MACK-BLACKWELL
Rural Transportation Center

NTSCOE NATIONAL TRANSPORTATION SECURITY CENTER OF EXCELLENCE

University of Arkansas
4190 Bell Engineering Center
Fayetteville, AR 72701
479.575.6026 – Office
479.575.7168 - Fax

MBTC DHS 1101 - Designing Resilient and Sustainable Supply Networks

Ed Pohl, Ph.D.
Scott Mason, Ph.D.
Chase Rainwater, Ph.D.
Ridvan Gedik
Hugh Medal
Jennifer Carter
Nick Martin
Chen Wang
Brittni King

March 2012

Prepared for
Mack-Blackwell Rural Transportation Center
National Transportation Security Center of Excellence
University of Arkansas

ACKNOWLEDGEMENT
This material is based upon work supported by the U.S. Department of Homeland Security under Grant Award Number 2008-ST-061-TS003.

Abstract

This report documents the outcome of project MBTC 1101: Designing Resilient and Sustainable Supply Networks. This project produced three main deliverables: a survey of the literature on networks subject to disruptions, a series of papers on locating facilities that are vulnerable to disruptions, and a case study on disruptions in the coal supply network.

The survey paper, presented in Chapter 2, reviews the literature on designing new networks that are subject to disruptions and reducing the risk of existing networks that are subject to disruptions. In addition to reviewing the literature on supply chains subject to disruptions, it also address other networks such as telecommunications networks. This review organizes a wide-ranging body of literature into a classification that should be useful to future researchers.

The series of papers, described in Chapter 3, focus on designing distribution networks that are subject to disruptions. Models are developed to mitigate against the worst case disruption scenario and system performance is measured as the maximum distance from a demand point to its closest facility after failures. First, a model is developed for locating facilities. Second, a model is developed that integrates location and hardening decisions. The models are used to trade off between multiple objectives as well as generate other managerial insights.

Finally, in Chapter 4, a case study is presented that examines vulnerabilities in the rail infrastructure used to transport coal in the United States. A model is presented that identifies the most critical components of the rail infrastructure.

Contents

Chapter 1

Introduction

This report describes the findings from project DHS 1101 titled "Designing Resilient and Sustainable Supply Chains". The objective of this project was to develop new methods for designing supply networks that are resilient and sustainable under the presence of disruptions. In this project we focused on the strategic decision-making level, where long-term decisions are made such as designing new networks and protecting existing networks. We considered two types of disruptions: random (e.g., natural disasters) and worst-case (e.g., disruptions caused by a terrorist attack). This report discusses the three main deliverables of our project: 1) a survey paper on reducing the risk of disruptions in networks, 2) a series of papers on locating and protecting facilities subject to disruptions, and 3) a case study to identify vulnerabilities in the coal supply chain.

Modern supply chains have evolved into complex systems because of globalization and decentralization. As with many complex systems, there are risks involved in supply chains. Of primary concerns are the risks associated with large-scale disruptions due to natural disasters, terrorist attacks, political instability, transportation and network failures etc. These risks can have a direct or an indirect impact on supply chain continuity and are important because they can dramatically reduce the effectiveness of the supply chain and result in significant economic loss and more importantly losses in human life. It is essential for organizations to assess these risks and develop strategies to mitigate them. The risk profile of a supply chain depends on the configuration of its primary components such as suppliers, warehouses, service centers, staging areas, ports of debarkation, transportation modes etc. The location, mode of transportation and choice of associates constituting these components is a strategic decision and hence there are significant costs associated with such decisions. Strategic location, transportation and selection decisions should make the supply chains robust, reliable and resilient and at the same time should not compromise on an organization's ability to meet its mission requirements. Poor decisions with regard to the location, transportation and selection of associates for these components can make the supply chain vulnerable to disruptions. For many organizations, these strategic decisions were made without consideration of the risk of disruption. The objective of this research is to develop models for resilient and reliable supply chain network design using tools and techniques from reliability and optimization. Previous research in supply chain management has focused on dealing with demand uncertainties and building "lean" supply chains. While these issues are important, the issue of large scale disruptions effecting supply chains cannot be overlooked. Events in the recent past such as the 9/11 terrorist attacks, hurricane Katrina, the 2002 West Coast port closure, Operation Iraqi Freedom, and Operation Enduring Freedom have brought to light the vulnerabilities in supply chains and a need for new models in supply chain design. This research aims to bridge this gap and will yield new supply chain network strategies that incorporate the risks associated with disruption.

Events in the last decade demonstrate the effect that disruptions can have on supply networks. One example 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. Another example is 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 (Rail Report: Rail Customer News and Information, 2005). Finally, in 2010, the eruption of a volcano in Iceland halted flights throughout Northeastern Europe for several days (Ulfarsson and Unger, 2011).

There are several potential reasons why these disruptions had such a large impact. First, networks are often designed to minimize cost and maximize efficiency. For example, many of the classic network optimization problems (e.g., shortest path, maximum flow, minimum cost spanning tree) have efficiency or cost objectives. When an efficiency or cost objective is used networks are often designed to be used at or near their maximum capacity, leaving them without the excess capacity needed to withstand 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 Barabási and Albert (1999) and Albert *et al.* (2000). Barabási 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.

The **goals** of this project were: (i) to contribute to the theoretical foundations of resiliency and sustainability for a complex supply chain system; (ii) to bring together and build upon the expertise and advances of the research team in the areas of reliability and maintainability, systems engineering, optimization and risk analysis to develop the science associated with the resiliency and sustainability of complex interdependent supply chain systems; (iii) to drive a paradigm shift from the traditional static deterministic approaches to the analysis of these systems to a dynamic interdependent view of the complex supply chain.

Our objective was to develop the mathematical models of resiliency and sustainability that support analysis and decision making in complex supply chains. We accomplished this through the following tasks: 1.) Identify a specific multi-modal supply chain network in the United States and identify the key strategic elements in the network. 2.) Develop an analytical model of the interdependent supply chain network that can incorporate the vulnerabilities and associated risks associated with key strategic elements. 3.) Study and analyze the problem of capacity degradation in the supply chain due to large scale disruptions and to extend the above ideas to design a reliable and sustainable robust supply chain transportation network.

This research produced three main deliverables, which are described in the remainder of this report. Section 2 contains a survey of the literature on mitigating against disruptions in networks. This survey classifies this body of literature and suggests opportunities for future work. Section 3 summarizes a series of papers on locating facilities that are vulnerable to failures. These papers present mathematical models for locating and hardening distribution networks that are subject to failures. In addition, we discuss some of the insights gained from this line of research. Section 4 discusses a case study on identifying vulnerabilities in the coal supply chain. A model was developed that identifies the set of railyards whose destruction results in the greatest increase in travel cost. This model is exercised using real data from the subituminous coal supply chain in the United States. Finally, Section 5 contains a summary of the findings of this report as well as opportunities for future work.

This project made several contributions to the academic literature on supply network disruptions. First, our survey paper is the first comprehensive review of reducing the risk of networks subject to disruptions.

The classification provided in the survey should help future researchers understand a broad area of research. Second, this report describes a series of papers that contributes to the literature on locating and protecting facilities subject to disruptions. In particular, these papers were the first to study an exact optimization procedure for locating facilities to mitigate against the worst-case disruption with the maximum distance objective. Further, these papers integrated location and hardening, two decisions that had been considered separately in the literature. Finally, the case study presented in this report is the first case study of the coal supply chain. Further, a new interdiction model, in which the second stage of the model mimics the movement of trains through a network over time, was developed for this case study.

Chapter 2

Literature Review[1]

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 consequences 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 (Rail Report: Rail Customer 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 about six days (Wikipedia, 2010). 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 Barabási and Albert (1999) and Albert *et al.* (2000). Barabási 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,

[1] This chapter is based on Medal, H., Sharp, S.J., Pohl, E., Mason, S.J., Rainwater, C. (2011) Models for networked infrastructure subject to disruptions. *International Journal of Risk Assessment and Management*, 15(2/3), 99-127.

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.* (2005) 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

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

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 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.

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

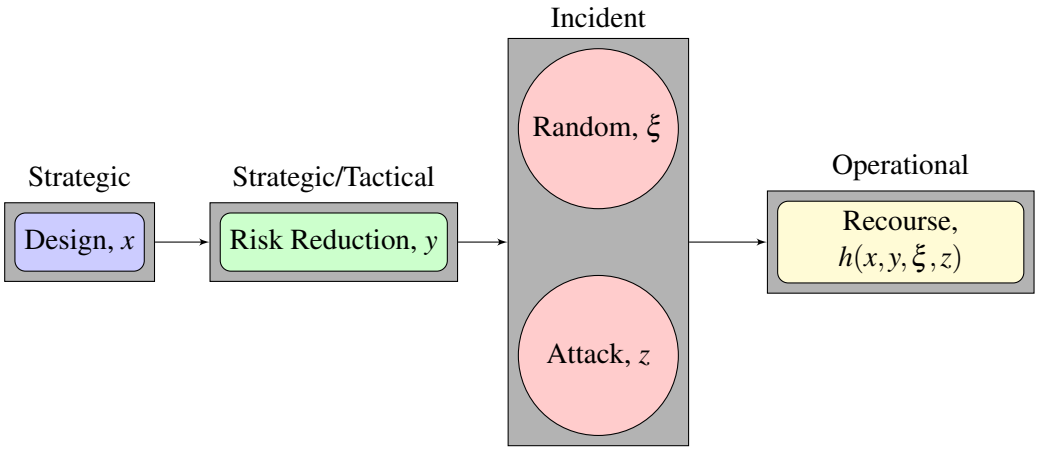


Figure 2.1: Network problems with disruptions: problem stages.

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 models 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 review has been to review as many different approaches as possible for dealing with risk from a network design and improvement perspective. All of the papers reviewed are analytical in nature and primarily deal with operations research models. A review of qualitative papers is available in Rao and Goldsby (2009).

While there is a vast body of literature on decision making under uncertainty, we only survey articles that explicitly consider disruptions. Another related area is that of strategic supply chain decisions under uncertainty, which includes facility location and network design among other areas. In fact, many of the models developed in these areas can be used for problems in this review. Snyder *et al.* (2006) provide a thorough survey of this topic.

This review deals with the body of knowledge pertaining to what to do before a potentially disruptive event occurs and not after. Thus, research areas such as emergency response and disaster relief were not considered. For more information on emergency response we recommend starting with the book by Larson and Odoni (1981). For a review of disaster relief see Altay and Green (2006) and for an general introduction to this research area see Ergun *et al.* (2010).

Another related area of research is that of economic input-output models, which focus on capturing the

inter-dependencies between economic sectors (see Santos (2006) for example). This differs from the topics that we cover in this survey in that input-output models are concerned with economic infrastructure while we are more concerned with physical infrastructure.

Papers are primarily divided into two categories: papers dealing with designing a new network considering risks to the network and papers seeking to improve an existing network's reaction to risks. Therefore, the papers reviewed in this work are more concerned with the strategic and tactical aspects of network design, rather than the operational component. This review seeks to answer the question "Where should facilities be located?" and not "How should facilities be operated?" So while the issue of facility location is discussed at great length, issues related to inventory (e.g. (Jodlbauer and Altendorfer, 2010), (Schmitt *et al.*, 2008)) and transportation (e.g. (Mehndiratta *et al.*, 2000), (Rodrigues *et al.*, 2008)) receive much less attention. For a review of that includes other topics relating to supply chain disruptions, the reader should see Snyder et al.'s 2010 survey (Snyder *et al.*, 2010) which additionally offers a more comprehensive review of supplier sourcing, inventory control, and contingent rerouting as supply chain disruption mitigation strategies.

We 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. And we are interested in studying system-wide impacts such as performance degradation rather than 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.

Our survey focuses on models that endogenously determine the redundancy levels for a system, excluding models that determine redundancy exogenously. Models of the latter type include vector assignment models (Weaver and Church, 1985), where facilities can only serve a predefined proportion of customer demand, and models that require each demand point to have a backup (Pirkul, 1989).

Finally, in this review we consider large scale disruptions as opposed to disruptions due to natural wear and tear. However, we acknowledge that there is a fine line between these two sources of disruptions. For more on the difference between large scale disruptions and disruptions due to natural wear and tear, see Sullivan *et al.* (2009).

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 Grubesic *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 (Barabási 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 Grubesic *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.

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*

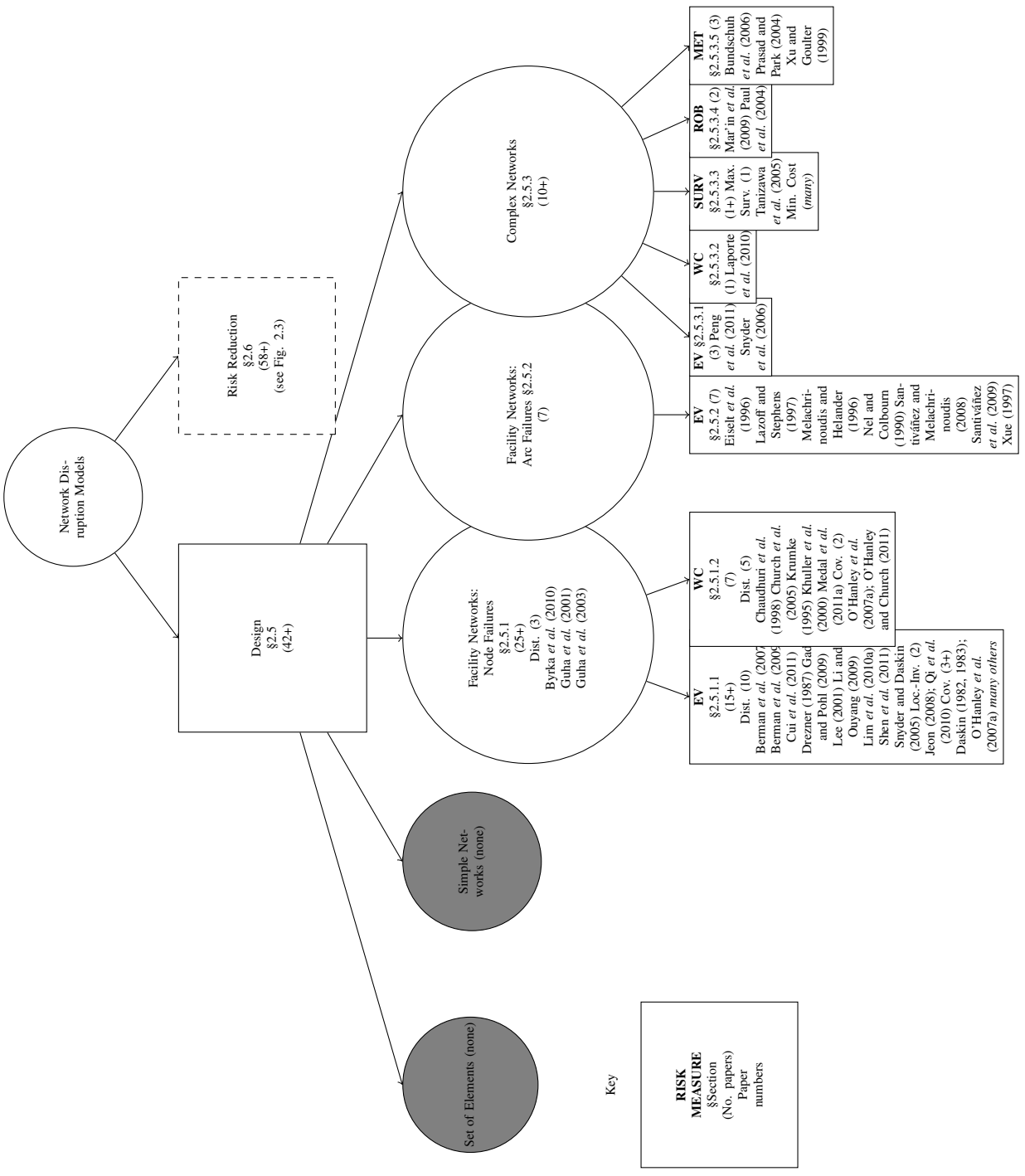


Figure 2.2: Classification diagram (left side)

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 $\lambda_{1j}\lambda_{2j} \leq \cdots \leq \lambda_{r_j j}$, 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 URPMP and URFLP, and otherwise assume that the probabilities are non-uniform.

