $b = 1,000,000$ and is flat for $b > 1,000,000$. This indicates that tradeoff curves, such as those shown in Figure 5.1, will be especially useful to decision makers when b is small because this is where the tradeoff between the two objectives is most pronounced.

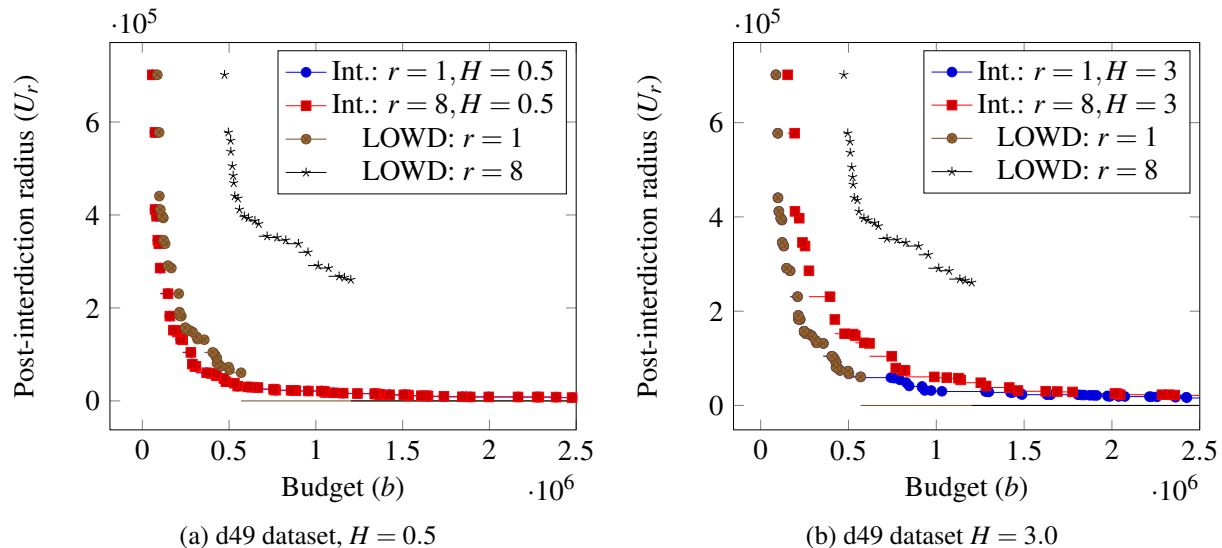


Figure 5.1: Budget vs. post-interdiction radius

5.8 Summary and Future Work

In this paper we studied the integration of facility location and hardening decisions. The main finding of this paper is that our integrated method for location and hardening facilities is better than several non-integrated methods. In particular, we found that the integrated method produced better quality solutions in a reasonable amount of time.

5.8.1 Summary

In order to perform our analysis, we developed an efficient method to model and solve the integrated minimax facility location-hardening problem (MFLHP) model and used our model to discover new insights about location and hardening. To model the problem, we first demonstrated that,

because of the problem's structure, the three-stage location-interdiction-distribution problem can be modeled as a single-stage problem. We then formulated this single-stage problem as a mixed-integer program (MIP). Rather than solving the MIP directly, we decomposed it into set-cover-like auxiliary problems and used a binary search algorithm to solve it.

In addition to the integrated model, we also described several special cases in Section 5.5. These special cases included location only without disruptions (LOWD), hardening only, location-then-hardening (LTH), and location-then-hardening with ideal proportion (LTHI). The integrated model was then compared with these special cases in a series of experiments.

Experimental results showed that the integrated model produced solutions that are of better quality than the other methods tested. Even when the ideal allocation proportion was used, the sequential method produced solutions that had a 65% higher average post-interdiction radius than the integrated method. Further, the r -all-neighbor p -center model, which only considers facility location to mitigate against facility disruptions, produced solutions that had, on average for a particular dataset, a 99% higher post-interdiction radius than the integrated MFLHP model; this shows that hardening can be a valuable strategy for reducing the vulnerability of a network.

The experimental results also showed that the integrated method is attractive from a computational standpoint. The results showed that the computational time required to solve the integrated model is comparable with the computation time of the LTH method and is much better than the computational time of the LTHI method. Also, the integrated model solved problems with up to 1323 nodes.

Managerial insights were gained through experimenting with the integrated MFLHP model. Results showed that the average relative objective function increase for not considering facility disruptions was 5.71 times higher than the average relative objective function increase for not considering non-disruption performance. This indicates that if hardening is an option, it is important for analysts to account for post-interdiction performance. The results also showed that increasing the budget for locating and hardening improves both non-disruption performance and post-interdiction performance.

5.8.2 Future Work

Although this study does provide some insights, there are several limitations and opportunities for future work. Interesting areas of future work include improving the solution method, extending the model, and studying related problems.

Although our proposed solution method does solve reasonably-sized problems in a reasonable amount of time, it would be useful to solve larger problems in less time. Analysts may wish to analyze datasets with more nodes for two reasons. First, they may want to analyze a system with a larger number of facilities. Second, they may want to reduce the amount of aggregation in the dataset; for example, rather than representing a city as a node, a dataset may represent a neighborhood as a node. Possible ways to improve the solution method include: decomposing the MIP model, decomposing the auxiliary sub problems, or finding better bounds.

There are several assumptions in our model that, if relaxed, would make the model more realistic. First, in our model a hardened facility cannot fail and a facility is either hardened or not. In reality, a facility can never be completely immune to disruptions and investing more protection resources into a facility makes it more resilient. Second, in our model the number of disruptions is known. In reality, the decision maker would have only an estimate of the number of disruptions. Third, in our model the facilities are uncapacitated. This may be a good approximation in some contexts. However, in other contexts it may be important to model facility capacity.

This research could also be extended to related problems. For example, it would be useful to study integrating location and hardening decisions for other common facility location objectives such as total weighted distance and maximum coverage. Although we did some multi-objective analysis, it would be useful to have an efficient method to generate the Pareto frontier for two or more objectives.

6 Imperfect Protection of Multi-State Supply Facilities

Abstract

Recent events such as the 2011 Tohoku earthquake and the 2010/2011 snowstorms in Western Europe have highlighted the significant impact that disruptions can have on systems such as supply chains and transportation systems. The presence of these disruptions presents a challenging problem for decision-makers: how to allocate a scarce budget to protecting elements of a system in order to minimize the effect of element failures. To aid this challenging decision, protection models can be used, which prescribe an optimal protection plan. Most of the models in the literature make two key assumptions: 1) protection actions are perfect and 2) the elements of the system have infinite throughput capacity. In this paper, these assumptions are relaxed and the notion of imperfect protection of a capacitated system is studied. In particular, a problem is studied in which protection resources are allocated to an element of the system in order to modify the probability distribution of that element's capacity. A two-stage stochastic programming model is presented and the model is solved using the L-shaped method. The results of computational experiments are presented which demonstrate the value of considering imperfect protection of capacitated systems.

6.1 Introduction

At 2:02pm EDT on August 14, 2003, a 345 kV overhead transmission line in southern Ohio tripped due to contact with a tree, causing a series of cascading failures (United States-Canada Power System Outage Task Force, 2004). By 4:13pm these cascading failures had caused many power plants to go off-line throughout the Northeast United States, leaving millions of people without power (United States-Canada Power System Outage Task Force, 2004). In all, this disruption affected 50 million people and the total economic loss is estimated at \$4 to \$10 billion (United States-Canada Power System Outage Task Force, 2004).

To address the need to mitigate against disruptions such as the 2003 blackout, researchers have developed models for allocating resources among networks to minimize the risk to the overall

network. Several researchers have studied the problem of defending a network against an attacker, often termed “defender-attacker” problems. For a survey of defender-attacker problems see, Brown and Carlyle (Brown et al., 2005b, 2006). Fan and Liu (2010) have studied how to protect a transportation network against random disruptions. Smith et al. (2007) modeled the protection of networks that are subject to an attack by an attacker that uses heuristics to allocate resources. Researchers have also examined how to protect a system of facilities against an attacker. This problem was first studied by Church and Scaparra (2007) (see also Scaparra and Church (2008a,b)). Scaparra and Church (2010) have recognized the importance of including facility capacity in protecting facilities against attacks. Alderson et al. (2011) have developed a generic solution methodology for defender-attacker problems.

In addition to protection models, researchers have also developed interdiction models, which seek to allocate resources to maximally disrupt the network, i.e., maximize the risk to the overall network. These networks can be useful in helping to disrupt adversarial networks or in identifying the critical elements of a network. For a recent survey see Wood (2011). Fulkerson and Harding (1977) were among the first to study how to interdict arcs in a network to maximally increase the length of the shortest path; they were later followed by others (Israeli and Wood, 2002). Variations on the shortest-path interdiction problem include stochastic networks (Hemmecke et al., 2003) and asymmetric information (Bayrak and Bailey, 2008). Wollmer (1964) was among the first to provide a model for interdicting a maximum-flow network. Maximum-flow interdiction was later given more rigor by Wood (1993).

Researchers have studied network interdiction and protection problems with stochastic aspects. For an up-to-date survey on stochastic network interdiction, see Morton (2011). Several authors have studied evader interdiction problems in which there is uncertainty about the evader’s objective (Pan et al., 2003; Morton et al., 2007; Pan and Morton, 2008). Liberatore et al. (2011) considered the problem in which a defender is protecting a network from an attacker with an unknown amount of attack resources. Others have studied interdiction problems in which the network topology is unknown or uncertain (Held et al., 2005; Held and Woodruff, 2005; Hemmecke et al., 2003).

A sub-area of the literature on stochastic protection and interdiction model imperfect protection and interdiction. That is, if a network element is protected (interdicted) it is not guaranteed to survive (fail). Cormican et al. (1998), and later Janjarassuk and Linderoth (2008) and Ramirez-Marquez and Rocco (2009), studied a network interdiction problem in which interdiction is successful according to a known probability. Cormican et al. (1998), and later Janjarassuk and Linderoth (2008) and Carrigy et al. (2010), also studied a problem variant in which network elements can be in multiple capacity states after interdiction. Church and Scaparra (2006) investigate a similar problem in facility interdiction. Losada et al. (2010) analyzed a facility interdiction problem where multiple resource units can be allocated to a facility. The more units allocated, the lower the probability of facility failure. Shen et al. (2010) considered a problem in which a decision-maker can invest resources to change the probability that tasks within a project management network will be completed on time. Ramirez-Marquez et al. (2009) modeled a network protection problem in which the probability that a link fails is a function of the amount of resources allocated to that link.

A summary of the existing literature on protection and interdiction with imperfect protection and multiple capacity states is shown in Table 6.1. The rows of the table classify the literature into the type of allocation decision. Under binary allocation, an element is either protected (interdicted) or not. The columns of the table classify the literature according to the way that capacity is modeled. In binary infinite capacity, an element is either completely failed or is fully available with infinite capacity. In binary finite capacity, an element is either completely failed or is fully available with finite capacity. In multi-state capacity, an element may be in multiple capacity states. As the table shows, researchers have studied multi-level allocation and multi-state allocation, but the present paper is the first to integrate both of these in the same model.

Imperfect protection and multiple capacity states are important areas of research for the following reasons. First, in many situations protection is indeed not perfect. Even if a large amount of resources are invested to protect an element, it often still has a nonzero failure probability. Further, unprotected elements may not fail completely if exposed to a hazard. Also, the benefit of protection

Table 6.1: Stochastic protection and interdiction: Modeling of resource allocation and post-disruption capacity

	Binary infinite capacity ($0, \infty$)	Binary finite capacity ($0, a_j$)	Multi-state capacity
Binary allocation	Church and Scaparra (2006)	Cormican et al. (1998) Janjarassuk and Linderoth (2008) Ramirez-Marquez and Rocco (2009)	Cormican et al. (1998) Janjarassuk and Linderoth (2008) Carrigy et al. (2010)
Multi-level allocation	Losada et al. (2010)	Cormican et al. (1998) Janjarassuk and Linderoth (2008) Ramirez-Marquez et al. (2009)	<i>The present paper</i>

is often nonlinear. For small allocation levels, the benefit may be convex but for large allocation levels it may be concave. Second, capacity is important to model in networks subject to disruptions because disruptions cause element failures, which increase the load on other network elements. It is often unrealistic to assume that network elements have infinite capacity and therefore can take on a limitless amount of additional load. Because these assumptions are relevant to practice it is important to study the impact of these assumptions on solution quality.

Although imperfect protection and multiple capacity states are important areas of study, these two extensions present several difficulties. First, in a protection model the likelihood of element failure is a function of the first-stage protection decisions. Under perfect protection, the probability function is simple: the probability of failure is 0 if the element is protected and q (the nominal probability of failure) otherwise. Under imperfect protection this probability function needs to be more detailed, often making the model more complicated. Second, including multiple capacity states is likely to require two stages, i.e., a bi-level optimization problem or a two-stage stochastic program. Most models for facilities subject to disruptions are uncapacitated and therefore only have one stage (Snyder and Daskin, 2005; Berman et al., 2007; Scaparra and Church, 2008b; Cui et al., 2011). On the other hand, most models for flow networks subject to disruptions are

capacitated and include two stages (Wood, 1993; Cormican et al., 1998; Smith et al., 2007; Smith, 2011; Rocco et al., 2010). Two-stage problems present a difficulty because they introduce a large number of scenarios into the problem.

In this paper we attempt to overcome the challenges with modeling both imperfect protection and multiple capacity states. We model the probability function using a discrete approximation to simplify the model and improve tractability. We model capacity states using a two-stage stochastic program. However, in our stochastic program the first-stage variables affect only the scenario probabilities and not the second-stage objective function. This property significantly improves the tractability of the model.

Thus, the contributions of this paper are

1. the first study of a problem that considers both imperfect protection and capacity;
2. a reformulation that removes the non-linearity from the initial formulation and allows the second-stage problems to be solved a priori;
3. an L-Shaped implementation for the reformulation;
4. insights regarding which features a model should have; and
5. preliminary evidence that a Local Search algorithm performs well for a special case of the problem.

The remainder of this paper is as follows. Section 6.2 describes the problem studied in this paper. Section 6.3 describes a natural two-stage stochastic programming formulation, which is likely intractable. A reformulated two-stage stochastic program is provided to improve tractability. Section 6.4.1 describes our implementation of the L-Shaped method to solve the reformulated stochastic program. Experimentation showed that solutions to the problem exhibited “nested” behavior. To take advantage of this, Section 6.4.2 describes a Local Search algorithm. In Section 6.5, experimental results are reported, which describe interesting observations about solution quality and computation time.

6.2 Notation and Problem Description

In this section, the Multi-State Median Fortification Problem (MS-MFP) is described. The purpose of the MS-MFP is to

prescribe an optimal allocation of protection resources to facilities subject to a budget constraint. The system of facilities is subject to random disruptive events. After a disruptive event, each facility is in a capacity state, where the capacity state probability distribution is a function of the amount of protection resources allocated to that facility.

The following notation is used throughout the remainder of the paper:

6.2.1 Preliminary Notation

Convention

- *parameters* are represented by the beginning letters of the latin alphabet ($a-e$)
- *functions* are represented by the latin letters $f-h$
- *indices* are represented by the latin letters $i-t$; corresponding index sets are the capital latin letters $I-T$
- *decision variables* are represented by the last letters of the alphabet ($u-z$)
- *dual variables* are represented by the beginning letters of the Greek alphabet ($\alpha-\theta$)
- *random variables* are represented by Greek letters
- *vectors* of parameters and variables are represented by the bold letter without the subscript (e.g., $y = (y_{jk})_{j \in J, k \in K}$)
- *probabilities and expectations* are represented using $\mathbb{P}[\cdot]$ and $\mathbb{E}[\cdot]$, respectively

Sets and Indices

- $I := \{1, \dots, i'\}$, set of demand points indexed by i
- $J := \{1, \dots, j'\}$, set of facilities indexed by j
- $K := \{0, 1, \dots, k' - 1\}$, set of protection allocation levels indexed by k
- $L := \{0, 1, \dots, \ell' - 1\}$, set of facility capacity levels indexed by ℓ
- $j' + 1$, the index for the dummy facility

Parameters

- $a_{j\ell} :=$ amount of capacity for level ℓ at facility j ($a_j = a_{j\ell'}$ is the full capacity for facility j)
- d' := penalty multiplier for service by the dummy facility
- $d_{ij} :=$ distance from customer i to facility j ; $d_{i,j'+1} = d' \max_{j \in J} \{d_{ij}\}$
- $e_i :=$ demand of customer i

Random Variables

- $\tilde{\xi}_{j\ell} :=$ 1 if facility j is in capacity state ℓ after the disruptive event and 0 otherwise

Functions

- $h :=$ the transportation cost objective for the second-stage transportation problem
- $h(\xi) :=$ the transportation cost given the capacity state vector ξ

Decision Variables

- $x_{ij} :=$ proportion of customer i 's demand satisfied by facility j

6.2.2 Problem Description

To understand the MS-MFP, it is helpful to first understand the underlying problem without facility failures. Without facility failures, the MS-MFP reduces to the classic transportation problem, which is to provide a minimum-cost allocation of demand points to capacitated facilities. The per-unit cost for facility j to service customer i is d_{ij} . The main component of d_{ij} is travel cost between i and j . To account for the case in which total facility capacity is insufficient to satisfy total customer demand (relevant in the presence of facility failures), a dummy facility, indexed $j' + 1$, is introduced for which $d_{ij'+1} = d' \max\{d_{ij}\}$, where d' is the penalty multiplier. This dummy facility may represent rush ordering or another mode of transportation such as air. A linear programming formulation of the transportation problem is

$$h = \min. \sum_{i \in I} \sum_{j=1}^{j'+1} e_i d_{ij} x_{ij} \quad (6.1a)$$

$$\text{s.t.} \quad \sum_{i \in I} e_i x_{ij} \leq a_j \quad \forall j \in J, \quad (6.1b)$$

$$\sum_{j=1}^{j'+1} x_{ij} = 1 \quad \forall i \in I, \quad (6.1c)$$

$$x_{ij} \geq 0 \quad \forall i \in I, j \in J. \quad (6.1d)$$

The objective of model (6.1) is to minimize the total transportation cost (6.1a). The constraints ensure that: total customer demand allocated to a facility does not exceed that facility's capacity (6.1b); all of the demand is satisfied for each demand point (6.1c); and the assignment variables are non-negative (6.1d).