The FTPMP and FTPMP can be used to model the URPMP 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 URPMP. 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 algorithm[3].

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 URPMP 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.

[3] A 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.

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 expected 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.

Santiváñez 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. Santiváñez *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 con-

nectivity after all single-element failures. Rather than attempting to review all of this literature here, we refer readers to the survey papers by Grotschel *et al.* (1995) and Kerivin and Mahjoub (2005).

2.5.3.4 Robustness Risk Measure

Mar'in *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.* (2000).

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, excluding papers that we didn't describe because there already exists another review paper on that specific topic. 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							
	Expected value	Worst case	Risk metric	Surviv.	Robust.	Other	Total	%
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 interdictor will never (optimally) attack a hardened element (Scaparra and Church (2008a)).

Remark 1 is rather intuitive; if the defender doesn't 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 made-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.

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.

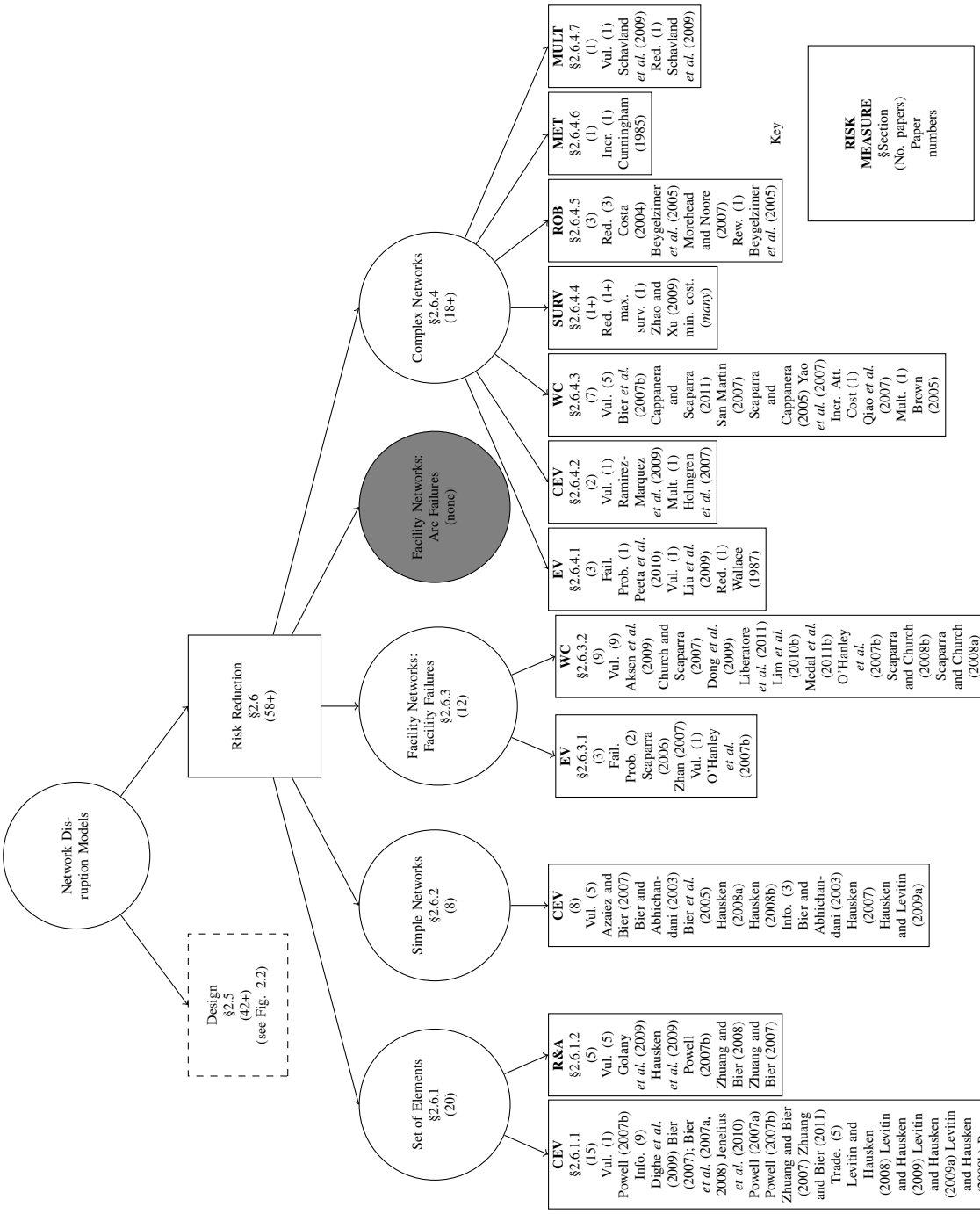


Figure 2.3: Classification diagram (right side)

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.* (2009) 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.* (2011b) 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 Bender's Decomposition, which is well-suited for nonlinear problems.

Redundancy Wallace (1987) 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 a 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 network 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 Bender's 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 Barabási 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 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 where 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.

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

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

	Primary risk reduction strategy											
	VUL	LI	ECR	FP	RED	REST	REW	INCR	INFO	MULT	TRADE	Sum	%
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%		

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

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

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 though 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.* (2011b) 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 Brittni King, for her assistance on this project.

Chapter 3

Locating and Protecting Facilities Subject to Disruptions

The decision of locating facilities is a strategic decision that 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 3.1 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 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

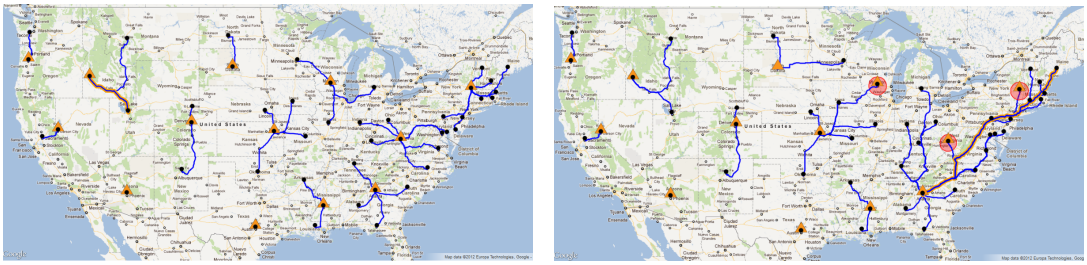
Table 3.1: Sectors with critical facilities

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

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 we use the *p*-center model to choose the locations of the facilities. Thus, we seek 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 3.1a. 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 283, which is the distance from Carson City to the facility located at Boise City.



(a) Demand point assignments without failures

(b) Demand point assignments after failures

Figure 3.1: Facilities located to optimize performance without facility failures

The solution to the *p*-center model, shown in Figure 3.1a, 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. In this solution, the radius is 294.

Figure 3.1b 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-failure radius*. Conversely, we call the maximum distance when no facilities have failed the *non-failure radius*.

In the rest of this Section, we discuss ways to reduce the post-failure radius. The results in Section 3.1 show that locating facilities in different places can reduce the post-failure maximum distance. In Section 3.2 the results show that facility hardening can also reduced the post-failure maximum distance. In particular when location decisions are considered simultaneously with hardening decisions, the solutions prescribed are significantly more resilient to facility failures than solutions prescribed by models that consider facility location and facility hardening separately.

3.1 Locating Facilities Subject to Failure[1]

Recall Figure 3.1b, which shows that the post-failure distance is 1071. Keep in mind that the facilities in Figure 3.1b were located with the objective of minimizing the non-failure radius. Thus, we see that in this case the non-failure radius is not a good replacement for the post-failure radius objective. Therefore, in this section we present a model that minimizes the post-failure radius.

3.1.1 Model

To facilitate our analysis, we developed a mathematical model for the r-all-neighbor p-center problem (RANPCP). In particular, our model prescribes how to optimally locate a set of facilities that are vulnerable to failures. The purpose of this model is to *locate a set of facilities in order to minimize the maximum consequence over all possible failure scenarios consisting of the failure of r facilities. The consequence of a failure scenario is the maximum distance from a demand point to its closest located and operating facility.*

The RANPCP has several applications. First, this model can be used to locate *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. Second, this model can be used to locate facilities in order to mitigate against the worst-case failure of r facilities.

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. 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. The location decisions can be made together or separately. In the disruption stage, the disruption causes exactly r facilities to fail. In the response stage, demand points are served by their closest located facility.

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 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 $\mathscr{I}$ be a set of potential facility locations and $\mathscr{J}$ be a set of demand points. We measure the effectiveness of a facility located at i serving the demand point located at j using a measure ϕ_{ij}. Let ϕ_{ij} be a measure of the effectiveness of a facility located at i serving the demand point located at j. This measure could represent 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. The cost of locating a facility at i is f_i. The total cost of locating facilities must be within a budget b.

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

The following variables are used in our model. 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 1 if a facility at i is located and 0 otherwise.

A MIP formulation of the RANPCP model is:

[1]This subsection is a summary of the results in Medal, H., Rainwater, C., Pohl, E., and Rossetti, M., 2011. On the R-All-Neighbor P-Center Problem. Working paper.

$$\min \quad U \tag{3.1a}$$
$$\text{s.t.} \quad U \geq \phi_{ij} W_{ij} \quad \forall i \in \mathscr{I}, j \in \mathscr{J} \tag{3.1b}$$
$$(r+1)W_{ij} \leq \sum_{i' : \phi_{i'j} \leq \phi_{ij}} X_{i'} \quad \forall j \in \mathscr{J}, i \in \mathscr{I} \tag{3.1c}$$
$$\sum_{i \in \mathscr{I}} W_{ij} = 1 \quad \forall j \in \mathscr{J} \tag{3.1d}$$
$$W_{ij} \leq X_i \quad \forall i \in \mathscr{I}, j \in \mathscr{J} \tag{3.1e}$$
$$\sum_{i \in \mathscr{I}} f_i X_i \leq b \tag{3.1f}$$
$$X_i \in \{0,1\} \quad \forall i \in \mathscr{I} \tag{3.1g}$$
$$W_{ij} \in \{0,1\} \quad \forall i \in \mathscr{I}, j \in \mathscr{J} \tag{3.1h}$$

The objective equation (3.1a), in conjunction with Constraints equation (3.1b), is to minimize the post-interdiction radius. Constraints equation (3.1c) model the requirement that i and j can only form a post-interdiction bottleneck pair if $r+1$ facilities are located that are at least as close to j than i. Constraints equation (3.1d) require that every demand point form a post-interdiction bottleneck pair with one facility. Constraints equation (3.1e), although not necessary because of the presence of Constraints equation (3.1c), tighten the LP relaxation. Constraint equation (3.1f) requires that the amount spent on location and hardening must be within a budget. Constraints equation (3.1g)–equation (3.1h) specify bounds on the variables.

3.1.2 Solution Procedure

Model (3.1) can be solved using an off-the-shelf MIP optimizer such as CPLEX. However, because of the bottleneck structure of the RANPCP 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 U^*. The auxiliary problem for the p-center problem is the set-cover problem with unitary costs. If some radius U 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 U. Let $p^*(U)$ be the optimal number of facilities needed to cover all demand points within U. If $p^*(U) \geq p$, then $U \leq U^*$, and U is a new lower bound. If $p^*(U) < p$, then $U \geq U^*$, and U is a new upper bound. Thus, a binary search can be performed over all values of U to find U^*. 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 RANPCP. The main extension is that the auxiliary problem is different. In this chapter, we solve the RANPCP using with a modified auxiliary problem.

To use a binary search algorithm for the RANPCP, the auxiliary problem must first be described. Define U as the radius for the auxiliary problem. (Note that U is now a parameter and not a variable, as it was in Model (3.1).) To evaluate whether a particular U is above or below the optimal post-interdiction radius, the multi-set-cover problem (MSCP) (Church and Gerrard, 2003) is used:

$$\text{MSCP}(U) \quad \min \quad \sum_{i \in \mathscr{I}} f_i X_i \tag{3.2a}$$

$$\text{s.t.} \quad \sum_{i:\phi_{ij} \leq U} X_i \geq r+1 \quad \forall j \in \mathscr{J} \tag{3.2b}$$

The MSCP minimizes the cost required for every demand point to have a post-interdiction assignment distance less than or equal to U. The objective equation (3.2a) is to minimize the total cost of location. Constraints equation (3.2b) require that for each demand point j, $r+1$ facilities within U of j must be located.

A binary search algorithm for the RANPCP is described in Appendix A.1.