In extending the transportation problem to include failures, it is helpful to think of it as a two-stage problem, with a disruption event occurring between the stages. The first stage is when decisions are made to protect the facility system from random disruptions. After the first stage, a random disruptive event occurs, which has the potential to damage facilities. In the second stage, the transportation problem is solved for the post-disruption system, in which some facilities may

have degraded capacity due to damage caused by the disruptive event.

In the first stage, the planner has the option of allocating resources to facilities in order to reduce their vulnerability to hazards. Specifically, allocating resources to a facility changes the probability distribution of that facility's post-hazard capacity. These resources may represent building flood walls, retrofitting a facility to reduce its vulnerability to earthquakes, adding a backup power generator, etc. A protection plan, represented by the vector y , prescribes a discrete amount of resources to each facility. The total cost of the resources may not exceed a budget. The objective of the planner in the first stage is to minimize the expected post-hazard transportation cost.

The first stage is followed by a random disruptive event. After a disruptive event, each facility is in exactly one capacity state. The facility capacity states are represented by the random vector $\tilde{\xi} = (\tilde{\xi}_{j\ell})_{j \in J, \ell \in L}$. The post-hazard capacity probabilistically depends on the amount of protection resources that were allocated to the facility in the first stage. That is, $\mathbb{P}[\tilde{\xi}_{j\ell} = \xi_{j\ell}]$ is a function of the amount of protection resources allocated to j .

After the disruptive event, the second stage occurs. In the second stage, the post-degradation transportation problem is solved given a realization, ξ , of the random capacity state vector $\tilde{\xi}$. Given the realization ξ , the amount of capacity for facility j is $\sum_{\ell \in L} a_{j\ell} \xi_{j\ell}$, where $a_{j\ell}$ is the amount of capacity available at facility j when j is in state ℓ . Thus, the post-degradation transportation problem is represented as

$$h(\xi) = \min. \sum_{i \in I} \sum_{j=1}^{j'+1} e_i d_{ij} x_{ij} \quad (6.2a)$$

$$\text{s.t. } \sum_{i \in I} e_i x_{ij} \leq \sum_{\ell \in L} a_{j\ell} \xi_{j\ell} \quad \forall j \in J, \quad (6.2b)$$

(6.1c)–(6.1d).

The only difference between h and $h(\xi)$ is that the capacity constraint 6.2b depends on the the capacity state vector.

In the MS-MFP the post-hazard facility capacity state vector is conditional upon both the first-

stage protection decisions, y , and the realization of the disruption, represented by the vector $\tilde{\xi}$. Therefore, the problem is to find a first-stage decision vector y that solves

$$\min_y \mathbb{E}_{\tilde{\xi}}[h(\tilde{\xi})]. \quad (6.3)$$

6.3 Model

In this section a more explicit formulation of the MS-MFP is presented. The following additional notation is used throughout the remainder of the paper:

6.3.1 Additional Notation

Sets and Indices

- $S :=$ the set of random scenarios, indexed by s

Parameters

- $b :=$ budget for protecting facilities
- $c_{jk} :=$ cost of allocating k resources to facility j
- $\mathbb{P}_{j\ell k} :=$ probability that element j is in capacity state ℓ after the disruption given that k resources were allocated to j
- $\hat{h}^s :=$ the transportation cost for scenario s

Variables

- $y_{jk} :=$ 1 if k resources are allocated to facility j and 0 otherwise
- $x_{ij}^s :=$ proportion of customer i 's demand satisfied by facility j in scenario s

6.3.2 Model

Representing y as a vector of binary variables, a natural formulation of the MS-MFP is as a two-stage stochastic program:

$$\min. \mathbb{E}_{\tilde{\xi}}[h(\tilde{\xi})] \quad (6.4a)$$

$$\text{s.t. } \sum_{k \in K} y_{jk} = 1 \quad \forall j \in J, \quad (6.4b)$$

$$\sum_{j \in J} \sum_{k \in K} c_{jk} y_{jk} \leq b, \quad (6.4c)$$

$$y_{jk} \in \{0, 1\} \quad \forall j \in J, k \in K. \quad (6.4d)$$

The objective (6.4a) minimizes the expected second-stage transportation cost. The vector y is a binary vector (6.4d) and its total cost must be within a budget (6.4c). Constraints (6.4b) ensure that each facility is at exactly one allocation level.

A major difficulty with solving (6.4) is computing the expectation in the objective function. Because $\tilde{\xi}$ is a discrete random variable with countable range, its outcome space can be enumerated as $\Xi = \{\xi^s\}_{s \in S}$. In stochastic programming nomenclature, S is called the set of scenarios. This allows the objective (6.4a) to be expressed as a weighted sum:

$$\sum_{s \in S} \mathbb{P}^s(y) h(\xi^s), \quad (6.5)$$

where $\mathbb{P}^s(y)$ is the likelihood of ξ^s given allocation vector y . Assuming that facilities fail independently,

$$\mathbb{P}^s(y) = \prod_{j \in J} \sum_{k \in K} \sum_{\ell \in L} \mathbb{P}_{j\ell k} \xi_{j\ell}^s y_{jk}. \quad (6.6)$$

Hence, (6.4) can be expressed in extensive form as

$$\min. \sum_{s \in S} \prod_{j \in J} \sum_{k \in K} \sum_{\ell \in L} \mathbb{P}_{j\ell k} \xi_{j\ell}^s y_{jk} \sum_{i \in I} \sum_{j=1}^{j'+1} e_i d_{ij} x_{ij}^s \quad (6.7a)$$

$$\text{s.t.} \quad \sum_{i \in I} e_i x_{ij}^s \leq \sum_{\ell \in L} a_{j\ell} \xi_{j\ell}^s \quad \forall j \in J, s \in S, \quad (6.7b)$$

$$\sum_{j=1}^{j'+1} x_{ij}^s = 1 \quad \forall i \in I, s \in S, \quad (6.7c)$$

$$x_{ij}^s \geq 0 \quad \forall i \in I, j \in J, s \in S, \quad (6.7d)$$

(6.4b)–(6.4d).

Model (6.7) is undesirable because of the nonlinear objective function and is presented only for illustration. The underlying reason for this non-linearity is to compute the expectation in (6.5). To eliminate the non-linearity, we rely on the following fact based on the structure of the MS-MFP:

Fact 1. *The optimal solutions to the subproblems $h(\xi^s)$ are independent of the first-stage vector, y .*

In most stochastic programs, the subproblems $h(\cdot)$ are a function of both the first-stage vector and the random variable. Thus, each time a new first-stage vector is identified, each of the subproblems must be solved again. However, because of Fact 1, we do not have to solve the subproblems $h(\xi^s)$ each time a new y is identified— we can solve them prior to constructing the model and use their fixed values, \hat{h}^s , in the construction of the model.

Thus, formulation (6.7) can be reformulated as

$$\min. \sum_{s \in S} \hat{h}^s \prod_{j \in J} \sum_{k \in K} \sum_{\ell \in L} \mathbb{P}_{j\ell k} \xi_{j\ell}^s y_{jk} \quad (6.8a)$$

$$\text{s.t.} \quad (6.4b)–(6.4d).$$

To remove the remaining non linearity in Model (6.8), we rely on the assumption that facilities fail independently. Thus, we can replace the product $\prod_{j \in J} \sum_{k \in K} \sum_{\ell \in L} \mathbb{P}_{j\ell k} \xi_{j\ell}^s y_{jk}$ with a recursive expression by using bookkeeping variables to calculate the product of probabilities. Let w_{rks} be a

bookkeeping variable that holds value of \hat{h}^s multiplied by the likelihood that facilities $1, \dots, r-1$ are in their corresponding capacity states defined by scenario s if k units have been allocated to facility r and zero otherwise. Thus, we want w_{rks} to hold the value

$$\hat{h}^s \prod_{j=1}^{r-1} \sum_{k \in K} \mathbb{P}_{r,\ell,k} \sum_{\ell \in L} \xi_{j\ell}^s y_{jk}.$$

To ensure that the values of w_{rks} are computed correctly, we use the following recursive equations

$$\hat{h}^s = \sum_{k \in K} w_{1ks} \quad \forall s \in S, \quad (6.9)$$

$$\sum_{k \in K} \mathbb{P}_{r-1,\ell,k} \sum_{\ell \in L} \xi_{j\ell}^s w_{r-1,ks} = w_{rks} \quad \forall r = 2, \dots, j'; s \in S, \quad (6.10)$$

along with the constraints

$$w_{rks} \leq y_{rk} \quad \forall r = 1, \dots, j'; k \in K; s \in S, \quad (6.11)$$

which ensure that w_{rks} is positive only if k units are allocated to facility j . Constraints (6.11) help ensure that the w_{rks} variables hold the correct values.

The MS-MFP can now be reformulated as an integer linear program:

$$\min. \sum_{s \in S} \sum_{k \in K} \sum_{\ell \in L} \mathbb{P}_{j'\ell k} \xi_{j\ell}^s w_{j'ks} \quad (6.12a)$$

$$\text{s.t. } \hat{h}^s = \sum_{k \in K} w_{0ks} \quad \forall s \in S,$$

$$\sum_{k \in K} \mathbb{P}_{r-1,\ell,k} \sum_{\ell \in L} \xi_{j\ell}^s w_{r-1,ks} = w_{rks} \quad \forall r = 1, \dots, j'; s \in S,$$

$$w_{rks} \leq \hat{h}^s y_{rk} \quad \forall r = 1, \dots, j'; k \in K; s \in S, \quad (6.12b)$$

$$w_{rks} \geq 0 \quad \forall r = 1, \dots, j'; k \in K; s \in S, \quad (6.12c)$$

(6.4b)–(6.4d).

The objective (6.12a) is to minimize the expected second-stage transportation cost and is equivalent to (6.5). Note that because the values of \hat{h}^s are computed a-priori, the second stage transportation does not need to be linear or even convex.

6.4 Solution Method

Although the non-linearity has been removed, formulation (6.12) still has several computational difficulties. Mainly, the number of variables and constraints depends on the number of scenarios, which can be quite large. In this section we describe solution strategies that exploit the structure inherent in the MS-MFP.

6.4.1 L-Shaped Method

Formulation (6.12) has the L-shaped structure of the constraint matrix that is common in two-stage stochastic programs. In particular, for a fixed y , the constraints and variables are separable by the scenarios $s \in S$. Thus, we can use the multi-cut L-shaped method for stochastic programs (Birge and Louveaux, 1988) to solve (6.12).

To take advantage of the L-shaped structure, we can reformulate (6.12) as

$$\begin{aligned} \min. \quad & \sum_{s \in S} g(y, \xi^s) & (6.13a) \\ \text{s.t.} \quad & (6.4b) - (6.4d), \end{aligned}$$

where $g(y, \xi^s)$ equals

$$\min. \sum_{k \in K} \sum_{\ell \in L} \mathbb{P}_{j'\ell k} \xi_{j\ell}^s w_{j'ks} \quad (6.14a)$$

$$\text{s.t. } \hat{h}^s = \sum_{k \in K} w_{0ks}, \quad [\alpha_s] \quad (6.14b)$$

$$\sum_{k \in K} \mathbb{P}_{r-1,\ell,k} \sum_{\ell \in L} \xi_{j\ell}^s w_{r-1,ks} = w_{rks} \quad \forall r = 1, \dots, j' \quad (6.14c)$$

$$w_{rks} \leq \hat{h}^s y_{rk} \quad \forall r = 1, \dots, j'; k \in K, \quad [\beta_{rks}] \quad (6.14d)$$

$$w_{rks} \geq 0 \quad \forall r = 1, \dots, j'; k \in K. \quad (6.14e)$$

Dual variables for constraints with nonzero right-hand sides are listed in brackets. Notice that because there are $\xi_{j\ell}^s$ terms in the objective function and the left-hand sides of the constraints, this stochastic program does not have fixed recourse (Birge and Louveaux, 1997). Also notice that given a fixed solution vector, \hat{y} , the problem $g(\hat{y}, \xi^s)$ can be solved by inspection. To see this, note in an optimal solution there exists exactly one $k \in K$ for which $w_{rks} \geq 0$ for all $r = 1, \dots, j'$ and $s \in S$ because $y_{rk} \in \{0, 1\}$ for all $r = 1, \dots, j'$ and $k \in K$ and $\sum_{k \in K} y_{rk} = 1$ for all $r = 1, \dots, j'$. Further, Constraints (6.14d) determine for all $r = 1, \dots, j'$ and $s \in S$ the $k \in K$ for which $w_{rks} \geq 0$. Knowing which w -variables are positive, it is easy to determine their values.

The idea behind the L-shaped method is that $g(y, \xi^s)$ is a convex function of y and thus we can form successive approximations of $g(y, \xi^s)$ using supporting hyperplanes, i.e., cutting planes. Let O^s be the set of all dual vertices for scenario s . Let α_s^o and β_{rks}^o be the dual variable values corresponding to vertex o . If we enumerate all possible dual vertices, we can reformulate (6.12) as the L-Shaped master problem:

$$\text{(MP) } \min. \sum_{s \in S} z^s \quad (6.15a)$$

$$\text{s.t. } z^s \geq \hat{h}^s \alpha_s^o + \sum_{r=1}^{j'-1} \sum_{k \in K} \beta_{rks}^o y_{rk} \quad s \in S, o \in O^s, \quad (6.15b)$$

(6.4b)–(6.4d).

Rather than enumerating all possible dual vertices, the L-shaped uses delayed row generation to successively improve the approximation of $g(y, \xi^s)$. Let $O^{s,m}$ be the set of dual vertices identified in iterations $1, \dots, m$ of the L-Shaped algorithm. Thus, the restricted L-shaped master problem at iteration m is

$$\text{RMP}_m \quad \min. \quad \sum_{s \in S} z^s \quad (6.16a)$$

$$\text{s.t.} \quad z^s \geq \hat{h}^s \alpha_s^o + \sum_{r=1}^{j^s-1} \sum_{k \in K} \beta_{rks}^o y_{rk} \quad o \in O^{s,m}, \quad (6.16b)$$

(6.4b)–(6.4d).

The L-Shaped algorithm is described in Algorithm 6.1.

Algorithm 6.1 L-Shaped algorithm

Step 1: Compute second-stage objective values

Compute the value of \hat{h}^s for all $s \in S$

Step 2: find optimal allocation vector

Choose an initial feasible allocation vector y_0 . Set $m = 0$, $UB = \infty$. Choose ε to be the relative optimality gap tolerance.
repeat

Evaluate $\sum_{s \in S} g(y_m, \xi^s)$. Collect α and β for all $s \in S$ and add to $O^{s,m}$.

$UB = \min\{UB, \sum_{s \in S} g(y_m, \xi^s)\}$.

Solve RMP_m to obtain solution y_{m+1} with optimal value $LB = \sum_{s \in S} z^s$.

$m = m + 1$

until $(LB - UB)/UB \leq \varepsilon$

6.4.2 Local Search Heuristic

In our experimentation, we found that the L-Shaped algorithm took a long time to solve. We also noticed that the solutions to the MF-MFP were often nested in the sense that as the budget

increased the allocation vector y only changed slightly. In particular, we found that if the budget is integer and the costs c_{jk} equal one, then for exactly one $\bar{j} \in J$, if $y_{\bar{j}\bar{k}}^*(b) = 1$ for some $\bar{k} \in K$, then

$$y_{\bar{j},\bar{k}+1}^*(b+1) = 1,$$

$$y_{\bar{j}k}^*(b+1) = 0 \quad \forall k \neq \bar{k} \in K,$$

and

$$y_{jk}^*(b) = y_{jk}^*(b+1) \quad \forall j \neq \bar{j} \in J, k \in K,$$

where $y_{jk}^*(b)$ is the optimal value of y_{jk} given that the budget equals b .

This property suggests the following Local Search heuristic. The algorithm starts with an empty allocation vector y^0 . Let $y^m(\bar{j})$ be a copy of y^m with one unit added to facility \bar{j} . That is,

$$y_{\bar{j},\bar{k}+1}^m(\bar{j}) = 1,$$

$$y_{\bar{j}k}^m(\bar{j}) = 0 \quad \forall k \neq \bar{k} \in K,$$

and

$$y_{jk}^m(\bar{j}) = y_{jk}^m \quad \forall j \neq \bar{j} \in J, k \in K.$$

At each iteration m , the Local Search algorithm determines on which facility to place the additional unit. The Local Search heuristic is described in Algorithm 6.2. Note that the Local Search algorithm allows facility failures to be correlated. Also note that the second-stage transportation problem is not required to be linear or even convex.

Algorithm 6.2 Local Search heuristic

Step 1: Compute second-stage objective values

 Compute the value of \hat{h}^s for all $s \in S$

Step 2: find optimal allocation vector

 Choose an initial feasible allocation vector y^0 such that $y_{jk} = 0$ for all $j \in J$ and $k \in K$. Set $m = 0$.

 for $m = 1$ to b

$$y^{m+1} = \operatorname{argmin}_{j \in J} \{ \sum_{s \in S} \hat{h}^s \mathbb{P}^s(y^m(j)) \}$$

$$m = m + 1$$

 return y_b as the optimal allocation vector

6.5 Experimentation

In this section we experiment with the MS-MFP model to provide decision-making and computational insights. Our experiments are defined as follows:

Datasets We use two common facility location datasets from Daskin (1995). The first dataset, which we call d49, contains demand points located at the capital cities in the lower-48 states in the United States as well as at Washington, D.C. The second dataset, which we call d88, contains demand points at the 88 largest cities in the US. To obtain a set of located facilities, we solve the p -median problem Hakimi (1964). For example, the d49-5 instance consists of 49 demand points with five facilities obtained by solving the p -median problem. Because d49 and d88 datasets do not contain capacity information, we set the capacity of each facility to

$$a_j = \frac{\sum_{i \in I} e_i}{(1 - 0.1)},$$

implying that the system is designed with 10% excess capacity.

Model Parameters We vary the following model parameters: number of allocation levels, number of capacity levels, and the budget. We set the penalty multiplier to $d' = 2.0$. The capacity for each level is set to $a_{j\ell} = \frac{\ell}{\bar{\ell}} a_j$.

The probabilities of capacity states given allocation amounts is generated using the binomial

distribution so that

$$\mathbb{P}_{j\ell k} = \text{binom}(\ell, \ell' - 1, p_{jk}),$$

where $\text{binom}(x, n, p)$ is the binomial probability mass function with x successes, n trials, and probability of success p . The parameter p_{jk} is computed using a contest success function (Skaperdas, 1996), a function often used in the economics literature. Thus,

$$p_{jk} = \frac{(k + 0.25)^{1.5}}{(k + 0.25)^{1.5} + 1.0^{1.5}}, \quad (6.17)$$

which makes up a sigmoid or ‘‘S-curve’’. By the shape of the sigmoid function, $p_{jk}(k)$ is convex when k is small and concave when k is large, a desirable property for resource allocation.

6.5.1 Comparison with Other Models

In this section the MS-MFP is compared with results from other related problems. Because each of these problems are special cases of the MS-MFP, we use the MS-MFP to generate solutions to these problems by choosing appropriate parameters. The problems are:

1. **The P-Median Fortification Problem (PMFP)** (Snyder et al., 2006). This model minimizes the expected travel distance from demand points to their closest operating facilities by selecting p facilities to protect. If a facility is protected, it is guaranteed to operate. Otherwise, it fails with known probability q . Facilities are uncapacitated. To convert the MS-MFP to the PMFP, set $k' = 2$ and $\ell' = 2$, set the capacity of each facility to infinity, and let $\mathbb{P}_{j,1,0} = 1 - q$, $\mathbb{P}_{j,1,1} = 1$, $\mathbb{P}_{j,0,1} = 0$, and $\mathbb{P}_{j,0,0} = q$. In our experimentation we used $q = 0.5$.
2. **The Capacitated PMFP (C-PMFP)**. A version of the PMFP where facilities are capacitated. We used the same capacity as in the MS-MFP.
3. **The Multiple-Resource PMFP (MR-PMFP)**¹. Like the PMFP, the MR-PMFP only has two capacity states: operating and failed. Like the MS-MFP, multiple units can be allocated to a

¹This model has not appeared in the literature but it was referenced in Losada et al. (2010)

facility. Further, the probability that a facility is in a capacity state depends on the number of units allocated to that facility. To convert the MS-MFP to the MR-PMFP, set $\ell' = 2$ and set the capacity of each facility to infinity.

A summary of the three models and the MS-MFP is shown in Table 6.2.

Table 6.2: Models included in experimentation

Model Name	Allocation levels	Capacity states	Facility capacity
PMFP	2	2	Infinite
C-PMFP	2	2	Finite
MR-PMFP	k'	2	Infinite
MS-MFP	k'	ℓ'	Finite

The goal of this section is to compare the different models regarding the allocation solutions they generate and the corresponding objective function values. Toward this goal, we examined an instance of the d49-7 dataset, which has seven facilities: Sacramento, CA (SAC); Albany, NY (ALB); Austin, TX (AUS); Tallahassee, FL (TAL); Springfield, IL (SPR); Olympia, WA (OLY); and Columbus, OH (COL). This d49-7 instance has five allocation levels and five capacity states and the allocation budget is eight units. The optimal allocation plan distributes the eight units as 1-2-1-1-1-1-1, producing an objective value of 4.51×10^6 . For the purposes of this section, we refer to this instance as the “true problem” and 4.51×10^6 as the “true objective value”.

Table 6.3 shows the solutions and objective values that are obtained when different numbers of allocation levels and capacity states are used. Each row represents the solution generated by a model instance. For the MR-PMFP the number of allocation levels, k' , is varied. For the MS-MFP, both k' and the number of capacity states, ℓ' , are varied.

Because k' and ℓ' vary among the model instances, the solutions for a model instance must be translated back to the true problem. After a model instance is solved the resulting solution is translated into a solution to the true problem via the transformation

$$y_j^{TRUE} = y_j \frac{(5 - 1)}{(k' - 1)},$$

where y_j^{TRUE} is the amount allocated to facility j in the translated solution and y_j is the amount allocated to j in the instance (k', ℓ') . This transformation implies that one unit of allocation in instance (k', ℓ') is equivalent to $(5 - 1)/(k' - 1)$ units in the true problem. (Note that $(5 - 1)$ is the number maximum number of units one facility can receive in the true problem and $(k' - 1)$ is the maximum number for the instance (k', ℓ') .) For example, in the first row, the solution to the PMFP instance 1-0-0-0-0-1-0 and the translated solution is 4-0-0-0-0-4-0. The translated solution is then evaluated in the true problem to obtain the *objective value to the true problem*. The *relative objective function increase* is then:

$$\frac{(\text{Objective value to the true problem}) - (\text{true objective value})}{\text{true objective value}}$$

Table 6.3: Solutions for different models

Model Name	Resources allocated to facility							Obj. val. to true problem	Rel. obj. fn. increase
	SAC	ALB	AUS	TAL	SPR	OLY	COL		
PMFP	4	0	0	0	0	4	0	8.33E6	0.85
C-PMFP	0	4	0	0	0	4	0	8.35E6	0.85
MR-PMFP ($k' = 3$)	2	2	0	0	2	2	0	6.15E6	0.36
MR-PMFP ($k' = 5$)	2	2	1	1	1	1	0	4.99E6	0.11
MS-MFP ($k' = 5$) ($\ell' = 5$)	1	2	1	1	1	1	1	4.51E6	0.00

Table 6.3 shows that the model used is important with regard to the solution produced and the corresponding objective function value. The PMFP and C-PMFP models have a high relative objective function increase, meaning that they are likely a poor surrogate for the MF-MFP. This high

relative objective function increase is probably due to the fact that they consider perfect protection while the MS-MFP allows for imperfect protection. Also, in this instance the PMFP and C-PMFP can only allocate “all-or-nothing” to a facility. Thus, the decision of the PMFP and C-PMFP for this problem instance is which two facilities to protect perfectly. Therefore, in the optimal solution two facilities have four units and the remaining facilities have zero units. The MR-MFP model has a lower relative objective function increase than the PMFP and C-PMFP. This is probably due to the fact that the MR-MFP model includes multiple allocation levels.

6.5.2 Sensitivity Analysis

In this section we examine how sensitive solutions are to three parameters in the MS-MFP: the budget, the number of allocation levels, and the number of capacity states.

6.5.2.1 Budget

In this section we investigate how budget increases change the optimal allocation plan. The d49-7 instance is used with five allocation levels and five capacity states. Table 6.4 shows solutions for budget amounts ranging from one to 28 (the maximum allocation). Each row shows how many units are allocated to each facility and the resulting objective value.

Table 6.4: Solutions for different budget amounts

Budget	Resources allocated to facility							Obj. value
	SAC	ALB	AUS	TAL	SPR	OLY	COL	
1	1	0	0	0	0	0	0	10.57E6
2	1	1	0	0	0	0	0	9.59E6
3	1	1	0	0	0	1	0	8.63E6
4	1	1	0	0	1	1	0	7.67E6
5	1	1	0	1	1	1	0	6.71E6
6	1	1	1	1	1	1	0	5.76E6
7	1	1	1	1	1	1	1	4.90E6
8	1	2	1	1	1	1	1	4.51E6
9	1	2	1	1	1	2	1	4.14E6
10	1	2	1	1	2	2	1	3.76E6
11	1	2	1	2	2	2	1	3.39E6
12	1	2	2	2	2	2	1	3.03E6
13	2	2	2	2	2	2	1	2.69E6
14	2	2	2	2	2	2	2	2.40E6
15	2	3	2	2	2	2	2	2.23E6
16	2	3	2	2	2	3	2	2.08E6
17	2	3	2	2	3	3	2	1.92E6
18	2	3	2	3	3	3	2	1.77E6
19	2	3	3	3	3	3	2	1.64E6
20	3	3	3	3	3	3	2	1.51E6
21	3	3	3	3	3	3	3	1.41E6
22	3	4	3	3	3	3	3	1.34E6
23	3	4	3	3	3	4	3	1.27E6
24	3	4	3	3	4	4	3	1.20E6
25	3	4	3	4	4	4	3	1.13E6
26	3	4	4	4	4	4	3	1.08E6
27	4	4	4	4	4	4	3	1.03E6
28	4	4	4	4	4	4	4	9.99E5
SUM	55	69	53	57	61	65	46	

The key characteristic of Table 6.4 is that the solutions are all nested. That is, $(y_j^b)^* \leq (y_j^{b+1})^*$ for all $j \in J$, where $(y_j^b)^*$ is the optimal number of units allocated to facility j when the budget is b . Unfortunately, this property does not hold in general. However, in 203 of 204 instances tested, the property held (see Section 6.5.3.2).

By summing the columns of Table 6.4 a metric can be obtained that indicates the criticality of each facility. The higher the metric, the more protection a facility receives. According to this metric, Table 6.4 indicates that the facility at Albany is the most critical.

6.5.2.2 Number of Allocation Levels and Capacity States

In this section we examine the sensitivity of the solution to the number of allocation levels and capacity states. We performed experiments on an instance of the d49-5 dataset that has five facilities: Sacramento, CA (SAC); Austin, TX (AUS); Tallahassee, FL (TAL); Springfield, IL (SPR); and Trenton, NJ (TRE). The instance has seven allocation levels and seven capacity states and the allocation budget is twelve units. The optimal allocation plan among the five facilities is 2-2-2-3-3, producing an objective value of 1.96×10^6 . As in Section 6.5.1, we refer to this instance as the “true problem” and 1.96×10^6 as the “true objective value”.

Table 6.5 shows the solutions and objective values that are obtained when different numbers of allocation levels and capacity states are used. Each row represents an instance defined by the number of allocation levels, k' , and the number of capacity states, ℓ' . Each instance is solved and the resulting solution is translated into a solution to the true problem via the transformation

$$y_j^{TRUE} = y_j \frac{(7-1)}{(k'-1)},$$

similar to the transformation Section 6.5.1. The translated solution is then evaluated by the true problem to obtain the *objective value to the true problem*. The *relative objective function increase* is defined in Section 6.5.1.

Table 6.5: Solutions for different numbers of allocation levels and capacity states

Resources allocated to facility							Obj. val. to true problem	Rel. obj. fn. increase
k'	ℓ'	SAC	AUS	TAL	SPR	TRE		
2	2	0	0	0	6	6	6.86E6	2.57
2	3	0	0	0	6	6	6.86E6	2.57
2	4	0	0	0	6	6	6.86E6	2.57
2	5	0	0	0	6	6	6.86E6	2.57
2	6	0	0	0	6	6	6.86E6	2.57
2	7	0	0	0	6	6	6.86E6	2.57
4	2	2	2	2	2	4	2.02E6	0.05
4	3	2	2	2	2	4	2.02E6	0.05
4	4	2	2	2	2	4	2.02E6	0.05
4	5	2	2	2	2	4	2.02E6	0.05
4	6	2	2	2	2	4	2.02E6	0.05
4	7	2	2	2	2	4	2.02E6	0.05
7	2	2	2	2	3	3	1.92E6	0.00
7	3	2	2	2	3	3	1.92E6	0.00
7	4	2	2	2	3	3	1.92E6	0.00
7	5	2	2	2	3	3	1.92E6	0.00
7	6	2	2	2	3	3	1.92E6	0.00
7	7	2	2	2	3	3	1.92E6	0.00

Table 6.5 shows that the relative objective function increase for the instances with $k' = 2$ is very high: 2.57. This indicates that the MS-MFP is very sensitive to changes in the number of allocation levels.

On the other hand, for a fixed k' , the relative objective function increase is the same for different numbers of capacity states. This indicates that the MS-MFP is insensitive to changes in the number of capacity states. Recall it is important to include facility capacities (Section 6.5.1). However, the results in this section indicate that the number of capacity states included in the model is much less important. This is a useful observation from a modeling standpoint because it suggests that a model can include less capacity states without sacrificing model accuracy. By including less capacity states, the modeler does not have to estimate as many facility state probabilities, which are likely to

be difficult to estimate in practice. This is also a useful observation from computational standpoint because the number of scenarios is $\ell^{j'}$; thus, including less capacity states can significantly reduce the size of the model.

6.5.3 Computational Performance

In this section several questions are addressed that relate to the computational performance of the L-shaped algorithm and Local Search algorithm:

- How good are the solutions produced by the Local Search algorithm?
- Does the Local Search algorithm require less time than the L-shaped algorithm?
- How large of instances can the Local Search algorithm solve?

6.5.3.1 Experimental Set-Up

Both the L-shaped method and the Local Search algorithm were programmed in Java. All linear programs were solved using the CPLEX linear programming solver with the network simplex algorithm. The subproblems (6.14) were solved via inspection. The master problem for the L-shaped method was solved using the CPLEX integer programming solver. Random numbers were generated using the Java Simulation Library (JSL).

The L-Shaped method was run on a 12-core compute node that is part of the Arkansas High Performance Computing Cluster (AHPCC). The node has two Xeon X5670 Intel processors, which each have six cores and a clock speed of 2.93GHz. The algorithm was allocated 5GB of virtual memory. A 64-bit Linux operating system was used for all computations. We used an optimality tolerance of 0.01 for the L-Shaped algorithm.

The Local Search algorithm was run on a single-core 2.66GHz AMD processor running 64-bit Linux with 5GB of virtual memory.

6.5.3.2 Local Search: Solution Quality

Given that the Local Search algorithm is a heuristic, an important question to ask is “How good are the solutions produced by the Local Search algorithm?” To answer this question, a large set

of problem instances were tested. These instances included the d49 and d88 datasets with seven and nine facilities, 2–5 allocation levels, and 2–5 capacity levels. There are both capacitated and uncapacitated instances and instances with perfect and imperfect protection. In all, 204 instances were tested. In addition to using the Local Search algorithm, each instance was solved using total enumeration to identify the optimal solution.

In 203 of the 204 instances, the Local Search algorithm produced the optimal solution. The sub-optimal instance was the d88-7 dataset with infinite facility capacity and perfect protection, 2 allocation levels, and 2 capacity levels. The optimality gap was 0.3%.

6.5.3.3 L-Shaped vs. Local Search

In this section the run times of the L-shaped algorithm and Local Search algorithm are compared. Table 6.6 displays the run times of the L-shaped algorithm and Local Search algorithm for selected problem instances.

Table 6.6 shows that the run time of the L-shaped algorithm increases exponentially in the number of allocation levels and capacity states. This is because the number of variables depends on the number of allocation levels and the number of scenarios depends on the number of capacity states.

Table 6.6: Comparison of L-shaped and Local Search algorithms

Dataset	Instance		Run time (s)	
	k'	ℓ'	L-Shaped	Local Search
d49-3	4	4	1	<1
d49-4	4	4	46	<1
d49-5	4	2	11	<1
d49-5	4	3	127	<1
d49-5	4	4	2130	<1
d49-5	5	4	13408	<1
d88-3	4	4	2	<1
d88-4	4	4	45	<1
d88-5	4	4	2071	<1

On the other hand, the Local Search algorithm solves all of the instances very quickly. This is because the complexity of the Local Search algorithm is on the order of the number of facilities times the number of units available to allocate. Based on the results in Table 6.6, the Local Search algorithm runs faster than the L-shaped algorithm. This is especially true given that the L-Shaped algorithm used twelve cores while the Local Search algorithm only used one. The slow runtime of the L-Shaped algorithm could be due to the fact that Model (6.14) does not have fixed recourse. Thus, the set of dual vertices for the subproblems may be different for each scenario. Under fixed recourse, the set of dual vertices for the subproblems are independent of the scenario.

6.5.3.4 Local Search: Scalability

Another question of interest is “How large of instances can the Local Search algorithm solve?” Table 6.7 shows run times for selected instances.

Each row of Table 6.7 lists the instance solved and the run time for the Local Search algorithm, broken down into the Step 1 run time and the Step 2 run time. As the table shows, the run time ranges from 149s for the smallest instance (d49-10) to 980s for the largest instance (d88-11). The sizes of the instances and the corresponding run times are comparable to other studies that examine facility disruptions (Snyder and Daskin, 2005; Scaparra and Church, 2008a).