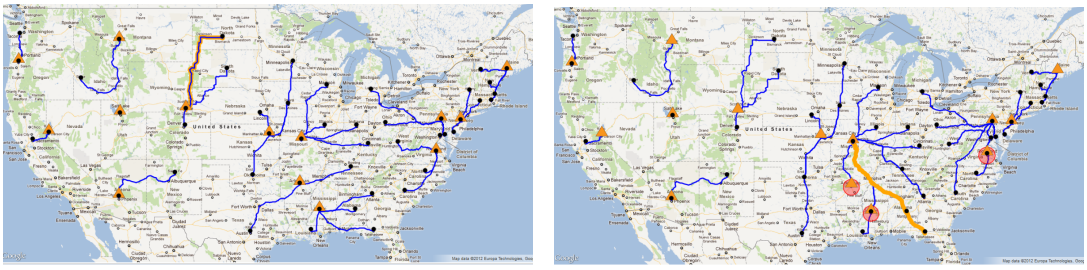
3.1.3 Example Continued

Continuing the example above, Figure 3.2 shows the solution generated by the RANPCP. The 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 3.1b.

Figure 3.2a 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 3.2b 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.



(a) Demand assignments without failures (b) Demand point assignments after failures

Figure 3.2: Facilities located to optimize post-failure radius

3.1.4 Insights

In this section we describe several insights gained from experimentation with the RANPCP. These insights involve 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 non-failure radius. The potential consequence is measured as the post-failure radius.

To generate the insights in this section, the model was tested using a variety of datasets from the facility location literature (for a complete list see Medal *et al.* (2011a)). For each dataset, several combinations of parameter settings were tested (Medal *et al.*, 2011a) and summary statistics were calculated across the combinations.

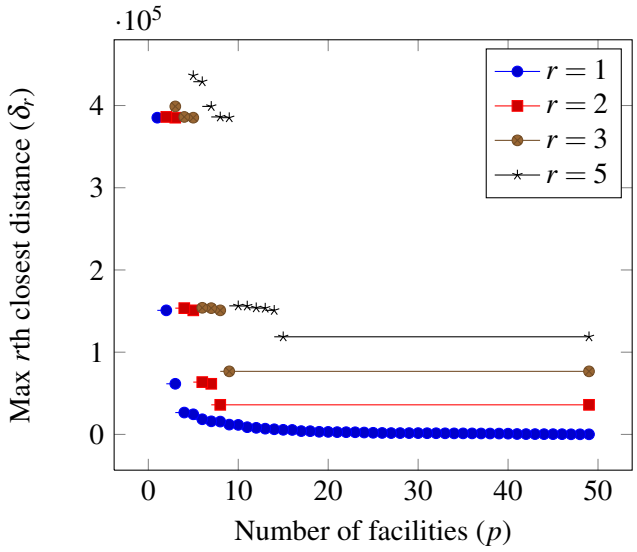
Insight 1: It is better to locate facilities in anticipation of a disruption and be wrong than to locate facilities without considering disruptions and be wrong. A weakness of the RANPCP model is that it only models potential consequence without modeling regular system performance. This is a problem 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. For a given instance, let $Y^*_{(1)}$ be the optimal facility configuration for the non-failure radius objective and let $Y^*_{(r)}$ be the optimal facility configuration for the post-failure 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 *penalty 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 *penalty for not considering facility unavailability* be $\gamma_{r,1} = \frac{f_{(r)}(Y^*_{(1)}) - f_{(r)}(Y^*_{(r)})}{f_{(r)}(Y^*_{(r)})}$.

Table 3.2 shows summary statistics for $\gamma_{1,r}$ and $\gamma_{r,1}$ over all of our datasets and instances. The table shows that, on average, the penalty for not considering facility unavailability is 0.62 times the penalty for not considering regular system performance. Thus, the max closest distance and max r^{th} closest distance objectives are conflicting.

Table 3.2: 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.	min	max	avg.	min	max	avg.
d49	1.10	3.10	1.90	0.54	11.00	5.20	0.16	4.10	1.10
sw55	0.18	0.96	0.51	0.32	1.70	1.10	0.21	1.00	0.53
d88	1.10	15.00	5.10	0.76	18.00	7.00	0.13	16.00	3.30
lor100	0.00	0.92	0.48	1.40	9.70	5.90	0.00	0.37	0.15
d150	1.90	19.00	6.20	1.10	21.00	7.50	0.14	17.00	2.30
lon150	0.00	0.69	0.39	0.59	5.80	2.00	0.00	0.94	0.33
lor200	0.00	0.84	0.47	0.90	21.00	10.00	0.00	0.58	0.12
lor300a	0.00	1.50	0.64	0.81	26.00	13.00	0.00	1.00	0.16
lor300b	0.00	1.50	0.64	0.81	26.00	13.00	0.00	1.00	0.16
lor400a	0.00	1.30	0.52	0.85	37.00	16.00	0.00	1.30	0.15
lor400b	0.00	1.30	0.52	0.85	37.00	16.00	0.00	1.30	0.15
beas500	0.15	0.39	0.30	0.18	2.10	1.00	0.10	1.60	0.52
beas600	0.16	0.37	0.25	0.08	2.70	1.30	0.07	2.10	0.48
beas700	0.03	0.49	0.25	0.11	3.30	1.80	0.01	3.90	0.66
beas800	0.03	0.31	0.20	0.15	3.40	1.90	0.02	1.90	0.27
lor818	0.28	1.30	0.61	0.46	11.00	3.90	0.03	1.30	0.37
beas900	0.08	0.38	0.17	0.38	4.90	3.00	0.02	0.99	0.15
u1060	0.00	0.86	0.53	0.50	8.00	4.60	0.00	0.91	0.22
ALL	0.00	19.00	1.09	0.08	37.00	6.34	0.00	17.00	0.62



(a) Daskin 49-node dataset (d49)

Figure 3.3: Budget vs. post-failure radius for Daskin 49-node dataset

Insight 2: Large Decreases in Vulnerability Can Be Obtained with Small Increases in Cost The budget, b, 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 post-failure radius. This curve can be generated by solving MSCP(δ, r) for all values of δ in the matrix $\{\phi_{ij} : i \in \mathscr{I}, j \in \mathscr{J}\}$.

Figure 3.3 shows tradeoff curves for b vs. post-failure radius for the USCities dataset with several values of r. The curves have several flat areas, at which locating additional facilities does not reduce the post-failure much. On the other hand, the curves also has several jumps where adding one additional facilities significantly reduces the post-failure radius.

3.2 Locating and Hardening Facilities Subject to Failure[2]

While the RANPCP model in the previous section minimizes the post-failure radius by strategically locating facilities, another possibility is to protect located facilities. Section 2.5.1 reviewed the literature on models for allocating protection resources among facilities. Typically, researchers have developed facility hardening models, in which a defender chooses a subset of the located facilities to harden. When a facility is hardened, it is immune to failures. Due to a budget constraint, the defender can only harden a limited number of facilities.

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

[2]This subsection is a summary of the results in Medal, H., Pohl, E., and Rossetti, M., 2011. On the R-All-Neighbor P-Center Problem. *Under revision*.

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 bi-level 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.

In this Section, we describe a model for making facility location and facility hardening decisions in an integrated way rather than a sequential way. 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-failure 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.

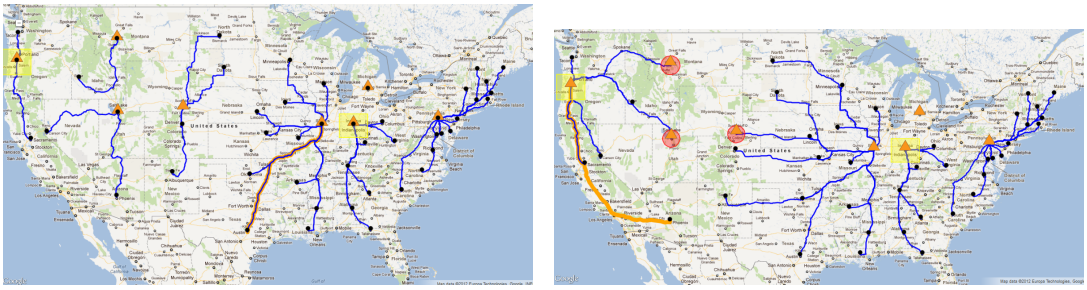
This Section will show that locating and hardening facilities with an integrated model produces much better solutions than when a sequential method is used. This is due to two reasons. First, the integrated method that we present subsumes the sequential method, so the integrated method guarantees solutions that are at least as good as the sequential method. Second, if using the sequential method, the decision-maker must first decide a proportion of the budget to allocate to location, leaving the remainder for hardening. In addition, optimizing this proportion is not straightforward.

To illustrate that the sequential method can produce bad solutions, let us continue the example from Section 3.1.3. Let the cost of hardening a facility be 2.5 times the cost of locating that facility. Figure 3.4 shows the optimal set of facilities to locate and harden given that 60% of the budget is allocated to locating facilities and 40% is allocated to hardening facilities. A version of the *r*-all-neighbor *p*-center model from Section 3.1 was used to decide where to locate facilities and a hardening model (Medal *et al.*, 2011b) was used to choose which facilities to harden. Since only 60% of the budget was used for location, only eight facilities are located in the first stage. The figure also shows that in the second stage, two facilities were hardened: one in the east and one in the west.

Figure 3.4a shows the assignments without failures. Because less facilities were located, the non-failure radius increases from 294 in the *p*-center solution to 798, a 171% increase (recall the 49% increase for the RANPCP model). The bottleneck pair (Springfield and Austin) is in the Midwest, as opposed to the west in the RANPCP solution.

Figure 3.4b shows the assignments after the failure of the three facilities whose failure maximizes the post-failure radius. These failed facilities are all located in the west, as opposed to the northeast in the RANPCP solution. The post-failure radius decreases from 1071 in the *p*-center solution to 979, a 9% increase (recall the 33% increase for the RANPCP model). Thus, the sequential solution produced a worse solution than the RANPCP model.

While a decision-maker would probably never used the solution prescribed by this sequential method, this example demonstrates that a sequential method can produce poor solutions. In the remainder of this section an integrated model is described, which produces better solutions than the sequential approach.



(a) Demand point assignments without failures (b) Demand point assignments after failures

Figure 3.4: Locate-then-harden solution (40% proportion)

3.2.1 Model

To facilitate our analysis, we developed a mathematical model for the minimax facility location-hardening problem (MFLHP). The MFLHP model is similar to the RANPCP model described in Section 3.1. The MFLHP 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 failure scenarios consisting of the failure of r facilities. The consequence of a failure scenario is the maximum distance from a demand point to its closest located and operating facility.*

3.2.2 MIP Model

In the MFLHP model, each demand point is treated as a post-interdiction bottleneck demand point and is assigned to its post-interdiction bottleneck facility. Let X_i be the number of facilities 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 $X_i > 0 \implies Z_i = 0$ and $Z_i = 1 \implies X_i = 0$.)

A MIP formulation of the integrated MFLHP model is:

$$\min \quad U \tag{3.3a}$$

$$\text{s.t.} \quad U \geq \phi_{ij} W_{ij} \quad \forall i \in \mathscr{I}, j \in \mathscr{J} \tag{3.3b}$$

$$(r+1)W_{ij} \leq (r+1)Z_i + \sum_{i' : \phi_{i'j} \leq \phi_{ij}} X_{i'} \quad \forall j \in \mathscr{J}, i \in \mathscr{I} \tag{3.3c}$$

$$\sum_{i \in \mathscr{I}} W_{ij} = 1 \quad \forall j \in \mathscr{J} \tag{3.3d}$$

$$W_{ij} \leq X_i + Z_i \quad \forall i \in \mathscr{I}, j \in \mathscr{J} \tag{3.3e}$$

$$\sum_{i \in \mathscr{I}} f_i X_i + \sum_{i \in \mathscr{I}} (f_i + g_i) Z_i \leq b \tag{3.3f}$$

$$X_i, Z_i \in \{0,1\} \quad \forall i \in \mathscr{I} \tag{3.3g}$$

$$W_{ij} \in \{0,1\} \quad \forall i \in \mathscr{I}, j \in \mathscr{J} \tag{3.3h}$$

The objective equation (3.3a), in conjunction with Constraints equation (3.3b), is to minimize the post-interdiction radius. Constraints equation (3.3c) 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 equation (3.3d) require that every demand point form a post-interdiction bottleneck pair with one facility. Constraints equation (3.3e), although not necessary because of the presence of Constraints equation (3.3c), tighten the LP relaxation. Constraint equation (3.3f) requires that the amount spent on location and hardening must be within a budget. Constraints equation (3.3g)–equation (3.3h) specify bounds on the variables.

3.2.3 Solution Procedure

Model (3.3) 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.

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 chapter, we solve the integrated MFLHP using a binary search algorithm with a modified auxiliary problem.

To use a binary search algorithm for the integrated MFLHP, the auxiliary problem must first be described. Define U as the radius for the auxiliary problem. (Note that U is now a parameter and not a variable, as it was in Model (3.3).) To evaluate whether a particular U is above or below the optimal post-interdiction radius, the set-cover problem with location and hardening (SCP-LH) is used:

$$\text{SCP-LH}(U) \quad \min \quad \sum_{i \in \mathscr{I}} f_i X_i + \sum_{i \in \mathscr{I}} (f_i + g_i) Z_i \tag{3.4a}$$

$$\text{s.t.} \quad (r+1) \sum_{i:\phi_{ij} \le U} Z_i + \sum_{i:\phi_{ij} \le U} X_i \ge r+1 \quad \forall j \in \mathscr{J} \tag{3.4b}$$

The SCP-LH minimizes the cost required for every demand point to have a post-interdiction assignment distance less than or equal to U. The objective equation (3.4a) is to minimize the total cost of location and hardening Constraints equation (3.4b) require that for each demand point j, either $r+1$ facilities within U of j must be located or at least one facility within U of j must be hardened.

3.2.4 Example Continued

The example at the beginning of Section 3.2 showed that the sequential method can produce poor solutions. As mentioned above, one of the reasons for the poor performance of the sequential method is that it is unclear what proportion of the budget to allocate to the location stage. Continuing the example from the beginning of Section 3.2, we obtained the ideal proportion to allocate to location by enumerating all proportions in the set $\{0, .01, 0.02, \ldots, 1\}$. We found that the ideal proportion is 1.0. That is, when using the sequential method, it is ideal to allocate all of the budget to location. Thus, the solution for the sequential method with the ideal proportion is identical to the solution generated by the RANPCP model in Figure 3.2.