As Table 6.7 shows, the largest instances solved included ten and eleven facilities with four capacity states. The Local Search algorithm ran out of memory when trying to solve instances with more than four capacity states. This is because ten facilities with four capacity states results in $4^{10} = 1.04 \times 10^6$ scenarios and eleven facilities with four capacity states results in $4^{11} = 4.19 \times 10^6$ scenarios. Increasing the number of capacity states beyond four results in more scenarios than the computer could store. Instances with 12 facilities also ran out of memory because of the large number of scenarios.

Further, the results in Table 6.7 show that Step 1 of the Local Search algorithm, in which a linear program is solved for each scenario, accounts for most of the run time. The run time of Step 1 can be reduced by using the Sample Average Approximation technique, in which Step 1 would be computed for a sample of scenarios, rather than the entire set of scenarios. Thus, the Sample

Average Approximation method would reduce both the run time and the memory requirements of the Local Search algorithm.

The results also show that the run time increases slowly with the number of demand points. This is because the transportation problems only need to be solved once for each scenario.

Table 6.7: Computational performance of the Local Search algorithm for larger datasets

Dataset	Instance		Run time (s)		
	k'	ℓ'	Step 1	Step 2	Total
d49-10	4	4	122	27	149
d49-10	5	4	122	23	145
d49-10	6	4	124	40	144
d49-11	4	4	510	143	653
d49-11	5	4	500	189	689
d49-11	6	4	507	235	742
d88-10	4	4	175	27	202
d88-10	5	4	172	34	206
d88-10	6	4	173	42	215
d88-11	4	4	742	143	885
d88-11	5	4	748	199	946
d88-11	6	4	742	238	980
all	>4	>4	out of memory		
d49-12	>4	>4	out of memory		
d88-12	>4	>4	out of memory		

6.6 Conclusion

In this research we studied the problem of allocating resources to a system of facilities, where allocating resources to a facility reduces its probability of failure. The objective is to minimize the expected total cost of transporting goods from facilities to demand points.

To study this problem, we formulated it as a two-stage stochastic program in which the first-stage decisions affect the scenario probabilities rather than the objective of the second-stage problem. Because the initial formulation was nonlinear, we assumed that facilities fail independently

and developed a reformulation in which the partial sums making up the expected value are computed in the second-stage problem rather than in the master problem. We then used the L-Shaped method to solve our reformulation. Because the execution of the L-Shaped method may take a long time, we also utilized a Local Search heuristic that takes advantage of the fact that the optimal solutions are often nested. The Local Search algorithm has two benefits in addition to a short run time: 1) the scenarios may be correlated, 2) the second-stage problem does not have to be linear or even convex.

This research contributes to the existing body of literature by relaxing two common assumptions. First, while many models in the literature only consider binary protection, i.e., a facility is either fully protected or fully unprotected, this research considers multiple levels of protection. Second, while many models in the literature assume that a facility can only be in one of two states—fully operating or completely failed—we study the case where disruptions can cause facilities to become partially degraded.

Our experimentation on a limited set of problem instances revealed several interesting observations. First, in Section 6.5.1 we found that assuming perfect protection was inadequate if the protection is not really perfect. We also found that it is important to consider facility capacity in facility disruption models. In Section 6.5.2.2 more light was shed on this insight: we found that the number of capacity states did not affect solution quality but that the number of allocation levels affects solution quality significantly. The number of allocation levels affects solution quality because the more allocation levels there are, the more a decision-maker is able to spread out the allocation of units. Second, in Section 6.5.2.1 we found that when the budget was integer and the costs were unitary, solutions were often nested, i.e., the solution for a budget of b varies only slightly from the solution for a budget of $b - 1$.

6.6.1 Future Work

There are several areas of future work directly relating to the current paper. First, we plan to further investigate the nested property, possibly finding conditions for which the solutions are always nested. Second, we plan to utilize the Sample Average Approximation method to avoid having to

evaluate a very large set of scenarios. Along with this, we plan to investigate the convergence of the Local Search algorithm in conjunction with Sample Average Approximation.

Because both the L-Shaped algorithm and the Local Search algorithm place few restrictions on the second-stage problem, we would like to investigate this approach on other problems. First, we plan to apply this methodology to an interdiction problem in which interdiction is imperfect. Second, we plan to test this approach on other multi-state reliability problems. Finally, we plan to test our approach on other stochastic programs in which the first-stage allocation can be viewed as changing the scenario probabilities, rather than the second-stage objective.

A natural extension of the problem studied in this paper is to consider cascading failures, i.e., when the amount of demand served by a facility reaches its capacity, the facility fails. This would be especially relevant in studying the protection and interdiction of power grids.

7 Conclusions

7.1 Summary

This dissertation studied how to locate and protect networks of facilities that are subject to disruptions. The dissertation focused on developing new models and solution strategies for these problems to generate both computational and decision-making insights.

Chapter 4 investigated the r -All-Neighbor p -Center Problem (RANPCP), which is to locate a set of facilities to serve a set of demand points given that the facilities are subject to attacks by an intelligent adversary. This chapter contributes to the literature by being the first rigorous analysis of the RANPCP. To facilitate the analysis, several MIP models were presented as well as a binary search algorithm, which outperformed the MIP models. Experimentation with the model generated the following insights.

1. optimizing either regular system performance in isolation or potential consequence in isolation produces solutions that perform poorly (at least 43% worse, on average) with respect to the other objective;
2. a significant reduction in potential consequences can be obtained by allowing a small reduction in regular system performance, and vice versa; and
3. significant reductions in potential consequences can be obtained by building a small number of additional facilities.

Chapter 5 extended the work on the RANPCP from Chapter 4 to also include facility hardening, a special case of facility protection. An important part this extension is that the facility location and facility hardening decisions are integrated. An MIP model was presented for this integrated problem as well as a binary search algorithm, which outperformed the MIP model. Experimentation with the model generated the following insights:

1. integrating facility location and hardening decisions rather than making the decisions in sequence generates much better solutions (40% smaller objective value);

2. considering both location and hardening (the MFLHP model) produces better solutions (50% smaller objective value) than when only location is considered (the RANPCP model);
3. a model for this problem should consider both performance (non-disruption radius) and vulnerability (post-disruption radius) objectives; and
4. if only one objective can be used, it should be the post-disruption radius because the average relative objective function increase for not considering facility disruptions was 5.71 times higher than the average relative objective function increase for not considering non-disruption performance.

Chapter 6 relaxed several assumptions made in most of the literature on facilities subject to disruptions and provided a preliminary investigation. The following assumptions were relaxed: 1) facility protection is perfect, 2) there is only one allocation level, and 3) facilities can only be in two capacity states. A natural two-stage integer nonlinear stochastic programming formulation was presented for this problem along with an integer linear reformulation. An implementation of the L-shaped algorithm was used to solve the model. For a special case of the problem, a Local Search algorithm was found to perform well. Experimentation with the model generated the following preliminary insights:

1. assuming perfect protection in a model will result in poor solution quality if the protection is not really perfect;
2. it is important to model capacity to achieve good solution quality, but the number of capacity levels does not affect solution quality;
3. the number of allocation levels affects solution quality; and
4. for a special case of the problem, the solutions had a special nested structure, meaning that a simple Local Search algorithm produced the optimal solution to the problem.

7.2 Major Insights Gained

In summary, the major insights generated by this dissertation are:

Modeling

1. For the problem with the maximum distance objective and the worst-case risk measure (Chapters 4 and 5), a three-stage problem can be reformulated into a single-stage problem. This idea could possibly be extended to other similar problems.
2. Chapters 5 and 6 generated results that indicate what features are important to include in modeling involving facility disruptions: (1) it is important to integrate the location and protection decisions, (2) it is important to consider imperfect protection, (3) it is important to model multiple levels of protection allocation, rather than assuming protection to be all-or-nothing, and (4) it is important to model capacity, but the number of capacity states is not critical.

Solution Methodology

1. A binary search algorithm is an efficient way of solving problems with the maximum distance objective and the worst-case risk measure (Chapters 4 and 5). This is a useful insight because there may be other important problems for which the maximum distance objective and worst-case risk measure are appropriate.
2. In Chapter 6, a local search algorithm performed well for a special case of the problem. This is a useful insight because there may be other problems for which the local search might also perform well.

Decision-Making

1. A large reduction in risk can be obtained for reasonable increase in cost.
2. A significant reduction in vulnerability can be obtained by allowing a small reduction in system performance.
3. A useful model of facilities subject to disruptions should consider both efficiency (e.g., operational cost) and vulnerability (e.g., operational cost after a disruption) because if only one

objective is optimized the value of the other objective is much worse than optimal (at least 43% higher, on average).

4. If only one objective is considered, it should be vulnerability because the average relative objective function increase for not considering vulnerability was much higher (at least 65% higher, on average) than the average relative objective function increase for not considering efficiency.
5. Assuming perfect protection in a model will result in poor solution quality (85% higher expected total distance) if the protection is not really perfect.
6. For good solution quality, it is important to consider facility capacity because experimentation showed that when capacity was ignored, the objective function value was 11-36% higher than the optimal value.

7.3 Recommendations for Future Work

This dissertation has furthered our understanding of networks subject to disruptions, but we still lack full understanding. In particular, the future work described involves developing models with more fidelity and studying new applications of networks subject to disruptions.

7.3.1 Increasing Model Fidelity

7.3.1.1 All-Hazards Protection

All of the literature on networks subject to disruptions considers only a single source of disruptions: a random source or an intelligent attacker. In reality, decision-makers in both the government and private sectors are concerned with both natural disasters and terror attacks. Thus, this new area of research would address this disconnect between theory and practice by including both random disruptions and intelligent attacks in the same model.

The major challenge of including two sources of disruptions in the same model is to find the most appropriate way to model the sequence of events. In one possible sequence, the attacker

would wait for the outcome of the random disruption and then adjust accordingly. This model is denoted as the attacker-wait-and-see model.

The attacker-wait-and-see problem could probably be modeled using a four-stage model. In the first stage, a defender prepares for disruptions by allocating three types of protection resources. The first type of resource, random protection resources, would only protect against random attacks. An example of this resource would be strengthening a levee to protect against flooding. The second type of resource, attacker protection resources, would only protect against the attacker. An example of this resource would be adding a security force to protect a bridge. The last type of resource, all-hazards protection, would protect against both random disruptions and intelligent attacks. An example of this resource would be retrofitting a bridge to better withstand shocks due to both earthquakes and bomb blasts. In the second stage of the model, the random disruption would occur. In the third stage, the attacker, observing the random disruption, would execute the attack that maximizes the damage to the network. In the fourth stage, an operator solves an optimization problem (e.g., the maximum flow problem) on the remaining network.

7.3.1.2 Multiple Time Periods

Most of the network disruptions literature also assumes that decisions are made at a single point in time. For example, all of the location and protection decisions are made at a single point in time. In reality, many decisions are made over a time horizon due to budget limitations and time constraints. For example, a company may have a limited annual budget for security. This new area of research would address this need for more model fidelity by modeling system design and protection decisions over a time horizon. These new models would generate decisions for an entire time horizon, prescribing decisions for each time period.

This is an interesting area of research because of the potential insights it could generate. For example, this research could reveal patterns of resource allocation over time. In addition, it could reveal insights about the benefit of being able to spread investments over a time horizon rather than having to spend all of the budget at once.

7.3.1.3 Sequential Decision-Making

This new area of research is an extension of the previous section. In the previous section, decisions could be made at different time periods but decisions still had to be made before observing any random outcomes. In contrast, sequential decision-making involves making decisions over time as random outcomes are realized. An example of sequential decision-making is the game of chess, in which moves are made after observing the moves of ones opponent. This is an even more realistic representation of how real decisions are made.

This problem can be modeled using dynamic programming. For the deterministic version of this problem, (deterministic) dynamic programming can be used. For the stochastic version, stochastic dynamic programming can be used to model it and approximate dynamic programming can be used to solve it.

7.3.1.4 Cascading Failures

A majority of the research on networks subject to disruptions assumes that failures only affect a single network element. However, in some networks failures can cascade, meaning that when one network element becomes overloaded, that element fails and its load is shifted to other elements, possibly overloading them. This chain of cascading failures take place until something intervenes or the entire network fails. Cascading failures are most common in power grids—the 2003 blackout in the Northeast US is a recent and notable example (United States-Canada Power System Outage Task Force, 2004).

This is an interesting area because it has the potential to generate new insights about protecting networks. For example, the optimal protection plan for a cascading network could be very different than that of a non-cascading network.

7.3.1.5 Interdependent Infrastructures

Most network disruption models only consider one type of infrastructure in isolation. In reality, infrastructure systems are interdependent of each other. For example, a supply chain is dependent on a power grid while a power grid has its own supply chain.

The main challenge in this area of research is deciding how to couple two infrastructure systems. This is likely to require in-depth knowledge about the two infrastructure systems.

One way to model this problem would be with binary variables that indicate whether or not an infrastructure system is functioning. These binary variables would then relate that infrastructure system to other systems. Another way of modeling inter-dependencies is by having one infrastructure system serve as a supply node for another infrastructure system. A third way would be to use an economic input-output model (Santos, 2006).

7.3.2 New Applications

7.3.2.1 Interdiction of Infectious Disease Outbreaks

Another important problem is that of preparing for and controlling infectious disease outbreaks, which relates to this dissertation because it can be thought of as interdicting the spread of an infectious disease. However, interdicting an infectious disease is more challenging than interdicting facilities because an infectious disease changes in time and space. Future research on infectious disease preparedness and control has at least three components:

1. preparing for a random outbreak;
2. preparing for a terrorist-induced outbreak; and
3. controlling an outbreak once it starts.

Indeed, this is a very challenging problem. The main challenge is that the spread of infectious disease is a complex process. Researchers have modeled disease spread using differential equations, Markov chains, and agent-based simulation, among other frameworks. This represents a new frontier for network disruptions models.

Work in this area could start with an analytical, but possibly non-convex, model of disease spread in the second-stage. This analytical two-stage model could then be solved using a cutting plane strategy. Extending the work on the analytical model, the second-stage could be modeled as a stochastic dynamic program or simulation model and solved using approximate dynamic pro-

gramming. Using this modeling paradigm, the infectious disease problem is an application of the general sequential decision-making paradigm described in Section 7.3.1.3.

7.3.2.2 Interdiction of Crime Hot-Spots

Another important area of future research related to this dissertation is controlling crime. Researchers in the social sciences have long studied crime behavior and how to prevent and control criminal activity, often taking an empirical or qualitative approach. Recently, researchers have begun to take a more quantitative approach in studying the spread of crime in a city (Short et al., 2010). In particular, they have studied crime using a partial differential equations (PDE) model and found that under certain assumptions, criminal activity forms clusters called crime hot-spots. These clusters are caused by a certain criminal behavior observed in the field: criminals tend to commit crimes in places with a lot of criminal activity. The work of Short et al. (2010) could be extended by adding interdiction decisions. These interdiction decisions would represent allocating police forces to different parts of the city. The objective would be to minimize the steady-state crime rate. Again, interdicting criminal activity is harder than interdicting facilities because criminal activity varies in time and space.

The main challenge with modeling and solving this problem is that the second-stage is a partial differential equation. More investigation is needed to determine if this PDE model is a convex function of the first-stage interdiction variables.

This area of research has the potential to generate interesting insights. For example, what are the characteristics of an optimal police patrol policy? Is it best to focus on interdicting crime hot spots or is it better to use another allocation strategy?

A Examples of Network Disruptions

The Powder River Basin, located in eastern Wyoming, supports about 40 percent of coal consumption in the United States. Unfortunately, there is only one rail line, shared by Union Pacific and Burlington Northern-Sante Fe, with which to move coal east from mines in the Powder River Basin. Thus, damages to this line in 2005 caused severe consequences. In particular, because these damages could not be repaired for several weeks, there was a shortage of coal throughout the United States. This shortage caused up to a 15 percent cost increase in some regions and led some power plants to seek other forms of energy (News and Information, 2005).

The incident in the Powder River Basin is one example illustrating that a small disruption to a network can cause significant system-wide consequences. In other words, the supply network for subbituminous coal is vulnerable to disruptions. Because economies throughout the world depend on networks such as transportation systems and supply chains, disruptions to networks can have a large effect on communities. Therefore, natural disasters, severe weather incidents, terrorist attacks, and other events have the potential to severely disrupt our daily lives.

The past decade has seen several natural disasters with major impacts on networks:

- In 2003, an earthquake in Bam, Iran resulted in the death of more than 26,000 people and reconstruction costs were estimated at over \$1 billion. The quake also impaired critical infrastructures such as electricity, telecommunications, and water. In addition, damages to transportation infrastructure made it difficult to deliver humanitarian aid (BBC News, 2003).
- In 2004, an earthquake and tsunami hit the Indian Ocean. This event resulted in the death of an estimated 280,000 people (BBC News, 2005). In addition, damages to ground and marine transportation infrastructure significantly slowed the fishing industry in Sri Lanka (News, 2005).
- In 2005, Hurricane Katrina in New Orleans, Louisiana, resulted in the death of an estimated 1,833 people and damages of \$108 billion (Knabb et al., 2005). Damages caused residents

of the area to be without electricity for weeks (Knabb et al., 2005).

- In 2010, a severe earthquake hit Haiti's capital city of Port-Au-Prince, killing approximately 230,000 people (Washington Post, 2010) and leaving many others homeless. Damages to the main airport and seaport hampered relief efforts (MSNBC, 2010). In addition, the nation's already struggling telecommunication infrastructure was impaired (Rhoads, 2010).
- In 2010, the eruption of a volcano in Iceland halted flights throughout Northeastern Europe for several days (Ulfarsson and Unger, 2011).
- The 2011 Tōhoku earthquake and tsunami in Japan caused significant damage to life and property, including a number of nuclear reactors. This earthquake had worldwide impacts, including delays in the deliveries for Japanese-based companies such as Toyota (Tabuchi, 2011).

Although less extreme than natural disasters, extreme weather events have also caused significant disruptions to networks:

- In 2010, a heat wave in Russia killed thousands of people. The resulting drought and wildfires caused by the heat wave disrupted the service and manufacturing industries (Kim and Levitov, 2010).
- Snowstorms in the United States in 2010 crippled air transportation for several days, leaving many passengers stranded (Carey, 2010)

Several recent, man-made disasters have also caused significant disruptions in networks. In addition to the aforementioned Powder River Basin derailment, the following disruptions are notable:

- In 2003, a series of cascading failures caused a power outage in the Northeast United States, leaving millions without power. This disruption affected many infrastructure networks including municipal water and transportation (CNN, 2003). Overall, this power outage affected about 50 million people (United States-Canada Power System Outage Task Force, 2004).

- In 2002, port authorities along the West Coast staged a labor lockout, endangering the large volume of US goods that are shipped through West Coast ports. Although government intervention ended the lockout, it is estimated that a major shutdown to West Coast ports could trigger a crisis in international financial markets (Cohen, 2005).
- In 2010, an explosion on the Deepwater Horizon drilling rig triggered a massive oil spill in the Gulf of Mexico. In the aftermath of the disaster BP agreed to pay \$7.8 billion to settle a lawsuit (BBC News, 2012). Fortunately, the spill has since been contained. However, the long-term effects on the environment and industries such as commercial fishing are unclear (Mervin, 2011).

Terror attacks are a particularly devastating type of man-made disaster. Several notable attacks have occurred in the past 15 years. The most notable attack was on the World Trade Center towers in New York City on September 11, 2001. Almost 3,000 people died and short-term economic losses are estimated at \$33 billion to \$36 billion (Bram et al., 2002). Another form of terrorism is the piracy that has taken place off the coast of Somalia. Researchers have estimated that this piracy costs about \$6 billion per year (Oceans Beyond Piracy, 2012). In addition, after the death of Osama Bin Laden, it was revealed that Al-Qaeda was planning an attack on the rail infrastructure in the United States (Entous et al., 2011).

B Extra Material for Chapter 4

B.1 Datasets Used in Experimentation

The following Table contains the 18 datasets used in the experimentation. The third column of Table B.1 indicates which datasets have weighted demand points and which do not. If demand points have weights, the distance values d_{ij} usually do not obey the triangle inequality.

Table B.1: Datasets used in experimentation

no.	name	$ \mathcal{S} = J $	weights	source of data	reference
1	s55	55	yes	population centers in Washington, D.C.	Swain (1971)
2	d49	49	yes	49 US state capitals and Washington, D.C.	Daskin (1995)
3	d88	88	yes	cities in US	Daskin (1995)
4	d150	150	yes	cities in US	Daskin (1995)
5	lor100	100	yes	population centers in San Jose Dos Campos, Brazil	Lorena and Senne (2004)
6	lon150	150	yes	population centers in London, Ontario	Alp et al. (2003)
7	lor200	yes	yes	population centers in San Jose Dos Campos, Brazil	Lorena and Senne (2004)
8	lor300a	300	yes	population centers in San Jose Dos Campos, Brazil	Lorena and Senne (2004)
9	lor300b	300	yes	population centers in San Jose Dos Campos, Brazil	Lorena and Senne (2004)
10	lor400a	402	yes	population centers in San Jose Dos Campos, Brazil	Lorena and Senne (2004)
11	lor400b	402	yes	population centers in San Jose Dos Campos, Brazil	Lorena and Senne (2004)
12	beas500	500	no	network distances	Beasley (1990)
13	beas600	600	no	network distances	Beasley (1990)
14	beas700	700	no	network distances	Beasley (1990)
15	beas800	800	no	network distances	Beasley (1990)
16	lor818	818	no	population centers in San Jose Dos Campos, Brazil	Lorena and Senne (2004)
17	beas900	900	no	network distances	Beasley (1990)
18	u1060	1060	no	drilling problem TSP	Reinelt (1991)

B.2 Effect of Weights on Binary Search Algorithm

In this section we try to determine if weighted or unweighted instances require more computation time for the binary search algorithm. We used our two largest weighted datasets, lor400a and

lor400b, for our analysis. Table B.2 shows the computation times and number of binary search iterations for the weighted and unweighted versions of several instances of the RANPCP. When the optimal solution is found after the lower and upper bounds stage because $lb = ub$, the number of iterations is listed as 0. Each row shows the computation time and number of iterations required for the weighted and unweighted versions of an instance. From the results in this table, it is difficult to discern if the weighted or unweighted problem requires more computation. In half of the instances, the unweighted version requires more computation time. This is also true for the number of iterations.

Table B.2: Binary search computational results for wtd. and unwtd. datasets

No.	Dataset	p	r	$\frac{p}{ \mathcal{D} }$	$\frac{r}{p}$	Time (s)			Number of iterations		
						wtd	unwtd	wtd-unwtd	wtd	unwtd	wtd-unwtd
1	lor402a	5	1	0.012	0.2	3.5	3.2	0.33	18	16	2
2	lor402a	5	2	0.012	0.4	4	3.9	0.08	14	16	-2
3	lor402a	10	1	0.025	0.1	2.9	3.8	-0.86	17	15	2
4	lor402a	10	2	0.025	0.2	3.5	3.4	0.15	13	15	-2
5	lor402a	40	1	0.100	0.025	5.4	14.0	-8.5	16	13	3
6	lor402a	40	2	0.100	0.05	0.1	15.0	-15	0	14	-14
7	lor402b	5	1	0.012	0.2	3.6	3.1	0.54	18	16	2
8	lor402b	5	2	0.012	0.4	4	3.9	0.1	14	16	-2
9	lor402b	10	1	0.025	0.1	3	3.8	-0.8	17	15	2
10	lor402b	10	2	0.025	0.2	3.6	3.4	0.15	13	15	-2
11	lor402b	40	1	0.100	0.025	5.4	14.0	-8.5	16	13	3
12	lor402b	40	2	0.100	0.05	0.1	15.0	-15	0	14	-14
Min						0.1	3.1	-15	0	13	-14
Max						5.4	15	0.54	18	16	3
Average						3.3	7.2	-3.9	13	15	-1.8

B.3 Saturation

In this section we prove several structural properties of the RANPCP. First, we explain a property called *saturation*. This means that an instance of the RANPCP is saturated if the r closest facilities are located for a given demand point and the distance from that demand point and its r^{th} closest located facility is equal to the optimal objective value. When an instance is saturated for a given

value of p and r , locating additional facilities does not improve the objective. We also show the implications that saturation has on the LP relaxations of models $M2$ and $M2 - C$. In particular, we show that when a problem instance is saturated, the LP relaxations of models $M2$ and $M2 - C$ form the convex hull of the RANPCP.

Let $V(j', r)$ be the optimal objective value for an instance of the RANPCP with j' facilities and r neighbors. An instance of the RANPCP is said to be *saturated* for a given j' and r if an optimal solution exists that has an objective value of $\max_j \{d_{i_j^r, j}\} = V(p, r)$. We call the quantity $\max_j \{d_{i_j^r, j}\}$ the *saturation objective*.

Lemma 1. *For an instance with r neighbors, the saturation objective is obtained when the r closest facilities to demand point j are located, where $j = \arg \max_j \{d_{i_j^r, j}\}$.*

Proof. By the definition of saturated, $\max_j \{d_{i_j^r, j}\} = V(p, r)$. Let $j = \arg \max_j \{d_{i_j^r, j}\}$. By the definition of the RANPCP, the distance from j and each of its r closest located facilities must be less than $V(p, r) = \max_j \{d_{i_j^r, j}\}$. However, the distance from j and each of its r closest located facilities can only be less than $V(p, r) = \max_j \{d_{i_j^r, j}\}$ if the r closest facilities to j are located. \square

Theorem 1. *If an instance is saturated for a given j' ($j' \leq |J| - 1$) and r , then $V(j', r) = V(j' + 1, r)$ and the instance is also saturated for $j' + 1$ and r .*

Proof. By the definition of saturation, there exists a $j \in \mathcal{J}$ such that $d_{i_j^r, j} = V(p, r)$. By Lemma 1, the r closest facilities to j have been located. Let $p = p + 1$. Thus, one new facility can be located. Wherever the new facility is located, it will be at least as far from demand point j as facility i_j^r . As a result, this additional facility location would not change the distance from j to its r^{th} closest located facility and the optimal objective value is not changed. Thus, $V(j', r) = d_{i_j^r, j} = V(p + 1, r)$ and therefore the instance is saturated for $j' + 1$ and r . \square

Let $V(LM2, p, r)$ be the optimal objective value for the linear programming relaxation of model $M2$.

Theorem 2. $V(LM2, p, r) \geq \max_j \{d_{i_j^r, j}\}$.

Proof. Because the safe assignment variables $Z_{i_j^\ell}$ are only defined for $\ell \geq r$, the left hand side of constraint (3b) for a demand point $j \in \mathcal{J}$ must equal a convex combination of $(d_{i_j^\ell})_{\ell=r, \dots, |\mathcal{J}|}$. This convex combination is minimized when $Z_{i_j^r} = 1$ with a minimum value of $d_{i_j^r}$. As a result, the optimal LP relaxation objective value $V(LM2, p, r)$ is greater than or equal to the maximum of these lower bounds, namely $\max_j \{d_{i_j^r}\}$. \square

The following corollaries follow from Theorem 2.

Corollary 2. *If an instance of the RANPCP is saturated for a given j' and r , then $V(LM2, p, r) = V(j', r)$.*

Proof. Because model $LM2$ is a relaxation, $V(LM2) \leq V(j', r)$. By the definition of saturation, $\max_j \{d_{i_j^r}\} = V(p, r)$. Because $V(LM2, p, r) \geq \max_j \{d_{i_j^r}\}$, (Theorem 2), we have that $V(LM2) = V(j', r)$. \square

Let $V(LM2 - C, p, r)$ be the optimal objective value of the LP relaxation of model $M2 - C$.

Corollary 3. *If an instance of the RANPCP is saturated for a given j' and r , then $V(LM2 - C, p, r) = V(j', r)$*

Proof. Because the instance is saturated, $V(j', r) = V(LM2, p, r)$ (Corollary 2). Because model $M2$ is a relaxation of $M2 - C$, $V(LM2, p, r) \leq V(LM2 - C, p, r)$. Because $LM2 - C$ is a relaxation, $V(LM2 - C, p, r) \leq V(p, r)$. Thus, $V(LM2 - C, p, r) = V(j', r)$. \square

Clearly Corollary 3 implies that for saturated instances, $M2 - C$ results in the same linear programming relaxation of $M2$.

An instance of the RANPCP with a given j' and r is said to be *LP saturated* for a particular model if the optimal objective value of the LP relaxation of that model is equal to $\max_j \{d_{i_j^r}\}$. Note that it is possible for an instance to be LP saturated for a given model when the instance is not saturated. The following corollary follows from Theorem 1.

Corollary 4. *If an instance of the RANPCP is LP saturated for model M2 with a given j' ($j' \leq |J| - 1$) and r , then $V(LM2, j', r) = V(LM2, j' + 1, r)$ and the instance is LP saturated for model M2 for $j' + 1$ facilities and r neighbors.*

Proof. By the definition of LP saturation, the optimal objective of the LP relaxation, $V(LM2, j', r)$, equals $\max_j \{d_{r,j}\}$. Let $j = \arg \max_j \{d_{r,j}\}$. Since the optimal objective is $\max_j \{d_{r,j}\}$, the r closest facilities to j must have been located. Otherwise, the distance from j to its r^{th} closest facility would be greater than $\max_j \{d_{r,j}\}$. Let $p = p + 1$. Thus, one new facility can be located. Wherever the new facility is located, it will be at least as far from demand point j as facility i'_j . As a result, this additional facility location would not change the distance from j to its r^{th} closest located facility and the optimal objective value is not changed. Thus, $V(LM2, j', r) = d_{r,j} = V(LM2, j' + 1, r)$ and therefore the instance is saturated for $j' + 1$ and r . □

C Extra Material for Chapter 5

C.1 Proofs

This section contains proofs for the propositions and corollaries that appear in Chapter 5.

Proposition 2. *If i' and j' form a bottleneck pair, then the set $\mathcal{O}^*(J, \mathcal{H})$ is composed of all of the facilities that are closer to j' than i' .*

Proof. (Proof by contradiction.) Suppose there exists an optimal strategy to the interdictor's problem such that a facility $i^0 \in J_{j'}$ is not destroyed. Because i^0 is not destroyed and it is closer to j' than i' , j' will be assigned to it after the interdiction stage. Because j' is no longer assigned to i' after the interdiction stage, i' and j' no longer form a bottleneck pair. \square

Corollary 5. *If i' and j' form a post-interdiction bottleneck pair, then either 1) i' is hardened, or 2) there are at least r located facilities that are closer to j' than i' .*

Proof. **Part (1):** If i' is hardened, then it can always be a post-interdiction bottleneck facility because it cannot fail. **Part (2):** (By contradiction.) Suppose there are $s < r$ facilities that are closer to i' than j' . By Proposition 1, the interdictor's optimal strategy is to destroy the s closest facilities to j' . Because $s < r$, the interdictor can increase his objective by destroying i' , causing j' to be assigned to some other facility after the interdiction stage, namely facility i^0 , which is closer to j' than i' . Because j' is not assigned to i' after disruptions, i' is not a bottleneck demand point. \square

Proposition 3. *If less than $\frac{1}{1+H}$ of the budget is used to locate facilities, then the total cost of locating and hardening will be less than the total budget.*

Proof. Suppose $\lambda^0 < \frac{1}{1+H}$ of the budget is used to locate facilities. Then $\lambda^1 = H\lambda^0$ of the budget will be needed to harden those facilities. The proportion of the budget used for locating and hardening will then be $\lambda^0 + \lambda^1 = \lambda^0 + H\lambda^0 < \frac{1}{1+H} + H\left(\frac{1}{1+H}\right) = 1$, because $\lambda^0 < \frac{1}{1+H}$. \square

C.2 Other Models Tested

In addition to formulation (5.4) (see Section (5.3.3)), we also tried an alternate formulation of the integrated MFLHP model in which X_i is a variable that is 1 if a facility is located at i and 0 otherwise, and Z_i is a variable that is 1 if a facility at i is hardened and 0 otherwise. (Note that when a facility is located and hardened at a location i , both X_i and Z_i equal 1.) This alternate model is as follows:

$$(ALT) \quad \min \quad U \quad (C.1a)$$

$$\text{s.t.} \quad \phi_{ij}W_{ij} \leq U \quad \forall i \in I, j \in J, \quad (C.1b)$$

$$(r+1)W_{ij} \leq (r+1)Z_i + \sum_{i': \phi_{i'j} < \phi_{ij}} X_{i'} \quad \forall i \in I, j \in J, \quad (C.1c)$$

$$\sum_{i \in \mathcal{I}} W_{ij} = 1 \quad \forall j \in \mathcal{J}, \quad (C.1d)$$

$$W_{ij} \leq X_i \quad \forall i \in I, j \in \mathcal{J}, \quad (C.1e)$$

$$Z_i \leq X_i \quad \forall i \in I, \quad (C.1f)$$

$$\sum_{i \in I} f_i X_i + \sum_{i \in I} g_i Z_i \leq b, \quad (C.1g)$$

$$X_i, Z_i \in \{0, 1\} \quad \forall i \in I, \quad (C.1h)$$

$$W_{ij} \in \{0, 1\} \quad \forall i \in I, j \in J. \quad (C.1i)$$

Formulation (C.1) is identical to formulation (5.4) except for the definition of the location and hardening variables, X_i and Z_i . To account for the new definition, Constraints (C.1e) allow a safe assignment to be made between i and j only if a facility is located at i . Constraints (C.1f) allow a facility to be hardened at location i only if a facility is located at i .

Formulation (C.1) gave inferior (lower) linear-programming relaxations and higher run times than Model (5.4). However, neither model was a clear winner with regard to run time.

We also tried formulating the SCP-LH (see Section 5.4.1) as a modified version of (5.5) in which X_i is a variable that is 1 if a facility is located at i and 0 otherwise, and Z_i is a variable that is 1 if a facility at i is hardened and 0 otherwise. However, this modified model gave inferior (lower)

linear-programming relaxations than Model (5.5). In addition, the binary search algorithm solved faster with Model (5.5) as the auxiliary problem than with this modified Model (See Appendix C.3).

C.3 Comparison of Auxiliary Problems for Integrated MFHLP Binary Search

Table C.1 shows run times for the various auxiliary problems described in Section C.2.

Table C.1: Run time (s) of binary search algorithm with different auxiliary problems for u1060 dataset with $B = 0.1$

B	H	r	Run time (s)		Std.-LWFV
			Std.	LWFV	
0.1	2	159	20	21	-2
0.1	0.5	159	31	30	-1
0.1	2	106	25	32	-8
0.1	1	106	31	32	-3
0.1	0.5	106	30	30	-1
0.1	2	53	61	35	24
0.1	3	2	76	28	47
0.1	3	1	47	39	6
0.1	1	1	2362	71	2289
0.1	2	1	61	44	16
0.1	0.5	1	119	48	70
Average					221

C.4 Binary Search Algorithm

The binary search algorithm discussed in Section 5.4.2 is as follows.