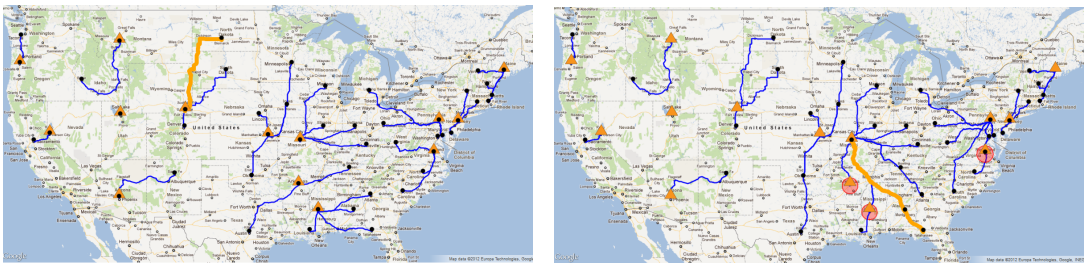
Thus, the question remains: can an integrated model produce better solutions than the sequential method with the ideal proportion. Continuing the example from the beginning of Section 3.2, Figure 3.6 shows the solution generated by the integrated MFLHP. Since the MFLHP includes hardening as an option, the MFLHP solution locates one less facility than the RANPCP solution, leaving enough resources to harden the facility at Harrisburg.

Figure 3.6a shows the assignments without failures for the MFLHP solution. The non-failure radius increases from 294 for the p-center solution to 466, a 59% increase (compare with the 49% increase for the sequential method with ideal proportion).

Figure 3.6b shows the assignments after failures for the MFLHP solution. The three failures occur in the west, perhaps partly because the hardened facility is on the east coast. The post-failure radius increases from

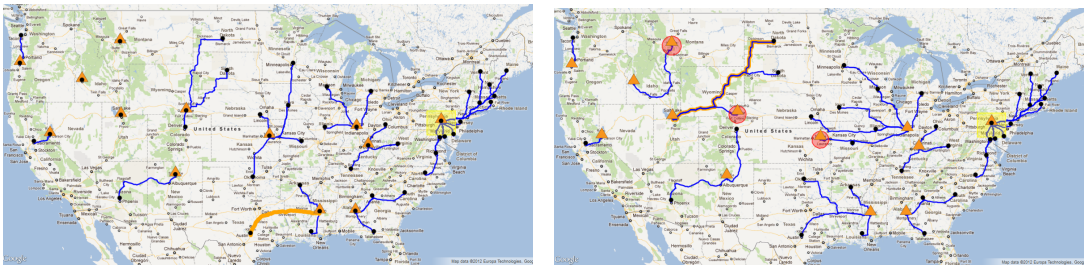
1071 for the p-center solution to 694, a 35% decrease (compare with the 33% decrease for the sequential method with ideal proportion).

In this example, the MFLHP model produced a solution with a slightly better post-failure radius than the sequential method solution. The next section shows that on average, the benefit from hardening is much higher.



(a) Demand point assignments without failures (b) Demand point assignments after failures

Figure 3.5: Locate-then-harden solution (ideal proportion)



(a) Demand point assignments without failures (b) Demand point assignments after failures

Figure 3.6: Integrated location-hardening solution

3.2.5 Insights

To generate the insights in this section, the model was tested using a variety of datasets taken from the facility location literature and adapted to include the cost of facility hardening (for a complete list see Medal *et al.* (2011b)). For each dataset, several combinations of parameter settings were tested (Medal *et al.*, 2011b) and summary statistics were calculated across the combinations.

Insight 1: Hardening Helps Make Networks Less Vulnerable Let $f_{(r)}(\cdot)$ be the optimal maximum post-failure maximum distance for a given solution. Let $X^*_{(r)}$ be the optimal solution to the RANPCP model described in Section 3.1, which does not consider facility hardening. Let $XZ^*_{(r)}$ be the optimal location and hardening solution produced by the integrated location-hardening model, described in Section 3.2.2. Let $\varepsilon^I = \frac{f_{(r)}(X^*_{(r)}) - f_{(r)}(XZ^*_{(r)})}{f_{(r)}(XZ^*_{(r)})}$ be the *penalty for not considering hardening*. For the USCities dataset, the average value of ε^I was 1.26. Thus, hardening significantly decreased the vulnerability of the network of located facilities.

Insight 2: Integrating Location and Hardening Decisions Helps Make Systems More Resilient In this section we report empirical results that indicate that there is a measurable benefit in integrating the location and hardening decisions. Let

$$\gamma(L) = \frac{f_{(r)}(X^*_{(r)}, Z^*_{S(L)}) - f_{(r)}(XZ^*_{(r)})}{f_{(r)}(XZ^*_{(r)})} \tag{3.5}$$

be the penalty for not integrating the location and hardening decisions when allocating L of the budget to location. In other words, this is the penalty incurred when the LTH method is used in place of the integrated MFLHP model. Table 3.3 summarizes the value of this penalty for several datasets.

Table 3.3: Average penalties for not integrating location and hardening decisions

Dataset	Avg. penalty for not integrating
sw55	0.55
lor100	3.12
lon150	1.31
lor200	3.15

3.3 Conclusions

This section described work on reducing the vulnerability of networked facilities that are subject to failures. In particular, this work focused on developing methods to minimize the post-failure radius, an objective that generates solutions that are equitable and risk-averse. Because these methods generate risk-averse solutions, they can be used to model a set of networked facilities that are vulnerable to an adaptive attacker. In this context, if the attacker attacks a facility it is always becomes completely destroyed.

First, we discussed a model for locating facilities in order to minimize the post-failure radius, called the r-all-neighbor p-center problem (RANPCP). We first presented an MIP formulation for this problem. Rather than solving the MIP formulation with available branch-and-bound codes, we solved the problem using a binary search algorithm. The binary search algorithm uses a modified set cover problem as sub-problems.

Experimenting with the RANPCP model led to two main insights:

1. It is better to locate facilities in anticipation of a disruption and be wrong than to locate facilities without considering disruptions and be wrong.

2. Large decreases in vulnerability can be obtained with small increases in cost.

Second, we discussed a model for integrating facility location and facility protection decisions. In the model, protected facilities are said to be hardened, or immune to failures. Again, we presented an MIP formulation for this problem and solved it using a binary search algorithm. Experimenting with the location-hardening model led to two main insights:

1. Hardening helps make networks less vulnerable.

2. Integrating location and hardening decisions helps make networks more resilient.

The work presented in this section points to several areas of future work. First, it would be useful to relax the two main assumptions made: 1) attacks are 100% successful and 2) Hardened facilities are immune to failures. In reality, allocating protection to a facility decreases its probability of failure and allocating attack resources to a facility increases its probability. Second, rather than modeling a facility network in isolation, it would be useful to model interdependent networks that are subject to cascading failures.

Chapter 4

Identifying Vulnerable Infrastructure Elements in a Unit Train Transportation System

4.1 Introduction

Today, our world depends on its transportation systems. A large percentage of the products we consume are transported long distances by road, rail, air, or a combination of modes. In addition, many people travel on roads to go to work every day.

The world's transportation systems are large and complex systems that are exposed to many types of risks. One of the risks is the failure of infrastructure elements such as bridges, tunnels, and facilities. These elements can fail due to natural disasters, terrorist attacks, or just because they are in bad condition. The failure of these elements can causes several different impacts including loss of life, economic loss, increased travel costs and congestion since the routes need to be changed to avoid the failed elements.

Rail transportation is an important and growing component of freight transportation in the United States. The benefits of rail transportation are that it is cheaper and produces less carbon emissions than road transportation. It is also easier to transport heavy loads on rail than on truck. Leaders in transportation are trying to increase the volume of goods transported by rail to alleviate load on the road transportation system and reduce carbon emissions. Large freight companies also are moving more of their transportation to a combination of rail and road.

There are several aspects of rail transportation that make it different than other transportation modes. First, the operations of a railroad are more centrally controlled than in road transportation. That is, train operators have less autonomy to choose their own routes and schedules. Second, compared to road transportation, there is not as much excess capacity in rail transportation. Thus, it is important to consider capacity when routing and scheduling.

Several events in the last 30 years illustrate that the freight rail transportation system in the United States is vulnerable to disruptions. In 1993, flooding of the Mississippi and Missouri rivers caused several railroads to experience delays and cancellations. The estimated total cost of the disruption was $ 182 million (Haefner, 1996). In 1996, a merger between Union Pacific and Southern Pacific railroads led to delays for many of Union Pacific's customers (Quillen, 1997). In 2005, a delrailment on a main line in Wyoming near the Powder River Basin led to a shortage of coal in many parts of the United States as well as price increases (Bleizeffer, 2006). Finally, 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 (Boyd, 2011).

Disruptions have a high impact in rail transportation because there are less alternate routes available when a disruption occurs. There are several reason for the lack of alternate routes. First, rail is not as ubiquitous as roads. Second, much of the track in the United States is single line track. Thus, only one train can be on the track at a time in either direction. This makes it more difficult to reroute trains after a disruption. Third, the operation of a railyard can be complex and therefore it is difficult for a railyard to accommodate a lot of excess capacity. Again, this must be taken into consideration when rerouting.

Deciding how to reduce the risk of the rail transportation system is difficult for economic reasons also.

Rail infrastructure is very expensive. Therefore, it is important to understand the cost benefit of risk reduction activities. Rail infrastructure is usually paid for by railroads. Thus, before investing in risk reduction measures such as additional security or upgraded infrastructure, railroads must be confident that it will help their profitability.

In this paper, we consider rail transportation of bulk commodities such as coal, grain, and scrap metal. Bulk commodities make up a large percentage of the volume transported on rail, and coal is 47% of the total volume. The transportation of bulk commodities is different than the transportation of other merchandise for several reasons. In bulk transportation, demand is in entire trains; therefore, there is no need to switch cars at intermediate classification yards. The demand in bulk transportation is also smoother than the demand for lower volume items. For example, several power plants in the southern United States place a fixed-quantity order one every month.

There have been many useful studies of how to reduce the risk of networks that can be applied to transportation networks. However, the mathematical models employed in these studies may not have enough detail to be directly applied to rail networks. Existing models mostly have modeled goods as continuous, i.e., divisible, quantities. However, in most settings trains can only be realistic modeled using discrete units of flow.

Existing models also are usually static, meaning that all flow happens at the same time. Although real flows are almost never static, modeling flows as static is appropriate for uncapacitated networks or networks in which there are capacity constraints over long time periods (e.g., a month) but there are not strict capacity constraints for shorter time intervals (e.g., day or hour). For example, it is not necessary to consider strict capacity constraints for short time periods for long-haul freight trucking networks in the United States (US) because US highways have a lot of excess capacity. Static flow models are also appropriate for strategic level decisions because routes can be modified at the operational level to account for capacity.

However, assuming static flow is probably not realistic for modeling the operational level of rail transportation. First, trains take a significant amount of time to travel a route. Second, there are strict capacity constraints on rail infrastructure for short time intervals. Because trains take a long time to stop, railroads prefer to only schedule one train on line in either direction every hour. Railyards also have limited capacity, with a large yard having between 60 and 80 tracks and a small yard having between 10 to 20 tracks. The number of tracks at a yard limits the number of trains that can pass through that yard in a period of time.

In this paper we first present a model for estimating the consequence of a disruption to a transportation network. Second, we present a model in which an attacker optimally chooses a set of infrastructure elements to attack in order to maximize the total disruption to the network, i.e., an interdiction model. In addition to modeling the threat of an attacker, this interdiction model can also be used to determine critical elements of the network. The consequence estimation model mimics a unit train transportation system by modeling trains as discrete units that stay intact from origin to destination. The model captures the movement of trains in time and space over a finite time horizon. Lines and railyards in the network have strict capacity constraints for short time periods. The interdiction model uses the consequence estimation model as a subroutine and identifies the set of lines and facilities whose unavailability causes the largest consequence.

There are several questions that we try to answer in this chapter. First, what is the solution-quality improvement of time-space consequence estimation model over simpler consequence estimation models? Also, is our time-space interdiction model more accurate than simpler interdiction models? Second, how much more computational resources do time-space consequence estimation and interdiction models require than simpler consequence estimation and interdiction models? Third, how vulnerable to disruptions are rail transportation networks?

Hence, the contributions of this paper as follows: 1) vulnerability assessment models for dynamic transportation systems, 2) an analysis of the freight unit train transportation system, 3) and a solution methodology for interdiction models in which the operator's problem is discrete.

4.2 Literature Review

There is an established body of research on the rail transportation system. Assad (1980), Ahuja *et al.* (2005), and Nemani and Ahuja (2011) provide surveys of this topic. Crainic (2000) has surveyed the research on freight transportation. He discussed three planning levels: strategic, tactical, and operational. In the strategic level, long-term decisions are made such as where to locate yards and where to build rail lines. In the tactical level, medium-term decisions are made such as the routing of trains and aggregate scheduling. The operational level includes shorter-term decisions such as crew scheduling and locomotive scheduling.

Many strategic rail decisions are informed by an estimate of current or future capacity. Capacity is influenced by many factors such as infrastructure and demand. Several authors have developed approaches to assess railway capacity (Kozan and Burdett, 2005; Burdett and Kozan, 2006; Mattsson, 2007; Abril *et al.*, 2008). This is related to our work because we include railyard and line capacity in our model.

There are two types of rail transportation: merchandise trains and unit trains. Merchandise trains are composed of cars with different destinations. Therefore, consolidation, or blocking, is a crucial part of merchandise train operations. Partly due to the challenging problems associated with the blocking process, most of the research on rail transportation from an operations research perspective has consider merchandise trains (see Nemani and Ahuja (2011)). Unit trains are composed of cars with the same destination; thus, blocking is no longer needed. There is not as much research on unit train transportation. Lawley *et al.* (2008) present a time-space routing and scheduling model for unit trains. Their model accounts for both loaded and empty trains. The second stage of our interdiction model is similar to this model except that we do not account for empty trains.

Because of the prevalence of disruptions in transportation networks, there has been a significant amount of work on managing the recovery from a disruption. Applications include machine scheduling (Qi *et al.*, 2006), production-inventory systems (Xia *et al.*, 2004), supply chains (Qi *et al.*, 2004), passenger air transportation (Kohl *et al.*, 2007), passenger rail transportation (Jespersen-Groth *et al.*, 2009), and project scheduling (Zhu *et al.*, 2005). Yu and Qi (2004) have written a book that discusses these topics.