1. **Initialize:** Let UB and LB be initial upper and lower bounds for the integrated MFLHP. Let $D = \{\phi_{ij} : i \in \mathcal{I}, j \in \mathcal{J}, \phi_{ij} \geq LB, \phi_{ij} \leq UB\}$ be the set of all inter-node distances that are within the initial upper and lower bounds. Set $lbIndex = 0$ and $ubIndex = |D| - 1$. Let (X^0, Z^0) denote the best feasible location and hardening solution.
2. Set $index = lbIndex + \lceil \frac{ubIndex - lbIndex}{2} \rceil$.

3. If $lbIndex = ubIndex$, RETURN D_{index} as the optimal post-interdiction radius and (X^0, Z^0) as the optimal location and hardening solution.
4. Obtain a heuristic solution to SCP-LH(D_{index}). Let $(\tilde{X}^*, \tilde{Z}^*)$ be the set of located and hardened facilities and $\chi(\tilde{X}^*, \tilde{Z}^*) = \sum_{i \in \mathcal{J}} f_i \tilde{X}^* + \sum_{i \in \mathcal{J}} g_i \tilde{Z}^*$ be the cost of the solution. IF $\chi(\tilde{X}^*, \tilde{Z}^*) \leq b$, set $ubIndex = index$ and return to Step 2.
5. Let $\delta(\tilde{X}^*, \tilde{Z}^*)$ be the post-interdiction radius for solution $(\tilde{X}^*, \tilde{Z}^*)$. IF $\delta(\tilde{X}^*, \tilde{Z}^*) < D_{ubIndex}$ set $ubIndex = index$ and return to Step 2.
6. Solve SCP-LH(D_{index}) to obtain solution (X^*, Z^*) . IF $\chi(X^*, Z^*) > b$, set $lbIndex = index + 1$; ELSE, set $ubIndex = index$ and set $(X^0, Z^0) = (X^*, Z^*)$. Return to Step 2.

In Step 4, a greedy heuristic is used to solve the auxiliary problem. The idea is that during the course of the binary search, the heuristic will sometimes find a new upper bound. Because these upper bounds are found heuristically, the number of times SCP-LH must be solved to optimality is fewer. This greedy heuristic is described in Appendix C.5.

C.5 Greedy Heuristic for Auxiliary Sub Problem

The greedy heuristic mentioned in Section 5.4.2 assigns scores to demand points and facilities and sequentially adds the facilities with the highest score. A demand point is *safe-covered* if the distance to its post-disruption bottleneck facility is within δ . Table C.2 shows some of the notation used for the heuristic. Let $m_j = |M_j|$, $n_j = |N_j|$, and $o_j = |O_j|$. To compute the *cover criticality score* for demand point j , first let $\mu_j = \frac{1}{1+1/H+1/(1+H)}$, $\nu_j = \frac{1/H}{1+1/H+1/(1+H)}$, and $\rho_j = \frac{1/(1+H)}{1+1/H+1/(1+H)}$ be the normalizing weights assigned to m_j , n_j , and o_j , respectively. The cover criticality score for demand point j is $q_j = \frac{p_j}{\mu_j m_j + \nu_j n_j + \rho_j o_j}$. Let the *location fitness score* of facility i be $r_i = \frac{\sum_{j \in B_i} p_j}{f_i / \sum_i f_i}$, the *hardening fitness score* be $s_i = \frac{\sum_{j \in C_i} p_j}{g_i / \sum_i g_i}$, and the *location-and-hardening fitness score* be $t_i = \frac{\sum_{j \in D_i} p_j}{(f_i + g_i) / \sum_i (f_i + g_i)}$.

Table C.2: Notation for set cover heuristic

M_j	the set of facilities whose location makes j safe-covered
N_j	the set of located facilities whose hardening makes j safe-covered
O_j	the set of facilities whose location and hardening makes j safe-covered
p_j	the smallest number of located facilities needed to make j safe-covered
I_j	the facilities whose location will make j safe-covered
B_j	the set of additional demand points that will be safe-covered if a facility is located at i
C_i	the set of additional demand points that will be safe-covered if located facility at i is hardened
D_i	the set of additional demand points that will be safe-covered if facility i is located and hardened

The greedy heuristic has the following steps:

1. **Initialize.** Let L be the set of located facilities and H be the set of hardened facilities. Let $L = H = \emptyset$.
2. Sort demand points by q_j .
3. Choose $\hat{i} = \arg \max_{i \in M_j} \{r_i\}$, $\bar{i} = \arg \max_{i \in N_j} \{s_i\}$, and $i^0 = \arg \max_{i \in N_j} \{t_i\}$. If $r_{\hat{i}} > \max\{s_{\bar{i}}, t_{i^0}\}$, locate facility \hat{i} and add \hat{i} to L . If $s_{\bar{i}} > \max\{r_{\hat{i}}, t_{i^0}\}$, harden facility \bar{i} , and add \bar{i} to H . If $t_{i^0} > \max\{r_{\hat{i}}, s_{\bar{i}}\}$, locate and harden facility i^0 , and add i^0 to L and H .
4. Continue Steps 2 and 3 until all demand points are safe-covered.
5. **Remove redundant locations and hardenings.**
 - (a) Remove all unhardened facilities from L whose location is not needed to safe-cover all demand points. Remove all hardened facilities from H whose hardening is not needed to safe-cover all demand points. Remove all hardened facilities from L and H for which the removal of the facility from L and H still allows all demand points to be safe-covered.

(b) Continue until there are no redundant facilities or hardenings.

6. Return solution (L, H) .

C.6 Hardening Model

The hardening model discussed in Section 5.5.1 prescribes the optimal set of existing facilities to harden in order to minimize the post-interdiction radius:

$$\min U \tag{C.2a}$$

$$\text{s.t. } U \geq \phi_{ij} W_{ij} \quad \forall j \in J, j \in \mathcal{J} \tag{C.2b}$$

$$W_{i^r, j} \leq Z_i \quad \forall j \in J, r = 1, \dots, r \tag{C.2c}$$

$$\sum_{j \in J} g_j Z_j \leq b \tag{C.2d}$$

$$Z_i \in \{0, 1\} \quad \forall i \in J \tag{C.2e}$$

$$W_{ij} \in \{0, 1\} \quad \forall j \in J, j \in \mathcal{J} \tag{C.2f}$$

The objective equation (C.2a), in conjunction with Constraints equation (C.2b), is to minimize the post-interdiction radius. Constraints equation (C.2c) model the requirement that a demand point j can only form a post-interdiction bottleneck pair with facility i if 1) facility i is hardened or 2) the facility is further from j than the r^{th} closest facility to j . Constraint equation (C.2d) requires that the amount spent on hardening must be within a budget. Constraints equation (C.2e)–equation (C.2f) specify bounds on the variables.

Model (C.2) can also be solved using a binary search algorithm, with the following axillary sub problem:

$$\text{SCP-H}(\delta) \quad \min \sum_{j \in J} g_j Z_j \quad (\text{C.3a})$$

$$\text{s.t.} \quad \sum_{i \in J: \phi_{ij} \leq \delta} Z_i \geq 1 \quad \forall j \in \{j : d_{i_j^{r+1} j} > \delta\} \quad (\text{C.3b})$$

$$Z_i \in \{0, 1\} \quad \forall j \in J, \quad (\text{C.3c})$$

where Constraints equation (C.3b) require that every demand point must either have $(r + 1)$ facilities closer than δ or a facility closer to j than δ must be hardened. The SCP-H problem is the same as the standard set-cover problem except that the cover constraints in equation (C.3b) are not present for every demand point. Therefore, algorithms that work well for the set cover problem are likely to work well for Model (C.3).

C.7 Datasets Used

Table C.3 describes the datasets used in the experimentation. Column 2 shows the number of nodes in the dataset. Columns 3 and 4 indicate whether the demand weights and location costs are homogeneous (H) or non-homogeneous (NH). Column 6 indicates the distance metric used: Euclidean, Great Circle, road distances measured from real data, or network distances.

Table C.3: Datasets used in experimentation

Name	$ \mathcal{S} = J $	Demand weights	Location costs	Source of data	Distance measure	Ref.
(1)	(2)	(3)	(4)	(5)	(6)	(7)
sw55	55	NH	H	population centers in Washington, D.C.	Euclidean	Swain (1971)
d49	49	NH	NH	49 US state capitals and Washington, D.C.	Great circle	Daskin (1995)
d88	88	NH	NH	cities in US	Great circle	(Daskin, 1995)
d150	150	NH	NH	cities in US	Great circle	(Daskin, 1995)
lor100	100	NH	H	population centers in San Jose Dos Campos, Brazil (SJDC)	Road	Lorena and Senne (2004)
lon150	150	NH	H	population centers in London, Ontario	Road	Alp et al. (2003)
lor200	yes	NH	H	SJDC	Road	(Lorena and Senne, 2004)
lor300a	300	NH	H	SJDC	Road	(Lorena and Senne, 2004)
lor402a	402	NH	H	SJDC	Road	(Lorena and Senne, 2004)
beas500	500	H	H	hypothetical network	Network distances	Beasley (1990)
beas600	600	H	H	hypothetical network	Network distances	(Beasley, 1990)
beas700	700	H	H	hypothetical network	Network distances	(Beasley, 1990)
beas800	800	H	H	hypothetical network	Network distances	(Beasley, 1990)
lor818	818	H	H	SJDC	Road	(Lorena and Senne, 2004)
beas900	900	H	H	hypothetical network	Network distances	(Beasley, 1990)
u1060	1060	H	H	drilling problem TSP	Euclidean	Reinelt (1991)
rl1323	1323	H	H	drilling problem TSP	Euclidean	Reinelt (1991)

C.8 Comparison of sequential and integrated methods

Table C.4 displays summary statistics for the LTHI method. Columns (2)–(5) show summary statistics for the relative objective function increase incurred when the LTH method is used instead of the integrated method. Columns (6)–(9) show summary statistics for the deviation in proportion. Columns (10)–(13) show summary statistics for the relative objective function increase incurred for using the LTHI method in place of the integrated method. Columns (2)–(5) show that the difference between the ideal proportion and the optimal proportion can be both positive and negative. The table also shows that the average relative objective function increase for not using the integrated method is between 0.55 and 3.15.

Table C.4: Comparison of sequential and integrated methods: summary

Dataset (1)	$\hat{L} - \lambda^*$				$\gamma(\hat{L})$			
	Avg. (2)	Min. (3)	Max. (4)	Std. dev. (5)	Avg. (6)	Min. (7)	Max. (8)	Std. dev. (10)
sw55	0.04	0.00	0.45	0.09	0.55	0.01	1.91	0.43
lor100	-0.03	-0.55	0.53	0.35	3.12	1.00	5.99	1.41
lon150	-0.13	-0.73	0.60	0.32	1.31	0.13	2.87	0.73
lor200	0.08	-0.51	0.53	0.35	3.15	1.00	8.02	1.62

Bibliography

- Ahuja, R. K., Magnanti, T. L., and Orlin, J. B. (1993) *Network Flows: Theory, Algorithms, and Applications*. Prentice Hall, Upper Saddle River, NJ, 1 edition.
- Aksen, D., Aras, N., and Piyade, N. (2011) A Bilevel P-Median Model for the Planning and Protection of Critical Facilities. *Journal of Heuristics*, pp. 1–26.
- Aksen, D., Piyade, N., and Aras, N. (2010) The Budget Constrained R-Interdiction Median Problem with Capacity Expansion. *Central European Journal of Operations Research*, 18(3), 269–291.
- Albert, R., Jeong, H., and Barabasi, A.-L. (2000) Error and Attack Tolerance of Complex Networks. *Nature*, 406(6794), 378–382.
- Alderson, D. L. (2008) Catching the "Network Science" Bug: Insight and Opportunity for the Operations Researcher. *Operations Research*, 56(5), 1047–1065.
- Alderson, D. L., Brown, G. G., Carlyle, W. M., and Wood, R. K. (2011) Solving Defender-Attacker-Defender Models for Infrastructure Defense. In *Proceedings of the 12th INFORMS Computing Society Conference: Research, Computing, and Homeland Defense*, Wood, R. K. and Dell, R. F., eds. INFORMS, pp. 28–49.
- Alp, O., Erkut, E., and Drezner, Z. (2003) An Efficient Genetic Algorithm for the P-Median Problem. *Annals of Operations Research*, 122(1), 21–42.
- Ambs, K., Cwilich, S., Deng, M., Houck, D. J., Lynch, D. F., and Yan, D. (2000) Optimizing restoration capacity in the AT&T network. *Interfaces*, 30(1), 26–44.
- Azaiez, M. N. and Bier, V. M. (2007) Optimal Resource Allocation for Security in Reliability Systems. *European Journal of Operational Research*, 181(2), 773–86.
- Bailey, M. D., Shechter, S. M., and Schaefer, A. J. (2006) SPAR: Stochastic programming with adversarial recourse. *Operations Research Letters*, 34(3), 307–315.
- Balakrishnan, A., Magnanti, T. L., and Sokol, J. S. (2001) Telecommunication Link Restoration Planning with Multiple Facility Types. *Annals of Operations Research*, 106(1), 127–154.
- Balakrishnan, A., Magnanti, T. L., Sokol, J. S., and Wang, Y. (2002) Spare-Capacity Assignment for Line Restoration Using a Single-Facility Type. *Operations Research*, 50(4), 617–635.
- Balas, E. and Ho, A. (1980) Set Covering Algorithms Using Cutting Planes, Heuristics, and Sub-gradient Optimization: a Computational Study. In *Combinatorial Optimization*, Padberg, M. W., ed., Elsevier North-Holland, The Netherlands, volume 12 of *Mathematical Programming Studies*, pp. 37–60.
- Ball, M. O., Colbourn, C. J., and Provan, J. S. (1995) Network Reliability. In *Network Models*, Ball, M. O., Magnanti, T. L., Monma, C. L., and Nemhauser, G. L., eds., Elsevier Science, The Netherlands, volume 7 of *Handbooks in Operations Research and Management Science*, chapter 11, pp. 673–762.

- Barabasi, A. L. and Albert, R. (1999) Emergence of scaling in random networks. *Science (New York, N.Y.)*, 286(5439), 509–512.
- Bard, J. F. (1998) *Practical bilevel optimization: algorithms and applications*. Kluwer Academic Publishers, The Netherlands.
- Batta, R., Dolan, J. M., and Krishnamurthy, N. N. (1989) The Maximal Expected Covering Location Problem: Revisited. *Transportation Science*, 23(4), 277–87.
- Bayrak, H. and Bailey, M. D. (2008) Shortest Path Network Interdiction with Asymmetric Information. *Networks*, 52(3), 133–140.
- BBC News (2003) Iran earthquake kills thousands. *BBC News*.
- BBC News (2005) Indonesia quake toll jumps again. *BBC News*.
- BBC News (2012) BP reaches \$7.8bn oil spill deal. *BBC News*.
- Beasley, J. E. (1990) OR-Library: Distributing test problems by electronic mail. *Journal of the Operational Research Society*, 41(11), 1069–1072.
- Berman, O. and Krass, D. (2002) Facility location problems with stochastic demands and congestion. In *Facility Location: Applications and Theory*, Springer, chapter 11, pp. 329–371.
- Berman, O., Krass, D., and Menezes, M. B. C. (2007) Facility Reliability Issues In Network P-Median Problems: Strategic Centralization And Co-Location Effects. *Operations Research*, 55(2), 332–350.
- Berman, O., Krass, D., and Menezes, M. B. C. (2009) Locating Facilities in the Presence of Disruptions and Incomplete Information. *Decision Sciences*, 40(4), 845–868.
- Beygelzimer, A., Grinstein, G., Linsker, R., and Rish, I. (2005) Improving Network Robustness by Edge Modification. *Physica A: Statistical Mechanics and its Applications*, 357(3-4), 593–612.
- Bier, V., Cox, L., and Azaiez, M. N. (2009) Why Both Game Theory and Reliability Theory Are Important in Defending Infrastructure against Intelligent Attacks. In *Game Theoretic Risk Analysis of Security Threats*, Bier, V. M. and Azaiez, M. N., eds., Springer, New York, International Series in Operations Research and Management Science, pp. 1–11.
- Bier, V., Oliveros, S., and Samuelson, L. (2007a) Choosing What to Protect: Strategic Defensive Allocation Against an Unknown Attacker. *Journal of Public Economic Theory*, 9(4), 563–587.
- Bier, V. M. (2007) Choosing What to Protect. *Risk Analysis*, 27(3), 607–620.
- Bier, V. M. and Abhichandani, V. (2003) Optimal Allocation of Resources for Defense of Simple Series and Parallel Systems from Determined Adversaries. In *Proceedings of the Tenth Risk-based Decisionmaking in Water Resources X*. American Society of Civil Engineers, p. 59.
- Bier, V. M., Gratz, E. R., Haphuriwat, N. J., Magua, W., and Wierzbicki, K. R. (2007b) Methodology for Identifying Near-Optimal Interdiction Strategies for a Power Transmission System. *Reliability Engineering and System Safety*, 92(9), 1155–61.

- Bier, V. M., Haphuriwat, N., Menoyo, J., Zimmerman, R., and Culpén, A. M. (2008) Optimal Resource Allocation for Defense of Targets Based On Differing Measures of Attractiveness. *Risk Analysis*, 28(3), 763–770.
- Bier, V. M., Nagaraj, A., and Abhichandani, V. (2005) Protection of Simple Series and Parallel Systems with Components of Different Values. *Reliability Engineering and System Safety*, 87(3), 315–323.
- Birge, J. R. and Louveaux, F. (1997) *Introduction to stochastic programming*, volume 9. Springer-Verlag, New York.
- Birge, J. R. and Louveaux, F. V. (1988) A Multicut Algorithm for Two-stage Stochastic Linear Programs. *European Journal of Operational Research*, 34(3), 384–392.
- Bolić, T. and Sivčev, v. (2011) Eruption of Eyjafjallajökull in Iceland. *Transportation Research Record: Journal of the Transportation Research Board*, 2214(-1), 136–143.
- Bram, J., Orr, J., and Rapaport, C. (2002) Measuring the Effects of the September 11 Attack on New York City. *Economic Policy Review*, 8(2), 5–20.
- Brotcorne, L., Laporte, G., and Semet, F. (2003) Ambulance Location and Relocation Models. *European Journal of Operational Research*, 147(3), 451–463.
- Brown, G., Carlyle, M., Salmerón, J., and Wood, K. (2006) Defending Critical Infrastructure. *Interfaces*, 36(6), 530–544.
- Brown, G. G., Carlyle, W. M., Royset, J. O., and Wood, R. K. (2005a) On the Complexity of Delaying an Adversary’s Project. In *The Next Wave in Computing, Optimization, and Decision Technologies*, Sharda, R., Voß, S., Golden, B., Raghavan, S., and Wasil, E., eds., Springer US, volume 29 of *Operations Research/Computer Science Interfaces Series*, chapter 1, pp. 3–17.
- Brown, G. G., Carlyle, W. M., Salmerón, J., and Wood, K. (2005b) Analyzing the Vulnerability of Critical Infrastructure to Attack and Planning Defenses. In *TutORials in Operations Research: Emerging Theory, Methods, and Applications*, Smith, J. C., ed., INFORMS, Hanover, MD, TutORials in Operations Research, chapter 4, pp. 102–123.
- Brown, P. S. (2005) *Optimizing the long-term capacity expansion and protection of Iraqi oil infrastructure*. Master’s thesis, Naval Postgraduate School.
- Bundschuh, M., Klabjan, D., Pei, P., and Thurston, D. L. (2006) Modeling Robust and Reliable Supply Chains. *Optimization Online*.
- Byrka, J., Srinivasan, A., and Swamy, C. (2010) Fault-tolerant facility location: a randomized dependent LP-rounding algorithm. In *Proceeding of the 14th Conference on Integer Programming and Combinatorial Optimization*, Eisenbrand, F. and Shepherd, B., eds. Springer-Verlag, volume 6080 of *Lecture Notes in Computer Science*, pp. 244–257.
- Cappanera, P. and Scaparra, M. P. (2011) Optimal Allocation of Protective Resources in Shortest-Path Networks. *Transportation Science*, 45(1), 64–80.

- Carey, S. (2010) Storm wreaks havoc on travel. *wsj.com*.
- Carrigy, A., Ramirez-Marquez, J. E., and Rocco, C. M. (2010) Multistate Stochastic Network Interdiction Via Reliability Modelling and Evolutionary Optimization. *Proceedings of the Institution of Mechanical Engineers, Part O: Journal of Risk and Reliability*, 224(1), 27–42.
- Chaudhuri, S., Garg, N., and Ravi, R. (1998) The P-Neighbor K-Center Problem. *Information Processing Letters*, 65(3), 131–134.
- Church, R. and ReVelle, C. (1974) The Maximal Covering Location Problem. *Papers in regional science*, 32(1), 101–118.
- Church, R. L. (2003) COBRA: A new formulation of the classic p-median location problem. *Annals of Operations Research*, 122(1), 103–120.
- Church, R. L. and Gerrard, R. A. (2003) The Multi-Level Location Set Covering Model. *Geographical Analysis*, 35(4), 277–290.
- Church, R. L. and Scaparra, M. P. (2006) Analysis of Facility Systems Reliability When Subject to Attack or a Natural Disaster. In *Critical Infrastructure: Reliability and Vulnerability*, Murray, A. T. and Grubestic, T. H., eds., Springer-Verlag, Berlin, Germany, chapter 11, pp. 221–241.
- Church, R. L. and Scaparra, M. P. (2007) Protecting Critical Assets: the R-Interdiction Median Problem With Fortification. *Geographical Analysis*, 39(2), 129–146.
- Church, R. L., Scaparra, M. P., and Middleton, R. S. (2004) Identifying Critical Infrastructure: The Median and Covering Facility Interdiction Problems. *Annals of the Association of American Geographers*, 94(3), 491–502.
- Church, R. L., Scaparra, M. P., and O’Hanley, J. R. (2005) Optimizing Passive Protection in Facility Systems. In *ISOLDE X. Spain*.
- CNN (2003) Major power outage hits new york, other large cities. *CNN*.
- Cohen, S. S. (2005) Economic Impact of a West Coast Dock Shutdown. Technical report, University of California at Berkeley.
- Cooper, L. (1963) Location-Allocation Problems. *Operations Research*, 11(3), 331–343.
- Corley and Chang, H. (1974) Finding the N Most Vital Nodes in a Flow Network. *Management Science*, 21(3), 362–4.
- Cormican, K., Morton, D. P., and Wood, R. K. (1998) Stochastic Network Interdiction. *Operations Research*, 46(2), 184–197.
- Costa, L. D. (2004) Reinforcing the Resilience of Complex Networks. *Phys. Rev. E*, 69(6).
- Cui, T., Ouyang, Y., and Shen, Z.-J. M. (2011) Reliable Facility Location Under the Risk of Disruptions. *Operations Research*, 58(4-Part-1), 998–1011.

- Cunningham, W. H. (1985) Optimal Attack and Reinforcement of a Network. *Journal of the Association for Computing Machinery*, 32(3), 549–61.
- Daganzo, C. F. (1984a) The Distance Traveled to Visit N Points With a Maximum of C Stops Per Vehicle: an Analytic Model and an Application. *Transportation Science*, 18(4), 331–50.
- Daganzo, C. F. (1984b) The Length of Tours in Zones of Different Shapes. *Transportation Research Part B*, 18B(2), 135–45.
- Daganzo, C. F. and Newell, G. F. (1986) Configuration of Physical Distribution Networks. *Networks*, 16(2), 113–22.
- Daskin, M., Hogan, K., and Reville, C. (1988) Integration of Multiple, Excess, Backup, and Expected Covering Models. *Environment And Planning B*, 15(1), 15–35.
- Daskin, M. and Stern, E. (1981) A hierarchical objective set covering model for emergency medical-service vehicle deployment. *Transportation Science*, 15(2).
- Daskin, M. S. (1982) Application of an Expected Covering Model to Emergency Medical Service System Design. *Decision Sciences*, 13(3), 416–39.
- Daskin, M. S. (1983) A Maximum Expected Covering Location Model: Formulation, Properties and Heuristic Solution. *Transportation Science*, 17(1), 48–70.
- Daskin, M. S. (1995) *Network and Discrete Location: Models, Algorithms, and Applications*. John Wiley and Sons, Hoboken, New Jersey.
- Daskin, M. S., Coullard, C. R., and Shen, Z. J. M. (2002) An Inventory-Location Model: Formulation, Solution Algorithm and Computational Results. *Annals of Operations Research*, 110(1), 83–106.
- Dighe, N., Zhuang, J., and Bier, V. M. (2009) Secrecy in Defensive Allocations as a Strategy for Achieving More Cost-effective Attacker Deterrence. *International Journal of Performability Engineering*, 5(1), 31–43.
- Dong, L., Xing-hua, K., and Xiang-tao, Y. (2009) A Model for Supply Chain Critical Facility Protection Planning Based on Time Satisfaction. In *Proceedings of the 2009 Second International Conference on Intelligent Computation Technology and Automation*. IEEE Computer Society, volume 3, pp. 903–906.
- Drezner, Z. (1987) Heuristic Solution Methods For Two Location Problems With Unreliable Facilities. *Journal of The Operational Research Society*, 38(6), 509–514.
- Eiselt, H. A., Gendreau, M., and Laporte, G. (1996) Optimal Location of Facilities On a Network With an Unreliable Node Or Link. *Information Processing Letters*, 58(2), 71–74.
- Elloumi, S., Labbé, M., and Pochet, Y. (2004) A New Formulation and Resolution Method for the P-Center Problem. *INFORMS Journal on Computing*, 16(1), 84–94.

- Entous, A., Barrett, D., and Gorman, S. (2011) Al Qaeda sought to target U.S. train network. *Wall Street Journal*.
- Erdos, P. and Renyi, A. (1959) On Random Graphs. *Publicationes Mathematicae Debrecen*, 6(156), 290–297.
- Fan, Y. and Liu, C. (2010) Solving Stochastic Transportation Network Protection Problems Using the Progressive Hedging-based Method. *Networks and Spatial Economics*, 10(2), 193–208.
- Francis, R. L., McGinnis, L. F., and White, J. A. (1992) *Facility layout and location: an analytical approach*. Prentice Hall, Englewood Cliffs, NJ.
- Fulkerson, D. R. and Harding, G. C. (1977) Maximizing Minimum Source-Sink Path Subject To A Budget Constraint. *Mathematical Programming*, 13(1), 116–118.
- Gade, D. and Pohl, E. A. (2009) Sample Average Approximation Applied to the Capacitated-Facilities Location Problem with Unreliable Facilities. *Journal of Risk and Reliability*, 223(4), 259–269.
- Golany, B., Kaplan, E. H., Marmur, A., and Rothblum, U. G. (2009) Nature Plays With Dice—Terrorists Do Not: Allocating Resources to Counter Strategic Versus Probabilistic Risks. *European Journal of Operational Research*, 192(1), 198–208.
- Grigg, C., Wong, P., Albrecht, P., Allan, R., Bhavaraju, M., Billinton, R., Chen, Q., Fong, C., Haddad, S., Kuruganty, S., and Others (1999) The IEEE reliability test system-1996: a report prepared by the reliability test system task force of the application of probability methods subcommittee. *IEEE Transactions on Power Systems*, 14(3), 1010–1020.
- Grotschel, M., Monma, C. L., and Stoer, M. (1995) Design of Survivable Networks. In *Network Models*, Ball, M. O., Magnanti, T. L., Monma, C. L., and Nemhauser, G. L., eds., Elsevier North-Holland, The Netherlands, volume 7 of *Handbooks in Operations Research and Management Science*, pp. 617–672.
- Grubestic, T. H. and Matisziw, T. C. (2008) Prospects for Assessing and Managing Vulnerable Infrastructures: Policy and Practice. *Growth and Change*, 39(4), 543–547.
- Grubestic, T. H., Matisziw, T. C., Murray, A. T., and Snediker, D. (2008) Comparative Approaches for Assessing Network Vulnerability. *International Regional Science Review*, 31(1), 88–112.
- Guha, S., Meyerson, A., and Munagala, K. (2001) Improved algorithms for fault tolerant facility location. In *Proceedings of the Annual ACM-SIAM Symposium on Discrete Algorithms*. Society for Industrial and Applied Mathematics, Society for Industrial and Applied Mathematics, Philadelphia, PA, USA, pp. 636–641.
- Guha, S., Meyerson, A., and Munagala, K. (2003) A constant factor approximation algorithm for the fault-tolerant facility location problem. *Journal of Algorithms*, 48(2), 429–440.
- Hakimi, S. L. (1964) Optimum Locations of Switching Centers and the Absolute Centers and Medians of a Graph. *Operations Research*, 12(3), 450–459.

- Hakimi, S. L. (1965) Optimum Distribution of Switching Centers in a Communication Network and Some Related Graph Theoretic Problems. *Operations Research*, 13(3), 462–475.
- Hausken, K. (2002) Probabilistic Risk Analysis and Game Theory. *Risk Analysis*, 22(1), 17–27.
- Hausken, K. (2007) Protecting Infrastructures from Strategic Attackers. In *Proceedings of the 2007 European Safety and Reliability Conference*. Stavanger, Norway, volume 1, pp. 881–887.
- Hausken, K. (2008a) Strategic Defense and Attack for Series and Parallel Reliability Systems. *European Journal of Operational Research*, 186(2), 856–881.
- Hausken, K. (2008b) Strategic Defense and Attack for Reliability Systems. *Reliability Engineering and System Safety*, 93(11), 1740–1750.
- Hausken, K., Bier, V., and Zhuang, J. (2009) Defending Against Terrorism, Natural Disaster, and All Hazards. In *Game Theoretic Risk Analysis of Security Threats*, Bier, V. and Azaiez, eds., Springer, New York, chapter 4, pp. 65–97.
- Hausken, K. and Levitin, G. (2009a) Protection vs. False Targets in Series Systems. *Reliability Engineering and System Safety*, 94(5), 973–981.
- Healy, A. and Malhotra, N. (2009) Myopic Voters and Natural Disaster Policy. *American Political Science Review*, 103(03), 387–406.
- Held, H., Hemmecke, R., and Woodruff, D. L. (2005) A Decomposition Algorithm Applied to Planning the Interdiction of Stochastic Networks. *Naval Research Logistics*, 52(4), 321–328.
- Held, H. and Woodruff, D. L. (2005) Heuristics for Multi-stage Interdiction of Stochastic Networks. *Journal of Heuristics*, 11(5-6), 483–500.
- Hemmecke, R., Schultz, R., and Woodruff, D. L. (2003) Interdicting Stochastic Networks With Binary Interdiction Effort. In *Network Interdiction and Stochastic Integer Programming*, Woodruff, D. L., ed., Kluwer, Boston, volume 22, pp. 69–84.
- Hochbaum, D. S. and Shmoys, D. B. (1986) A Unified Approach to Approximation Algorithms for Bottleneck Problems. *Journal of the ACM*, 33(3), 533–550.
- Holmgren, A. J., Jenelius, E., and Westin, J. (2007) Evaluating Strategies for Defending Electric Power Networks Against Antagonistic Attacks. *IEEE Transactions on Power Systems*, 22(1), 76–84.
- Horst, R., Pardalos, P. M., and Van Thoai, N. (2000) *Introduction to Global Optimization*, volume 48. Kluwer, The Netherlands, 2 edition.
- Israeli, E. and Wood, R. K. (2002) Shortest-Path Network Interdiction. *Networks*, 40(2), 97–111.
- Janjarassuk, U. and Linderoth, J. (2008) Reformulation and Sampling to Solve a Stochastic Network Interdiction Problem. *Networks*, 52(3), 120–32.

- Jenelius, E., Westin, J., and Holmgren (2010) Critical Infrastructure Protection Under Imperfect Attacker Perception. *International Journal of Critical Infrastructure Protection*, 3(1), 16–26.
- Jeon, H. M. (2008) *Location-inventory models with supply disruptions*. Ph.D. thesis, Lehigh University.
- Kerivin, H. and Mahjoub, A. R. (2005) Design of Survivable Networks: A survey. *Networks*, 46(1), 1–21.
- Khuller, S., Pless, R., and Sussmann, Y. J. (2000) Fault Tolerant K-Center Problems. *Theoretical Computer Science*, 242(1-2), 237–245.
- Kim, C. (2012) Toyota: supply chain will be ready by autumn for next big quake. *Reuters*.
- Kim, L. and Levitov, M. (2010) Russia heat wave may kill 15,000, shave \$15 billion of GDP. <http://www.bloomberg.com/news/2010-08-10/russia-may-lose-15-000-lives-15-billion-of-economic-output-in-heat-wave.html>.
- Knabb, R. D., Rhome, J. R., and Brown, D. P. (2005) Tropical Cyclone Report: Hurricane Katrina. Technical report, National Hurricane Center.
- Krumke, S. O. (1995) On a Generalization of the P-Center Problem. *Information Processing Letters*, 56(2), 67–71.
- Kuo, W., Prasad, V. R., Tillman, F. A., and Hwang, C. L. (2001) *Optimal Reliability Design: Fundamentals and Applications*. Cambridge University Press, Cambridge, UK.
- Laporte, G., Mesa, J. A., and Perea, F. (2010) A Game Theoretic Framework for the Robust Railway Transit Network Design Problem. *Transportation Research Part B: Methodological*, 44(4), 447–459.
- Lazoff, D. M. and Stephens, A. B. (1997) Fault-tolerant Replication in Networks With Asynchronous Communication Link Failures. In *Proceedings of the IEEE International Performance, Computing, and Communications Conference*. pp. 131–136.
- Lee, S.-D. (2001) On Solving Unreliable Planar Location Problems. *Computers and Operations Research*, 28(4), 329–44.
- Levitin, G. and Hausken, K. (2008) Protection vs. Redundancy in Homogeneous Parallel Systems. *Reliability Engineering and System Safety*, 93(10), 1444–1451.
- Levitin, G. and Hausken, K. (2009) False Targets Efficiency in Defense Strategy. *European Journal of Operational Research*, 194(1), 155–162.
- Levitin, G. and Hausken, K. (2009a) Meeting a Demand vs. Enhancing Protections in Homogeneous Parallel Systems. *Reliability Engineering and System Safety*, 94(11), 1711–1717.
- Levitin, G. and Hausken, K. (2009b) False Targets vs. Redundancy in Homogeneous Parallel Systems. *Reliability Engineering and System Safety*, 94(2), 588–595.