In the last few decades there has been more studies on assessing the vulnerability and reliability of networks. For more information, one can see several recent surveys in (Berdica, 2002; Grubesic *et al.*, 2008; Murray *et al.*, 2008; Sullivan *et al.*, 2009).

One way to study the vulnerability of a network is to identify the critical nodes and edges of the network. Interdiction models identify critical nodes and edges by modeling a game between an adversary and the operator of the network, who routes flow through the network after the adversary makes his attack. 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 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. Others have extended this problem to consider probabilistically successful attacks (Cormican *et al.*, 1998; Janjarassuk and Linderoth, 2008) and multiple objectives (Royset and Wood, 2007; Rocco *et al.*, 2009, 2010). Researchers have also considered other objectives such as minimizing the maximum reliability path (Pan and Morton, 2008), minimizing the maximum profit (Lim and Smith, 2007). Further, Church *et al.* (2004) presented models for interdicting a set of facilities. A survey of interdiction models is given in Smith and Lim (2008) and Smith (2011).

Researchers have also begun to study the vulnerability of freight rail networks. Peterson and Church (2008) described models for the impact of a disruption to the United States freight transportation network. They present an uncapacitated model that is a modification of the shortest path problem. They also present a continuous multicommodity network flow model that has line capacities. Babick (2009) modeled the allocation of security resources to the rail network in the state of California as a defender-attacker-operator problem, represented by a bilevel mixed-integer programming formulation of the problem. In the first stage,

the defender chooses arcs in the network to harden, or protect from failure. In the second stage, an attacker chooses arcs to destroy. In the final stage, the operator solves a continuous multicommodity network flow problem on the elements of the network that have not been destroyed. Both of these models model the rail transportation as a continuous static network flow problem. In our work we model the rail transportation system as a discrete dynamic network flow problem.

4.3 Consequence Estimation Model

4.3.1 Notation

In this section we introduce the sets, parameters and decision variables that are used in the two-stage integer-programming formulation of the interdiction model.

Sets

- N set of all loading/unloading stations (nodes)
- $M \subseteq N$ set of mines
- $P \subseteq N$ set of plants
- A set of all track segments
- R set of feasible routes between all O–D pairs for all trains
- $RT(r) \subseteq A$ set of track segments included in the route of O–D pair $r \in R$
- $RN(r) \subseteq N$ set of nodes included in the route of O–D pair $r \in R$
- D set of days making up the planning horizon
- $T(d)$ set of time periods in day $d \in D$

Parameters

- $\triangle \in Z^+$ length of planning period
- K number of planning periods in planning horizon
- T_{beg} beginning time of planning horizon
- T_{end} ending time of planning horizon, $T_{end} = T_{beg} + K\triangle$
- T set of all time periods $\{T_{beg}, T_{beg} + \triangle, T_{beg} + 2\triangle..., T_{beg} + K\triangle = T_{end}\}$
- $o(r) \subseteq M$ origin station of route $r \in R$
- $h(r) \subseteq P$ destination station of route $r \in R$ (Plants)
- TC_{at} track capacity of segment $a \in A$ at time $t \in T$
- TC_{it} track capacity of node $i \in N$ at time $t \in T$
- H_i holding capacity at node $i \in P \cup M$
- L_i loading/unloading capacity at node $i \in P \cup M$
- U_{it} loading/unloading time for a train arriving at time $t \in T$ in node $i \in h(r)$, $r \in R$

- τ_r total travel time of route $r \in R$ in multiples of $\triangle$
- τ_{ra} travel time on route $r \in R$ to reach track segment $a \in RT(r)$ in multiples of $\triangle$
- τ_{ri} travel time on route $r \in R$ to reach node $i \in RN(r)$ in multiples of $\triangle$
- d_r distance between route $r \in R$
- c_r cost of route r, which includes:
  - fixed costs: labor cost, cost of using cars and engines
  - variable costs: fuel cost, maintenance costs, etc.
- h_i demand of station i over the planning horizon
- l_t length of the time period of $t \in T$ (in hrs)
- q cost of incurred when 1 train is delayed 1 hour

Decision variables

- X_{rt} number of trains departing from $o(r)$ on route $r \in R$ in time period $t \in T$
- O_{rt} number of trains waiting (or being loaded/unloaded) at $o(r)$ of route $r \in R$ in time period $t \in T$

4.3.2 Formulation with IP Second Stage

We propose a pure integer model for the second stage decisions after the disruptions occurred. Basically, routing and scheduling decisions of the trains need to be made given a network with available nodes and arcs after the scenario realization in the first stage. Therefore, our second stage IP model aims to satisfy the demands of plants with minimum cost and without eliminating the capacity restrictions of network elements while dispatching trains from mines to plants through predetermined routes. A time-indexed formulation captures the true capacity limitations of nodes and arcs in any given period. Flexibility of being able to arrange the length of planning period provides great control on the scale of the problem as well. For the sake of simplicity, we only consider the flow of identical unit trains that carry the same amount of coal regardless of the origin destination pair they are assigned to. Even though a set of feasible routes is used as an input, the model selects the cheapest route first and then schedule trains according to capacity and demand requirements. Finally, we do not consider the routing and scheduling of empty trains from plants to mines. It is assumed that there is only one way flow from mines to plants.

$$\ell(\hat{T}) = \min \sum_{r\in R}\sum_{t\in T} c_r X_{rt} + \sum_{r\in R}\sum_{t\in T} ql_t O_{rt} \tag{4.1a}$$

$$\text{s.t.} \quad O_{rt} + X_{rt} \le O_{r,t-\triangle} \qquad \forall r \in R\, t = \Delta, \Delta+1, \ldots, T \tag{4.1b}$$

$$\sum_{t\in T(d)}\sum_{r|i\in RN(r)} X_{r,t-\tau_{ri}} \le TC_{id} \qquad \forall d \in D\, i \in N \tag{4.1c}$$

$$\sum_{t\in T(d)}\sum_{r|a\in RT(r)} X_{r,t-\tau_{ra}} \le TC_{ad} \qquad \forall d \in D\, a \in A \tag{4.1d}$$

$$X_{rt} \le 1 - \hat{z}_i \quad \forall i \in N, r|i \in RN(r), t \in \mathscr{T} \tag{4.1e}$$

$$X_{rt} \le 1 - \hat{z}_a \quad \forall a \in A, r|a \in RT(r), t \in \mathscr{T} \tag{4.1f}$$

$$\sum_{t\in T}\sum_{r|i=h(r)} X_{r,t} = h_i \qquad \forall i \in P \tag{4.1g}$$

$$\sum_{r\in R}(X_{rt} + O_{rt}) \le n \quad \forall t \in \mathscr{T} \tag{4.1h}$$

$$X_{rt}, O_{rt} \in \{0, 1, \ldots, n\} \qquad \forall r \in R, t \in \mathscr{T} \tag{4.1i}$$

Objective function (4.1a) is the total cost incurred by 1) the total distance traveled and 2) the total delay incurred when trains have to wait at their origin stations. Constraint set (4.1b) balances flow at the origin station of route r. For each planning period $(\triangle)$ and route (r), number of trains waiting at the origin node and departing the origin node to travel on route r at time period t can not be greater than number of trains available at the origin node Δ time units before, at time period $(t-\triangle)$. Constraint sets (4.1c) and (4.1d) guarantee that the total number of trains sent on a track segment and through a node in any given day does not exceed the daily track segment and node capacity, respectively. If node $i \in RN(r)$ or arc $a \in RT(r)$ have failed on a route r, constraints (4.1e) and(4.1f) assure that trains are not dispacthed through route r for each time period. Constraints (4.1g) state that the demand of each plant should be satisfied in the planning period. Finally, constraints (4.1h) ensure that number of trains waiting at the origin node and departing that node must be less than or equal to n, the number of trains, for each route and time period.

4.3.3 Formulation with Binary Second Stage

The second stage pure IP model (4.1) described in Section Section 4.3.2 can also be formulated as a binary IP in model (4.2). But first, we need to define set K as the set of unit trains circulating in the system. The following are the modified decision variables;

- $x_{rt}^k = 1$ if train k is departed from $o(r)$ at time period t, 0 otherwise
- $o_{rt}^k = 1$ if train k is waiting (or being loaded/unloaded) at $o(r)$ of route r in time period t, 0 otherwise

The 0-1 second stage formulation is as follows:

$$\ell(\hat{T}) = \min \sum_{k\in K}\sum_{r\in R}\sum_{t\in T} c_r x^k_{rt} + \sum_{k\in K}\sum_{r\in R}\sum_{t\in T} q l_t o^k_{rt} \tag{4.2a}$$

$$\text{s.t.} \quad o^k_{rt} + x^k_{rt} \le o^k_{r,t-\triangle} \quad \forall r \in R\, t \in T\, k \in K \tag{4.2b}$$

$$\sum_{r\in R}\sum_{t\in T} x^k_{rt} \le 1 \qquad \forall k \in K \tag{4.2c}$$

$$o^k_{rt} + x^k_{rt} = 1 \qquad \forall k \in K\, o(r) \in R\, t \in T \tag{4.2d}$$

$$\sum_{k\in K}\sum_{t\in T(d)}\sum_{r|i\in RN(r)} x^k_{r,t-\tau_{ri}} \le TC_{id} \qquad \forall d \in D\, i \in N \tag{4.2e}$$

$$\sum_{k\in K}\sum_{t\in T(d)}\sum_{r|a\in RT(r)} x^k_{r,t-\tau_{ra}} \le TC_{ad} \qquad \forall d \in D\, a \in A \tag{4.2f}$$

$$x^k_{rt} \le 1 - \hat{z}_i \quad \forall k \in K, i \in N, r|i \in RN(r), t \in \mathscr{T} \tag{4.2g}$$

$$x^k_{rt} \le 1 - \hat{z}_a \quad \forall k \in K, a \in A, r|a \in RT(r), t \in \mathscr{T} \tag{4.2h}$$

$$\sum_{k\in K}\sum_{t\in T}\sum_{r|i=h(r)} x^k_{r,t} = h_i \qquad \forall i \in P \tag{4.2i}$$

In Model (4.2), each occurence of X_{rt} and O_{rt} in Model (4.1) is replaced with $\sum_{k\in K} x^k_{rt}$ and $\sum_{k\in K} o^k_{rt}$, respectively. Like Constraints (4.1b), Constraint set (4.2b) ensures that flow is balanced for each unit train k. Constraints (4.2c) assure that a train can not be assigned to more than one route in time period T. Constraints (4.2d) require that every train k must be either leaving or waiting on some route in every time period.

Note that there are more decision variables in model 4.2 than in model 4.1. This is because, in addition to sets T and R, we also incorporated unit trains in order to address the new decision variables, x^k_{rt} and o^k_{rt}, in the binary formulation. Even though model 4.2 provides ability to control each unit train's movement on a timely basis, it takes much more time to create and solve the binary model due to memory problems. In the next section, we will discuss the solution methodology adapted to solve the version of the two-stage interdiction model in which the second stage has integer variables.

4.4 Identifying Critical Elements

In this section we model the problem of identifying critical elements as a two-player game. In this game an interdictor acts first and destroys a set of nodes and arcs. An operator follows the interdictor and chooses routes and schedules for trains given network elements that have not failed. This game can be modeled as a bi-level integer program. Let T be a vector of interdiction variables in which T_i is 1 if node i is destroyed and 0 otherwise. Let f_i be the cost of interdicting node i. The interdictor has a budget of b to spend on interdicting nodes. Let Y be the feasible region of T defined by the following constraints:

$$T_i \in \{0,1\} \quad \forall i \in \mathscr{N} \tag{4.3a}$$

$$\sum_{i\in\mathscr{N}} f_i T_i \le b \tag{4.3b}$$

In the following section, we demonstrate a two-stage interdiction model, first stage of which is represented by the constraints 4.3a and 4.3b.

4.4.1 Interdiction Model

First, we present a capacity-interdiction model for the routing and scheduling of trains. This model is a *capacity-interdiction* model because the destruction of a node or arc causes that node or arc to have zero capacity.

Consider the following bi-level capacity-interdiction model:

$$\max_{T\in Y} \ell(T) = \quad \min \quad \sum_{r\in R}\sum_{t\in T} c_r X_{rt} + \sum_{r\in R}\sum_{t\in T} ql_t O_{rt} \tag{4.4a}$$

$$\text{s.t.} \quad O_{rt} + X_{rt} \geq O_{r,t-\triangle} \qquad \forall r \in \mathscr{R}, t = \Delta, \Delta+1, \ldots, T \tag{4.4b}$$

$$\sum_{t\in T(d)} \sum_{r|i\in RN(r)} X_{r,t-\tau_{ri}} \leq c_{id}(1-T_i) \qquad \forall d \in \mathscr{D}, i \in \mathscr{N} \tag{4.4c}$$

$$\sum_{t\in T(d)} \sum_{r|a\in RT(r)} X_{r,t-\tau_{ra}} \leq c_{ad} \qquad \forall d \in \mathscr{D}, a \in \mathscr{A} \tag{4.4d}$$

$$\sum_{t\in T} \sum_{r|i=f(r)} X_{r,t} \geq w_j \qquad \forall i \in \mathscr{P} \tag{4.4e}$$

$$\sum_{r\in\mathscr{R}} (X_{rt} + O_{rt}) \leq n \quad \forall t \in \mathscr{T} \tag{4.4f}$$

$$X_{rt}, O_{rt} \in \{0,1,\ldots,n\} \qquad \forall r \in \mathscr{R}, t \in \mathscr{T} \tag{4.4g}$$

This model adds another level to Model (4.1) and replaces the z variables with interdiction variables T. Constraints (4.4c) force an interdicted node to have zero capacity.

4.4.2 Reformulation

One approach to solve the program (4.4) involves reformulating the problem as a single level MIP. One way to do this is to fix the T variables, relax the integrality restriction on the X and O variables, add constraints to form the convex hull of the inner minimization (further details about this will be added later in this section; for now just assume that we have the convex hull), and take the dual of the inner minimization. Since both levels are then maximization after taking the inner dual, the bi-level problem reduces to a single level mixed-integer program.

First, relax the inner minimization problem and add constraints to form the convex hull:

$$\max_{T\in Y} \tilde{\ell}(T) = \quad \min \quad \sum_{r\in R}\sum_{t\in T} g_r X_{rt} + \sum_{r\in R}\sum_{t\in T} ql_t O_{rt} \qquad [\text{duals}] \tag{4.5a}$$

$$\text{s.t.} \quad O_{rt} + X_{rt} \geq O_{r,t-\triangle} \qquad \forall r \in \mathscr{R}, t = 1,\ldots,|T| \qquad [\alpha_{rt}] \tag{4.5b}$$

$$\sum_{r|i\in\mathscr{N}(r)} \sum_{\substack{t\in\mathscr{T}(d)\\ t\geq\tau_{ri}}} X_{r,t-\tau_{ri}} \leq c_{id}(1-T_i) \qquad \forall i \in \mathscr{N}, d \in \mathscr{D} \qquad [\beta_{id}] \tag{4.5c}$$

$$\sum_{r|a\in\mathscr{A}(r)} \sum_{\substack{t\in\mathscr{T}(d)\\ t\geq\tau_{ra}}} X_{r,t-\tau_{ra}} \leq c_{ad} \qquad \forall a \in \mathscr{A}, d \in \mathscr{D} \qquad [\gamma_{ad}] \tag{4.5d}$$

$$\sum_{t\in T} \sum_{r|i=f(r)} X_{r,t} \geq w_j \qquad \forall i \in \mathscr{P} \qquad [\zeta_i] \tag{4.5e}$$

$$\sum_{r\in\mathscr{R}} (X_{rt} + O_{rt}) \leq n \quad \forall t \in \mathscr{T} \qquad [\phi_t] \tag{4.5f}$$

$$X_{rt}, O_{rt} \geq 0 \qquad \forall r \in \mathscr{R}, t \in \mathscr{T} \quad [\delta_{rt}, \eta_{rt}] \tag{4.5g}$$

$$\tag{4.5h}$$

We now take the dual of the inner minimization. The resulting model is then:

$$\max_{T\in Y} \quad \sum_{i\in\mathcal{N}}\sum_{d\in\mathcal{D}} c_{id}(1-T_i)\beta_{id} + \sum_{a\in\mathcal{A}}\sum_{d\in\mathcal{D}} c_{ad}\gamma_{ad} + \sum_{i\in\mathcal{P}} w_j\zeta_i + \sum_{t\in\mathcal{T}} n\phi_t \tag{4.6a}$$

$$\text{s.t.} \quad \sum_{i\in\mathcal{N}(r)} \mathbb{I}\left\{\beta_{i,d(0+\tau_{ri})}\right\} + \sum_{a\in\mathcal{A}(r)} \mathbb{I}\left\{\gamma_{a,d(0+\tau_{ra})}\right\} + \zeta_{f(r)} + \phi_0 + \delta_{r0} \le g_r \quad \forall r\in\mathcal{R} \tag{4.6b}$$

$$\alpha_{rt} + \sum_{i\in\mathcal{N}(r)} \mathbb{I}\left\{\beta_{i,d(t+\tau_{ri})}\right\} + \sum_{a\in\mathcal{A}(r)} \mathbb{I}\left\{\gamma_{a,d(t+\tau_{ra})}\right\} + \zeta_{f(r)} + \phi_t + \delta_{rt} \le g_r \quad \forall r\in\mathcal{R}, t=1,\ldots,|T| \tag{4.6c}$$

$$-\alpha_{r,1} + \phi_0 + \eta_{r0} \le ql_0 \quad \forall r\in\mathcal{R} \tag{4.6d}$$

$$\alpha_{rt} - \alpha_{r,t+\Delta} + \phi_t + \eta_{rt} \le ql_t \quad \forall r\in\mathcal{R}, t=1,\ldots,|T|-1 \tag{4.6e}$$

$$\alpha_{r|T|} + \phi_{|T|} + \eta_{r|T|} \le ql_{|T|} \quad \forall r\in\mathcal{R} \tag{4.6f}$$

$$\alpha_{rt} \le 0 \quad \forall r\in\mathcal{R}, t=1,\ldots,|T| \tag{4.6g}$$

$$\beta_{id} \le 0 \quad \forall i\in\mathcal{N}, d\in\mathcal{D} \tag{4.6h}$$

$$\gamma_{ad} \le 0 \quad \forall d\in\mathcal{D}, a\in\mathcal{A} \tag{4.6i}$$

$$\zeta_i \ge 0 \quad \forall i\in\mathcal{P} \tag{4.6j}$$

$$\phi_t \le 0 \quad \forall t\in\mathcal{T} \tag{4.6k}$$

$$\delta_{rt}, \eta_{rt} \ge 0 \quad \forall r\in\mathcal{R}, t\in\mathcal{T} \tag{4.6l}$$

where $d(t)$ is the day of time period t, $|T|$ is the last time period, $\mathbb{I}\left\{\beta_{i,d(t+\tau_{ri})}\right\} = \beta_{i,d(t+\tau_{ri})}$ if $t+\tau_{ri} \le T$ and 0 otherwise, and $\mathbb{I}\left\{\gamma_{a,d(t+\tau_{ra})}\right\} = \gamma_{a,d(t+\tau_{ra})}$ if $t+\tau_{ra} \le T$ and 0 otherwise.

Notice that when we take the dual of the inner minimization problem, it changes the inner minimization problem to a maximization problem. Thus, we eliminate the maximization sign for the inner problem. Therefore, we have a single-level model.

Also notice that our single-level model has nonlinear terms $T_i\beta_{id}$. Since these nonlinear terms are a product of a binary variable and a continuous variable, we can linearize them by applying a technique described by Sherali and Alameddine (1992). First, substitute the non-negative variable $\kappa_{id} = T_i\beta_{id}$. Then, add the constraints:

$$\kappa_{id} - \underline{\beta}_{id}T_i \ge 0 \quad \forall i\in\mathcal{N}, d\in\mathcal{D} \tag{4.7a}$$

$$\kappa_{id} - \beta_{id} \ge 0 \quad \forall i\in\mathcal{N}, d\in\mathcal{D}, \tag{4.7b}$$

with $\underline{\beta}_{id}$ denoting a lower bound of β_{id}.

Then we have the following single-level MIP:

$$\max_{T\in Y} \quad \sum_{i\in\mathcal{N}}\sum_{d\in\mathcal{D}} (c_{id}\beta_{id} - c_{id}\kappa_{id}) + \sum_{a\in\mathcal{A}}\sum_{d\in\mathcal{D}} c_{ad}\gamma_{ad} + \sum_{i\in\mathcal{P}} w_j\zeta_i + \sum_{t\in\mathcal{T}} n\phi_t \tag{4.8a}$$

$$\text{s.t.} \quad \text{(4.6b)–(4.6j)}$$

$$\text{(4.7a)–(4.7b)}$$

4.5 Case Study: Coal Transportation by Rail

4.5.1 Coal Supply Chain

Coal combustion has been commonly used to generate electricity and provide power for many kinds of operations in the United States. In 2008, it was announced that 48.2 % of the electricity consumed in the US was produced by the combustion of coal in coal power plants. The electricity generated in these plants is being used in many fields such as: hospital operations, vaccine storage, security and surveillance systems, as well as water treatment. Hence, in order to keep this source of electricity safe for such important services in case of a disruption or disaster, operations in the coal supply chain must be secured.

Rail transportation is the most popular transportation mode used to supply coal to power plants from mines. For instance, in 2008, 70% of coal were transported by rail throughout the US. On the other hand, truck and river transportations are the next two most popular modes of coal transportation with 16% and 14%, respectively. After the coal is mined, it is sent to a processing facility in where the coal pieces are crushed into more manageable chunks. The trains typically consist of 125 to 150 cars loaded with between 110-120 tons of coal in each rail car. These trains are dispatched on their routes towards associated power plants. Even though the primary objective in the coal supply chain is to meet electricity demand, reducing the transportation and storage costs of coal as much as possible also is a major consideration.

While transporting coal from mines to power plants, many important constraints are observed in the coal supply chain. The amount of coal that can be carried by a train is restricted by the size of the trains used in the system. Also, depending on the sizes of these trains, some trains can only travel on special tracks. The availability of coal in different time periods for loading/unloading operations requires extra planning.

Sub-bituminous is the most common type of coal mined in Wyoming's Powder River Basin (PRB). The PRB accounts for about 40 % of all consumption within the US. This particular coal type has significantly low SO_2 emissions and can not produce high energy output. However, many energy companies are automatically attracted by the low emissions level and abundance of supply of this type of coal.

Most power plants are designed in such a way that they can only use a single type of coal in order to generate electricity. Hence, there could be serious results of a disruption or a disaster that occurs in the coal supply chain, especially for the areas of the country that rely heavily on electricity generated from the coal mined in the PRB. In this case study, the sub-bituminous coal transportation network is used in order to test model 4.8 in Section 4.4.2. We now introduce the data source and required manipulation steps on the network in detail.

4.5.2 Network and Data Construction Process

The coal network is refined under the assumption that the small mines and plants will not make much of a difference in the large-scale coal supply chain if one or more are not able to operate. Many mines and plants only produce or consume a very small fraction of the amount required for the larger ones. There are also a great number of smaller mines that operate in close proximity to a larger mine which may be represented by a single node. The coal from these smaller mines is almost certainly going to travel on the same railroad as the coal from the larger mine(s) nearby. Hence, many smaller coal plants are also clustered together around a larger one.

Through focusing on large suppliers and consumers of coal, the refined network provides an initial representative model of the entire coal supply chain. The coal plants and coal mines are retrieved from the USCA data. The average production rate for each mine as well as the average consumption rate for each plant is analyzed. It is found that the average consumption rate for each mine is approximately 2,000 tons, while the average production rate over all of the plants is 1,400 tons. There were a great number of mines and plants to consider based upon these averages. Therefore, the network is narrowed down by using the constraint that the average production and consumption rate at each mine or plant has to be greater than 5,000 tons. This value is chosen as a threshold for network reduction. Accordingly, there were 39 mines

and 49 plants that fit the constraint. Then the mines and plants are found in which sub-bituminous coal is being mined and burned, respectively. Out of the 49 originally selected plants, the sub-bituminous plants burn 57% of the total consumption. Also, the production at the sub-bituminous mines account for 76 % of the total amount produced across all the chosen 39 mines.

Based on the rail network data obtained through the steps described above, we labeled 8 sub-bituminous coal mines and 23 sub-bituminous coal plants together with 37 rail yards as "real nodes". Overall, it is observed that there are 135,655 nodes (incluiding "real nodes", tunnels, bridges and other connection points) and 172,888 arcs (connecting nodes together) in the network. Hence, in order to generate a connected manageable rail network based on this dataset, trimming algorithm 1 is developed. Following notation introduced below includes some definitions of the terms used in trimming algorithm 1.

- M =set of coal mines
- P =set of coal plants
- R =set of railyards
- $\mathscr{N}$ =set of nodes "Real Nodes"
- B =set of elementary edges that have bridges (an elementary arc may or may connect two nodes)
- $\mathscr{A}$ =set of edges that connect two nodes
- n_{ij}^k= set of elementary nodes on the kth shortest path between i and j
- a_{ij}^k= set of elementary edges on the kth shortest path between i and j
- m_{ij} =the number of paths between i and j
- δ_{ij}^k =the length of the kth shortest path between i and j
- d_{ij} = length of edge (i, j)

The first two steps connect each real node to its closest real node neighbor. In the first step, the algorithm checks if any two real nodes are connected without any other real node in between. If so, the edge connecting that pair of real node is added to the edge set $\mathbb{E}$ (step 1) and for each member of this set, a dummy path is created (step 2). In step 3, If there is no other real node found in between, then the original path is preserved with its original components (rail line distances, nodes etc.).

After step 2, the real nodes become connected to one another. However, actual distances and nodes in between real nodes are still unknown. In the rest of the steps, the algorithm generates the K-th shortest paths for each node pairs connected by the arcs in edge set $\mathbb{E}$. During this stage, critical elements embedded in the paths $\in \mathbb{E}$ such as tunnels, bridges are also identified. Once a critical element is detected (i.e. a bridge with m as beginning node and n as end node) between node i and j, edges (i,m), (m,n) and (n, j) are created and added to $\mathscr{A}$. Nodes defining the critical element (m and n), are also added to $\mathscr{N}$ as well.

Algorithm 1 Procedure for constructing a network from USCA data

Set $\mathcal{N} = M \cup P \cup R$
Let $\mathbb{E} = \emptyset$ be an empty set of edges.
1. Connect each node to its closest neighbors

for each node pair $(i,j) \in \mathcal{N}$ such that $i \neq j$

IF n_{ij}^1 does not contain a node in the set $\mathcal{N}$, THEN add (i,j) to $\mathbb{E}$

2. Add dummy paths

for each arc $(i,j) \in \mathbb{E}$

add a dummy path from i to j that is composed of the single edge (i,j)
set $a_{ij}^{m_{ij}+1} = (i,j)$
set δ_{ij}^k to a large number

3. Add alternates for routes that have vulnerable elementary arcs

for each arc $(i,j) \in \mathbb{E}$
Set $k = 1$
Set interdictionCost=0
while $k \leq m_{ij}$+1
for each $(\ell,p) \in a_{ij}^k \cap B$
interdictionCost+=$f_{\ell m}$
IF $a_{ij}^k \cap B \neq \emptyset$ and interdictionCost$\leq b$
for each $(\ell,m) \in a_{ij}^k \cap B$
add (i,ℓ), (ℓ,p), and (p,j) to $\mathcal{A}$
add ℓ and p to $\mathcal{N}$
set $d_{i\ell} = \delta_{i\ell}^1$, $d_{\ell p} = \delta_{\ell p}^1$, and $d_{pj} = \delta_{pj}^1$
ELSE
add (i,j) to $\mathcal{A}$
set $d_{ij} = \delta_{ij}^k$
break from while loop
$k = k+1$

4. Compute capacity of arcs

for each arc $(i,j) \in \mathcal{A}$

RETURN the graph defined by nodes $\mathcal{N}$, edges $\mathcal{A}$, and distances $(d_{ij})_{(i,j)}$

Trimming algorithm 1 produced a connected graph with 456 nodes (8 mines, 23 plants, 37 yards, 388 beginning and end nodes of critical elements) and 36935 arcs connecting these nodes. Using this reduced network, multiple routes between coal plants and mines are generated to be used as input routes (R) for the model (4.8). K-th shortest paths between any combinations of mines and plants with $K = 3,5,7$ and 10 are calculated to obtain different sets of routes with sizes $R = 552, 920, 1288$ and 1840, respectively. In

the next section, we discuss the impacts of these networks that have different set and size of routes on the interdictor's and attacker's decisions.