- Li, X. and Ouyang, Y. (2009) A continuum approximation approach to reliable facility location design under correlated probabilistic disruptions. *Transportation Research Part B*, 44(4), 535–548.
- Liberatore, F., Scaparra, M. P., and Daskin, M. (2011) Analysis of Facility Protection Strategies Against Uncertain Numbers of Attacks: The Stochastic R-Interdiction Median Problem with Fortification. *Computers and Operations Research*, 38(1), 357–366.
- Lim, M., Daskin, M., Chopra, S., and Bassamboo, A. (2010a) A Facility Reliability Problem: Formulation, Properties, and Algorithm. *Naval Research Logistics*, 57(1), 58–70.
- Lim, M., Daskin, M., Chopra, S., and Bassamboo, A. (2010b) Solving Interdicator-Defender Bilevel Optimization: A Two Population Genetic Algorithm Approach. Technical report, University of Illinois.
- Liu, C., Fan, Y., and Ordóñez, F. (2009) A Two-Stage Stochastic Programming Model for Transportation Network Protection. *Computers and Operations Research*, 36(5), 1582–90.
- Lorena, L. A. N. and Senne, E. L. F. (2004) A Column Generation Approach to Capacitated P-Median Problems. *Computers and Operations Research*, 31(6), 863–876.
- Losada, C., Scaparra, M. P., Church, R., and Daskin, M. S. (2010) The Multiple Resource Probabilistic Interdiction Median Problem. Technical report, University of Kent.
- Ma, Y. and Wu, H. (2006) Definitions and curve fitting of time satisfaction functions in facility location problems. In *Proceedings of the 2006 International Conference on Management Science and Engineering*. Piscataway, NJ, USA, volume 1, pp. 429–33.
- Marin, Mesa, J., and Perea, F. (2009) Integrating Robust Railway Network Design and Line Planning under Failures. In *Robust and Online Large-Scale Optimization*, Ahuja, R. K., ed., Springer, pp. 273–292.
- Medal, H., Rainwater, C., Pohl, E., and Rossetti, M. (2011a) On the R-All-Neighbor P-Center Problem. *Working paper*.
- Medal, H., Sharp, S. J., Pohl, E., Mason, S. J., and Rainwater, C. (2011b) Models for Reducing the Risk of Critical Networked Infrastructures. *International Journal of Risk Assessment and Management*, 15(2/3), 99–127.
- Medal, H. R., Pohl, E. A., and Rossetti, M. D. (2011c) An Integrated Model for Facility Location and Hardening. *Working paper*.
- Melachrinoudis, E. and Helander, M. E. (1996) A single facility location problem on a tree with unreliable edges. *Networks*, 27(3), 219–237.
- Mervin, J. (2011) Counting the cost of the BP disaster one year on. *BBC News*.
- Mladenović, N., Labbé, M., and Hansen, P. (2003) Solving the p-Center problem with Tabu Search and Variable Neighborhood Search. *Networks*, 42(1), 48–64.

- Morehead, R. and Noore, A. (2007) Novel hybrid mitigation strategy for improving the resiliency of hierarchical networks subjected to attacks. *Physica A: Statistical Mechanics and its Applications*, 378(2), 603–612.
- Morton, D. P. (2011) Stochastic Network Interdiction. In *Wiley Encyclopedia of Operations Research and Management Science*, Cochran, J. J., Cox, L. A., Keskinocak, P., Kharoufeh, J. P., and Smith, J. C., eds., Wiley, Hoboken, NJ.
- Morton, D. P., Pan, F., and Saeger, K. J. (2007) Models for Nuclear Smuggling Interdiction. *IIE Transactions*, 38(1), 3–14.
- MSNBC (2010) U.S.: haiti aid bottleneck is easing up. *msnbc.com*.
- Murray, A. T., Matisziw, T. C., and Grubestic, T. H. (2008) A Methodological Overview of Network Vulnerability Analysis. *Growth and Change*, 39(4), 573–592.
- Nel, L. D. and Colbourn, C. J. (1990) Combining monte carlo estimates and bounds for network reliability. *Networks*, 20(3), 277–298.
- News, B. (2005) Tsunami fishermen struggle to cope. *BBC News*.
- News, R. R. R. C. and Information (2005) Rails Cause Utility Fuel Shortages, Electricity Rate Hikes.
- Oceans Beyond Piracy (2012) The Economic Cost of Somali Piracy 2011. Technical report, Oceans Beyond Piracy.
- O’Hanley, J. R., Church, R., and Gilless, J. K. (2007a) Locating and Protecting Critical Reserve Sites to Minimize Expected and Worst-Case Losses. *Biological Conservation*, 134(1), 130–141.
- O’Hanley, J. R. and Church, R. L. (2011) Designing Robust Coverage Networks to Hedge Against Worst-Case Facility Losses. *European Journal of Operational Research*, 209(1), 23–36.
- O’Hanley, J. R., Church, R. L., and Gilless (2007b) The importance of *In Situ* site loss in nature reserve selection: Balancing notions of complementarity and robustness. *Biological Conservation*, 2(1), 170–180.
- Pan, F., Charlton, W., and Morton, D. P. (2003) Interdicting Smuggled Nuclear Material. In *Network Interdiction and Stochastic Integer Programming*, Woodruff, D. L., ed., Kluwer Academic Publishers, Boston, pp. 1–20.
- Pan, F. and Morton, D. P. (2008) Minimizing a stochastic maximum-reliability path. *Networks*, 52(3), 111–119.
- Paul, G., Tanizawa, T., Havlin, S., and Stanley, H. E. (2004) Optimization of robustness of complex networks. *European Physical Journal B*, 38(2), 187–191.
- Peeta, S., Salman, F. S., Gunneç, D., and Viswanath, K. (2010) Pre-Disaster Investment Decisions for Strengthening a Highway Network. *Computers and Operations Research*, 37(10), 1708–1719.

- Peng, P., Snyder, L. V., Liu, Z., and Lim, A. (2011) Reliable Logistics Networks Design with Facility Disruptions. *Transportation Research-Part B*, 45(8), 1190–1211.
- Peng, R., Levitin, G., Xie, M., and Ng, S. H. (2010) Defending simple series and parallel systems with imperfect false targets. *Reliability Engineering and System Safety*, 95(6), 679–688.
- Powell, R. (2007a) Allocating defensive resources with private information about vulnerability. *American Political Science Review*, 101(04), 799–809.
- Powell, R. (2007b) Defending against terrorist attacks with limited resources. *American Political Science Review*, 101(03), 527–541.
- Prasad, T. D. and Park, N. (2004) Multiobjective genetic algorithms for design of water distribution networks. *Journal of Water Resources Planning and Management*, 130(1), 73–82.
- Qi, L., Shen, Z. J. M., and Snyder, L. V. (2010) The Effect of Supply Disruptions on Supply Chain Design Decisions. *Transportation Science*, 44(2), 274–289.
- Qiao, J., Jeong, D., Lawley, M., Richard, J. P. P., Abraham, D. M., and Yih, Y. (2007) Allocating Security Resources to a Water Supply Network. *IIE Transactions*, 39(1), 95–109.
- Ramirez-Marquez, J. E., Rocco, C. M., and Levitin, G. (2009) Optimal Protection of General Source-Sink Networks via Evolutionary Techniques. *Reliability Engineering and System Safety*, 94(10), 1676–1684.
- Ramirez-Marquez, J. E. and Rocco, C. M. S. (2009) Stochastic Network Interdiction Optimization via Capacitated Network Reliability Modeling and Probabilistic Solution Discovery. *Reliability Engineering and System Safety*, 94(5), 913–21.
- Reinelt, G. (1991) TSPLIB- a traveling salesman problem library. *ORSA Journal of Computing*, 3(4), 376–384.
- Rhoads, C. (2010) Quake sets back haiti's efforts to improve telecommunications. *Wall Street Journal*.
- Rocco, C. M., Ramirez-Marquez, J. E., and Salazar, D. E. (2010) Bi- and Tri-Objective Optimization in the Deterministic Network Interdiction Problem. *Reliability Engineering and System Safety*, 95(8), 887–896.
- Ryoo, H.-S. and Sahinidis, N. V. (2003) Global optimization of multiplicative programs. *Journal of Global Optimization*, 26(4), 387–418.
- San Martin, P. A. (2007) *Tri-level optimization models to defend critical infrastructure*. Master's thesis, Naval Postgraduate School.
- Santiviáñez, J. and Melachrinoudis, E. (2008) Location of a reliable center on a tree network. *Operational Research*, 7(3), 419–445.

- Santiviáñez, J., Melachrinoudis, E., and Helander, M. E. (2009) Network location of a reliable center using the most reliable route policy. *Computers and Operations Research*, 36(5), 1437–1460.
- Santos, J. R. (2006) Inoperability input-output modeling of disruptions to interdependent economic systems. *Systems Engineering*, 9(1), 20–34.
- Scaparra, M. P. (2006) Optimal resource allocation for the facility protection in median systems. Working paper, University of Kent.
- Scaparra, M. P. and Cappanera, P. (2005) Optimizing Security Investments in Transportation and Telecommunication Networks. Technical report, University of Kent, Canterbury, UK.
- Scaparra, M. P. and Church, R. L. (2008a) A Bilevel Mixed-Integer Program for Critical Infrastructure Protection Planning. *Computers and Operations Research*, 35(6), 1905–1923.
- Scaparra, M. P. and Church, R. L. (2008b) An Exact Solution Approach for the Interdiction Median Problem with Fortification. *European Journal of Operational Research*, 189(1), 76–92.
- Scaparra, M. P. and Church, R. L. (2010) Protecting Supply Systems to Mitigate Potential Disaster: A Model to Fortify Capacitated Facilities. Technical Report 209, University of Kent, Canterbury, UK.
- Schavland, J., Chan, Y., and Raines, R. A. (2009) Information Security: Designing a Stochastic-network for Throughput and Reliability. *Naval Research Logistics*, 56(7), 625–641.
- Shen, S., Smith, J. C., and Ahmed, S. (2010) Expectation and Chance-Constrained Models and Algorithms for Insuring Critical Paths. *Management Science*, 56(10), 1794–1814.
- Shen, Z. J. M., Coullard, C., and Daskin, M. S. (2003) A joint location-inventory model. *Transportation Science*, 37(1), 40–55.
- Shen, Z. J. M., Zhan, R., and Zhang, J. (2011) The Reliable Facility Location Problem: Formulations, Heuristics, and Approximation Algorithms. *INFORMS Journal on Computing*, 23(3), 470–482.
- Short, M. B., Brantingham, P. J., Bertozzi, A. L., and Tita, G. E. (2010) Dissipation and displacement of hotspots in reaction-diffusion models of crime. *Proceedings of the National Academy of Sciences*, 107(9), 3961–3965.
- Skaperdas, S. (1996) Contest success functions. *Economic Theory*, 7(2), 283–290.
- Smith, Lim, C., and Sudargho, F. (2007) Survivable Network Design Under Optimal and Heuristic Interdiction Scenarios. *Journal of Global Optimization*, 38(2), 181–199.
- Smith, J. C. (2011) Basic Interdiction Models. In *Wiley Encyclopedia of Operations Research and Management Science*, Cochran, J., Cox, L. A., Keskinocak, P., Kharoufeh, J. P., and Smith, J. C., eds., Wiley, Hoboken, NJ.

- Smith, J. C. and Lim, C. (2008) Algorithms for Network Interdiction and Fortification Games. In *Pareto Optimality, Game Theory And Equilibria*, Chinchuluun, A., Pardalos, P. M., Migdalas, A., and Pitsoulis, L., eds., Springer New York, New York, NY, volume 17 of *Springer Optimization and Its Applications*, chapter 24, pp. 609–644.
- Snyder, L. V. (2006) Facility location under uncertainty: a review. *IIE Transactions*, 38(7), 537–554.
- Snyder, L. V., Atan, Z., Peng, P., Rong, Y., Schmitt, A. J., and Sinsoysalk, B. (2010) OR/MS models for supply chain disruptions: A review. Technical report, Lehigh University.
- Snyder, L. V. and Daskin, M. S. (2005) Reliability Models for Facility Location: The Expected Failure Cost Case. *Transportation Science*, 39(3), 400–416.
- Snyder, L. V. and Daskin, M. S. (2007) Models for Reliable Supply Chain Network Design. In *Critical Infrastructure: Reliability and Vulnerability*, Murray, A. T. and Grubescic, T. H., eds., Springer-Verlag, Berlin, Germany, chapter 13, pp. 257–289.
- Snyder, L. V., Scaparra, M. P., Daskin, M. S., and Church, R. L. (2006) Planning for Disruptions in Supply Chain Networks. In *TutORials in Operations Research: Models, Methods, and Applications for Innovative Decision Making*, Johnson, M. P., Norman, B., and Secomandi, N., eds., INFORMS, Baltimore, MD, TutORials in Operations Research, chapter 9, pp. 234–257.
- Sullivan, J. L., Aultman-Hall, L., and Novak, D. C. (2009) A review of current practice in network disruption analysis and an assessment of the ability to account for isolating links in transportation networks. *Transportation Letters*, 1(4), 271–280.
- Swain, R. W. (1971) *A Decomposition Algorithm for a Class of Facility Location Problems*. Ph.D. thesis, Cornell University.
- Tabuchi, H. (2011) Toyota says no full production until year’s end. *The New York Times*.
- Tanizawa, T., Paul, G., Cohen, R., Havlin, S., and Stanley, H. E. (2005) Optimization of network robustness to waves of targeted and random attacks. *Physical Review E*, 71(4), 047101+.
- The Department of Homeland Security Risk Steering Committee (2008) DHS Risk Lexicon. Technical report, Department of Homeland Security, Washington, D.C.
- Toregas, C. and ReVelle, C. (1972) Optimal location under time or distance constraints. *Papers in Regional Science*, 28(1), 131–143.
- Toregas, C., Wwain, R., Revelle, C., and Bergman, L. (1971) Location of Emergency Service Facilities. *Operations Research*, 19(6), 1363–1373.
- Ulfarsson, G. and Unger, E. (2011) Impacts and Responses of Icelandic Aviation to the 2010 Eyjafjallajökull Volcanic Eruption. *Transportation Research Record: Journal of the Transportation Research Board*, 2214(2011), 144–151.
- United States-Canada Power System Outage Task Force (2004) Final Report on the August 14th Blackout in the United States and Canada.

- Van Slyke, R. (1982) Redundant set covering in telecommunications networks. In *Proceedings of the 1982 IEEE International Large Scale Systems Symposium*. New York, NY, USA, pp. 217–222.
- Veerasamy, J., Venkatesan, S., and Shah, J. C. (1999) Spare capacity assignment in telecom networks using path restoration and further improvement using traffic splitting. *Journal of Systems and Software*, 47(1), 27–33.
- Von Stackelberg, H. and Peacock, A. T. (1952) *The Theory of the Market Economy*. Oxford University Press, Oxford, UK.
- Wallace, S. W. (1987a) A Piecewise Linear Upper Bound on the Network Recourse Function. *Mathematical Programming*, 38(2), 133–146.
- Wallace, S. W. (1987b) Investing in Arcs in a Network to Maximize the Expected Max Flow. *Networks*, 17(1), 87–103.
- Washington Post (2010) Haiti raises earthquake toll to 230,000. *The Washington Post*.
- Watts, D. J. and Strogatz, S. H. (1998) Collective dynamics of ‘small-world’ networks. *Nature*, 393(6684), 440–442.
- Wollmer, R. (1964) Removing Arcs from a Network. *Operations Research*, 12(6), 934–940.
- Wood, R. K. (1993) Deterministic Network Interdiction. *Mathematical and Computer Modelling*, 17(2), 1–18.
- Wood, R. K. (2011) Bilevel Network Interdiction Models: Formulations and Solutions. In *Wiley Encyclopedia of Operations Research and Management Science*, Cochran, J. J., Cox, L. A., Keskinocak, P., Kharoufeh, J. P., and Smith, J. C., eds., John Wiley & Sons, Inc., Hoboken, NJ.
- Xu, C. and Goulter, I. C. (1999) Reliability-based optimal design of water distribution networks. *Journal of Water Resources Planning and Management*, 125(6), 352–362.
- Xue, G. (1997) Linear time algorithms for computing the most reliable source on an unreliable tree network. *Networks*, 30(1), 37–45.
- Yao, Y., Edmunds, T., Papageorgiou, D., and Alvarez, R. (2007) Trilevel Optimization in Power Network Defense. *IEEE Transactions On Systems Man And Cybernetics Part C*, 37(4), 712–718.
- Zhan, R. L. (2007) *Models and Algorithms for Reliable Facility Location Problems and System Reliability Optimization*. Ph.D. thesis, University Of Florida.
- Zhao, J. and Xu, K. (2009) Enhancing the robustness of scale-free networks. *Journal of Physics A: Mathematical and Theoretical*, 42(19), 195003+.
- Zhuang, J. and Bier, V. (2011) Secrecy and Deception at Equilibrium, with Applications to Anti-Terrorism Resource Allocation. *Defence and Peace Economics*, 22(1), 43–61.

- Zhuang, J. and Bier, V. M. (2007) Balancing terrorism and natural disasters-defensive strategy with endogenous attacker effort. *Operations Research*, 55(5), 976–991.
- Zhuang, J. and Bier, V. M. (2008) Katrina vs. 9/11 How should we optimally protect against both? In *Post-Katrina: Economics, Social Aspects, and Risk*, Richardson, H., Gordon, P., and Moore, J., eds., Edward Elgar Publishing, chapter 4, pp. 71–83.
- Zhuang, J. and Bier, V. M. (2010) Reasons for Secrecy and Deception in Homeland-Security Resource Allocation. *Risk Analysis*, 30(12), 1737–1743.