4.6 Computational Results

In this section, we present results for solving the two-stage interdiction program as a single-stage MIP (see Section 4.4.2) using CPLEX. The following subsections demonstrate computational results for applying the single-level MIP model (4.8) to different networks prepared by the steps in Section 4.5.2.

4.6.1 Interdictions with Different Network Sizes and Budget Levels

There are three main costs included in the objective function of model (4.8). These cost items are total transportation (operation), total delay, and total interdiction costs. Note that we only allow certain numbers of interdictions ($b = 1,3,5,7,10,15$) with no fixed cost assigned to interdicting a specific node. In other words, regardless of which node is interdicted, it is counted as one interdiction ($f_i = 1$) and the number of remaining nodes that can be interdicted becomes $b-1$. Hence, the total cost incurred by transporting coal from mines to plants and the total cost due to delays are the two main costs that the defender wants to minimize, while the attacker wants to keep them at their maximum levels.

Figures 4.1 and 4.2 (see also Figures B.1, B.2 in APPENDIX) show total transportation and delay costs on four different networks where total number of routes in the network (R) is 552 and 920 (1288, 1840), respectively. These cost terms are also observed by running the model (4.8) in case of different cost ratios (c)[1] and attacker's budget levels (b). It is commonly seen in both Figure 4.1 and 4.2 that in all c zones ($c = 50,100,150,200$), total transportation costs do not change significantly as the number of interdicted nodes in the network is increased. However, delay costs increase dramatically compared to the total transportation cost as more nodes are interdicted. This means that as the interdictor manages to disable more routes and nodes in the network, it only yields extra delays but train transportation can be handled at similar costs in each scenario. On the other hand, total transportation cost remains level despite different levels of c and b. One expects to see similar transportation costs with different c values since c is changed only by varying the cost of delaying a train for an hour not the cost of operating a route r. However, based on the cost terms in Figures 4.1 and 4.2, we can see that neither higher budget levels nor larger c values are able to increase total transportation costs. in different networks. Model (4.8) finds similar minimum transportation costs regardless of which/how many nodes are interdicted for different c values.

Tables 4.1, 4.2 (see also Tables B.1 and B.2 in APPENDIX) include indices of nodes that are interdicted in the scenarios provided in Figures 4.1 and 4.2 (Figures B.1 and B.2), respectively. It is commonly seen in that most of the nodes interdicted with smaller budget levels are also attacked when the budget levels are increased. For instance, in Table 4.1, nodes 72, 74, and 80 are attacked when $c = 50$ and $b = 3$. These three nodes are also taken out of the network when $c = 50$ and $b = 7,10$, or 15. For the scenario with $b = 5$, two of these nodes (72, 74) are disabled by the attacker. More importantly, from these tables, we can observe the frequency of interdicting a specific single or a combinations of nodes in different networks with different values of b and c.

It is noted that model (4.8) is pretty consistent in terms of interdicting similar nodes for comparable b levels even though the number of routes in the network (R) or cost ratios (c) are different (i.e. see Tables 4.1 and 4.2). Hence, this can also be an explanation of observing same cost pattern in the same network but in different c zones. On a network with same R values and same interdictions, as c increases, we obtain higher total (or delay) costs.

[1] Cost ratio $= c = \frac{\text{cost of delay per time period}}{\text{cost of delay per unit distance}} = \frac{q}{c_r}$

Figure 4.1: Cost components with number of interdictions when R=552

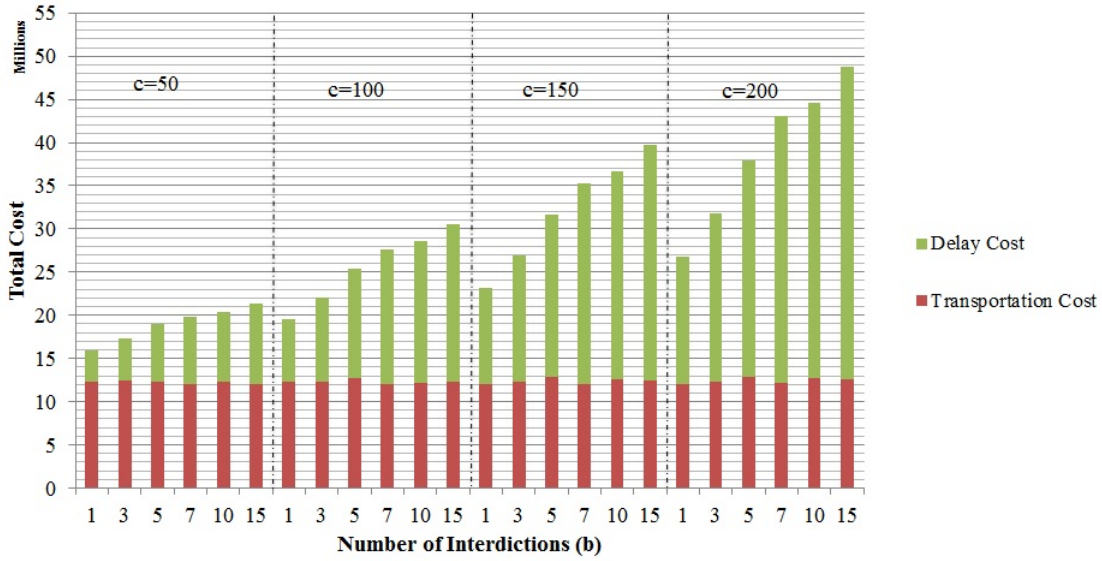
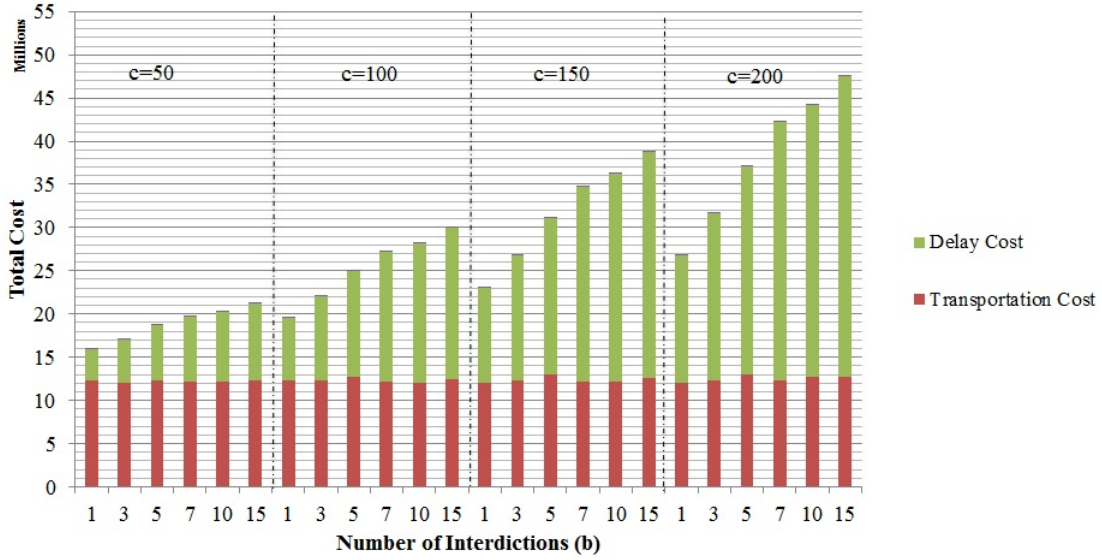


Figure 4.2: Cost components with number of interdictions when R=920



4.6.2 Solution Time and Integrality of Second Stage Relaxed IP

Figures 4.3 and 4.4 (Figures B.3 and B.4 in APPENDIX) demonstrate solution times of model (4.8) for the scenarios in Figures 4.1 and 4.2 (B.1 and B.2 in APPENDIX), respectively. It takes more time to solve model (4.8) as the network size (R) increases. On the other hand, it seems that neither the cost ratio (c) nor the budget level (B) have significant impacts on solution times for a given network. Overall, it can be seen that model (4.8) can be solved very efficiently in each scenario. Even for the network with 1840 routes, our two stage model is solved within less than 6.3 seconds.

Table 4.1: Interdicted nodes with varying c when R=552

B	c=50	c=100
1	74	74
3	72-74-80	73-81-187
5	70-72-74-79-81	70-72-74-80-187
7	68-70-72-74-78-80-187	68-70-72-74-78-80-187
10	69-70-72-74-78-80-86-88-91-187	68-70-72-74-78-80-86-88-103-187
15	51-69-70-72-74-78-80-82-84-86-88-91-93-103-187	51-68-70-72-74-78-80-83-84-86-88-91-93-102-187

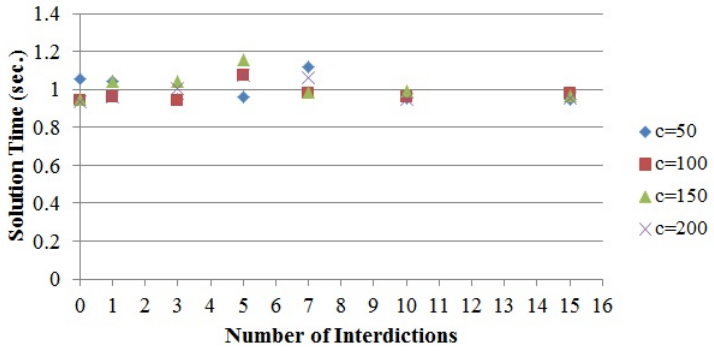
B	c=150	c=200
1	80	80
3	72-80-187	72-81-187
5	70-72-74-80-187	70-72-74-81-187
7	69-70-72-74-78-80-187	68-70-72-74-78-80-187
10	68-70-72-74-78-80-86-88-103-187	68-70-72-74-78-80-86-88-103-187
15	51-68-70-72-74-78-80-83-84-86-88-91-92-103-187	51-68-70-72-74-78-80-83-84-86-89-91-92-103-187

Table 4.2: Interdicted nodes with varying c when R=920

B	c=50	c=100
1	74	74
3	72-80-187	72-80-187
5	70-72-74-80-187	70-72-74-81-187
7	69-70-72-74-78-80-187	68-70-72-74-78-80-187
10	68-70-72-74-79-80-86-88-187-311	69-70-72-74-78-80-88-103-187-311
15	50-51-68-70-72-74-78-80-84-86-88-90-93-187-311	50-51-68-70-72-74-78-80-84-86-88-91-92-187-311

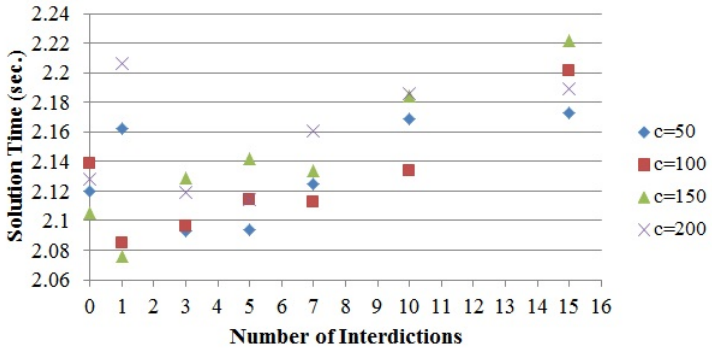
B	c=150	c=200
1	80	80
3	72-80-187	73-81-187
5	70-72-74-80-187	70-72-74-80-187
7	69-70-72-74-78-80-187	69-70-72-74-78-80-187
10	68-70-72-74-78-80-88-103-187-311	68-70-72-74-78-80-88-103-187-310
15	50-51-69-70-72-74-78-80-84-86-88-90-93-187-311	50-51-69-70-72-74-78-80-84-86-88-90-93-187-310

Figure 4.3: Solution times of model (4.8) when R=552



Solution times increase gradually as the interdictor's budget (b) increases in Figure 4.4. In contrast to this increase, solution times are observed to be stable around 1 second when the number of available routes (R) is 552 in Figure 4.3. However, for a given budget level, there is not any pattern observed in between solution times and c levels in neither Figure 4.3 nor Figure 4.4.

Figure 4.4: Solution times of model (4.8) when R=920



Certainly, one of the most important factors in solving model (4.8) very efficiently is the relaxed second stage integer problem. Note that we relaxed decision variables X_{rt} and O_{rt} first and took the dual to obtain a single level MIP. In theory, it is possible that the solution of MIP can have fractional X_{rt} and O_{rt} values. However, our experiments demonstrated that all X_{rt} and O_{rt} variables are observed to be integers. Proving that the constraints of our second stage pure IP construct the convex hull of the solution space is left as a future work.

4.6.3 Rerouting decisions after interdiction(s)

In previous sections, we demonstrated how total transportation and delay costs respond due to varying b, c, and R. In this section, we introduce a Google Maps-based toll that displays the routes along which the unit trains are moving (X_{rt}) and waiting (O_{rt}) at different time periods $\triangle$. As it can be seen in Figure 4.5, one can select the number of routes (R), number of interdictions allowed (b) and the specific time period ($\triangle$) as input parameters. Then, the solution of model (4.8) (i.e. the values of X_{rt} and O_{rt} decision variables) is displayed on a map.

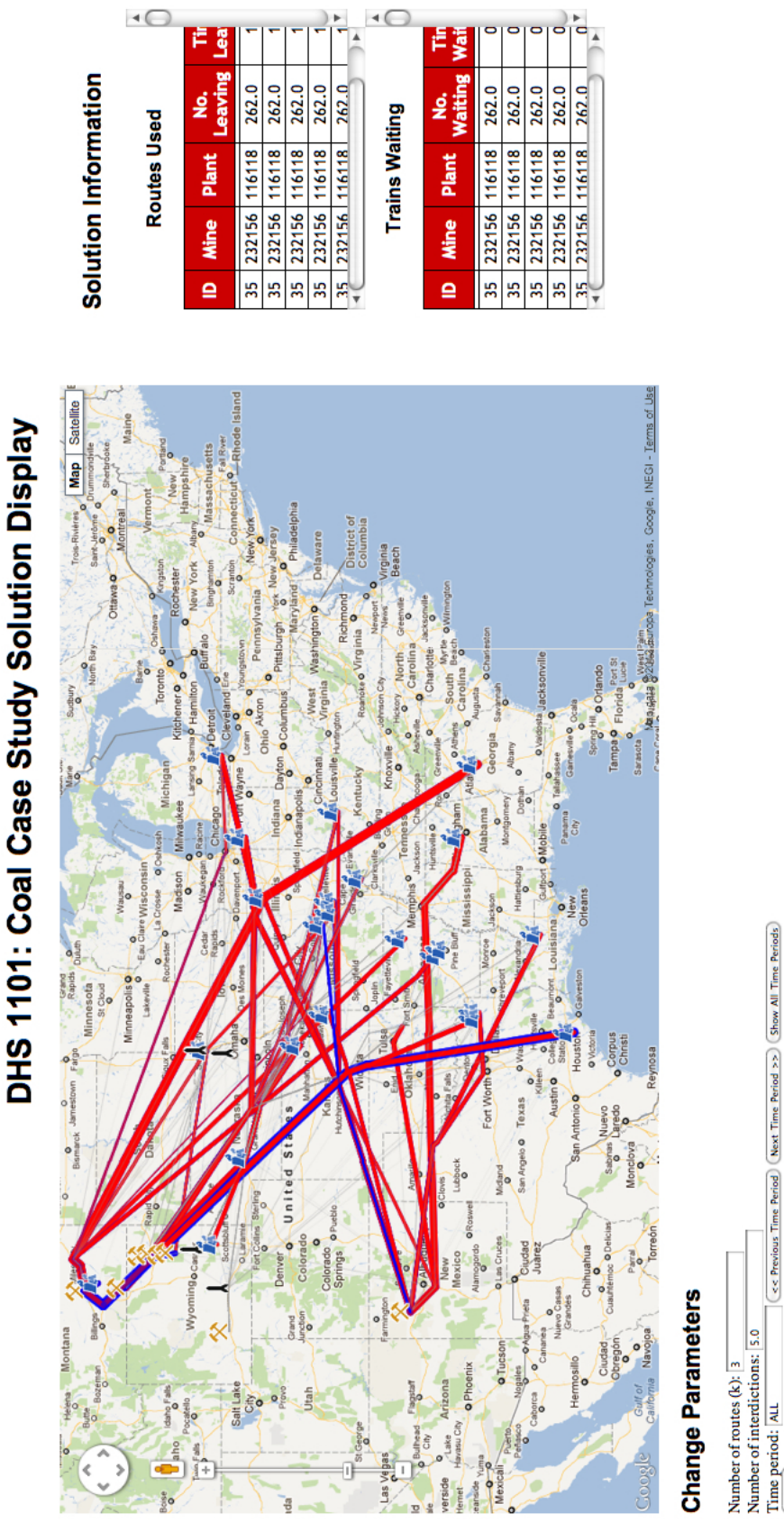
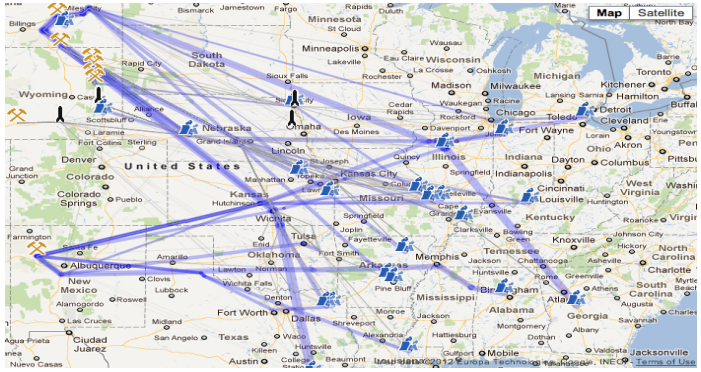
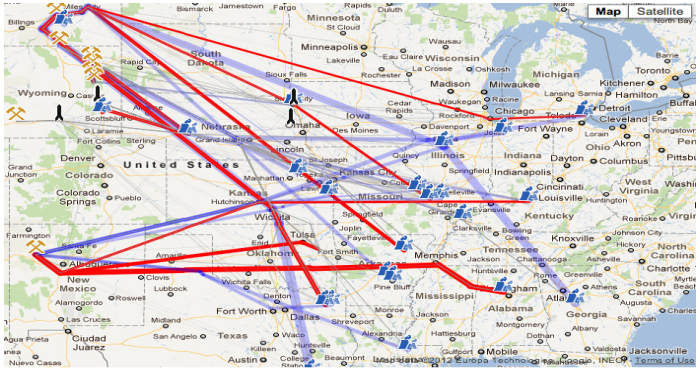


Figure 4.5: Web based Network Interdiction and Resilience Visualization Tool

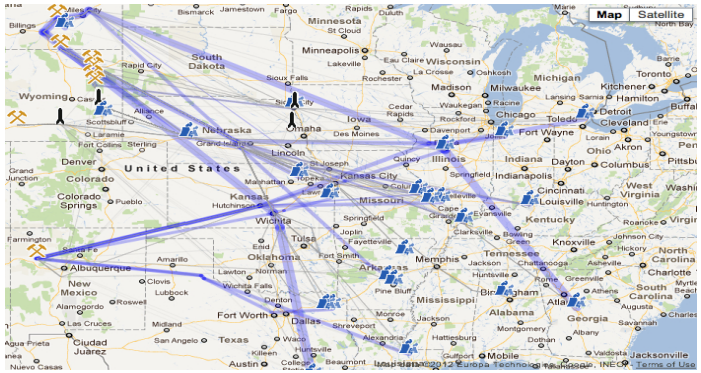
Figure 4.5 shows mines, plants, and interdicted nodes, as well as the routes along which the trains are dispatched and delays occurred due to interdictions. In order to display the delays and movements of trains clearly, straight lines are used to draw the routes between mines and plants. However, in reality, those straight lines stand for shortest paths between selected origin and destination nodes. Moreover, the more frequently a route is being used, the thicker the red straight line becomes to represent the intensity of the route. Similarly, the more frequently delays occur on a route, the thicker the blue straight line is drawn to highlight the intensity of the delays on that route. The blue and red routes seen on Figure 4.5 demonstrate all train delays and dispatches over the planning period.



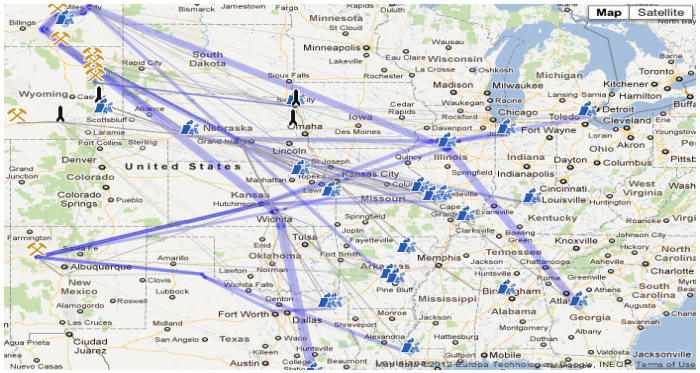
(a) O_{rt} and X_{rt} at $\triangle$=0



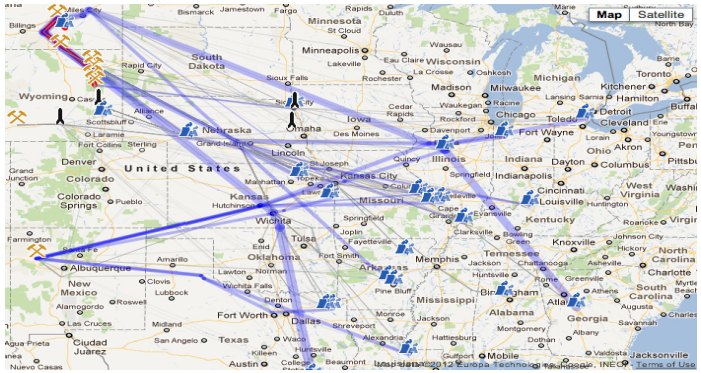
(b) O_{rt} and X_{rt} at $\triangle = 5$



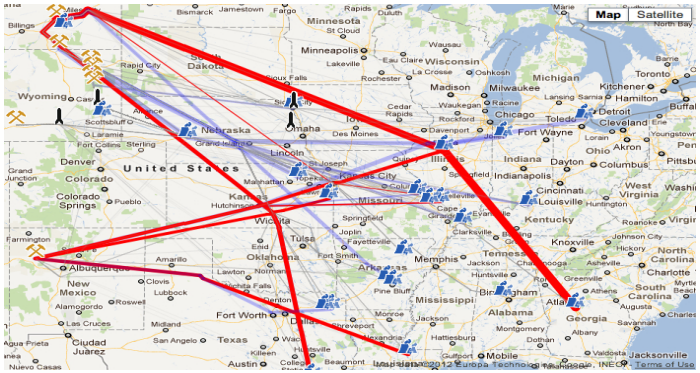
(c) O_{rt} and X_{rt} at $\triangle = 6$



(d) O_{rt} and X_{rt} at $\triangle = 8$

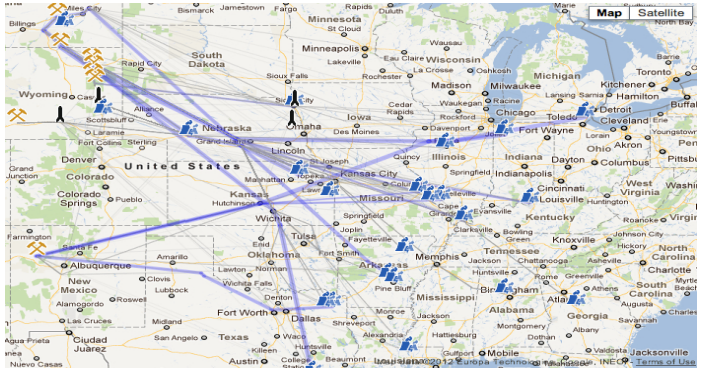


(e) O_{rt} and X_{rt} at $\triangle = 9$

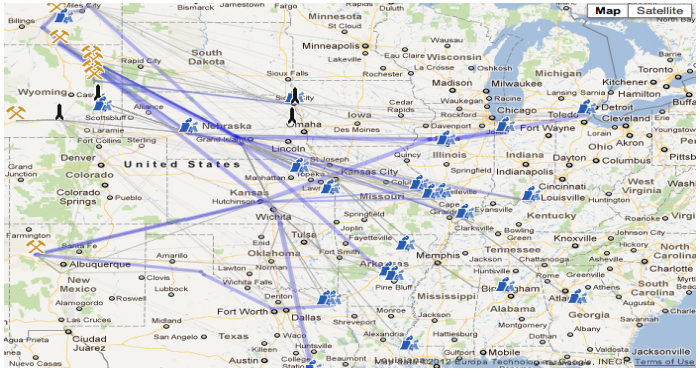


(f) O_{rt} and X_{rt} at $\triangle = 10$

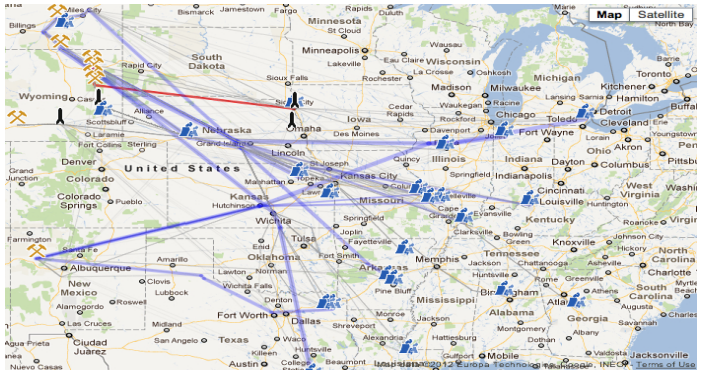
Figure 4.6: Number of trains waiting and/or moving on routes when $R = 552$ and $b = 5$



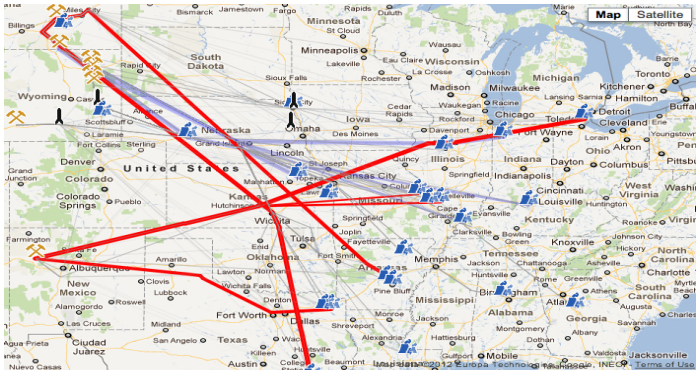
(g) O_{rt} and X_{rt} at $\triangle = 11$



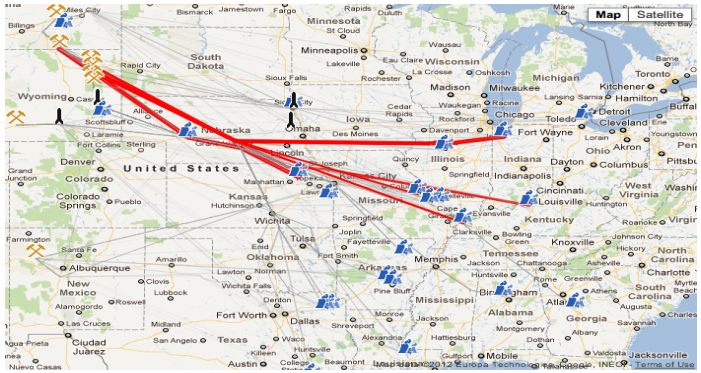
(h) O_{rt} and X_{rt} at $\triangle = 13$



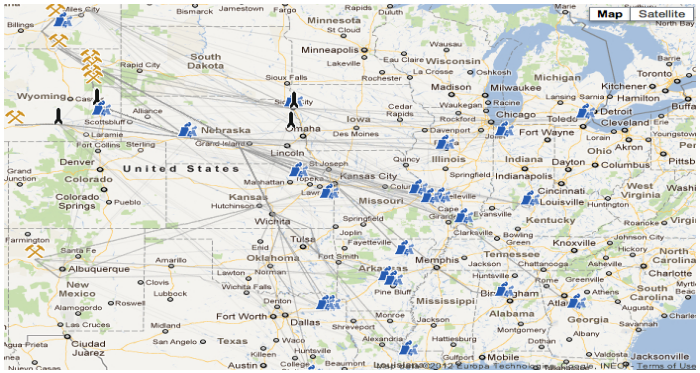
(i) O_{rt} and X_{rt} at $\triangle = 14$



(j) O_{rt} and X_{rt} at $\triangle = 15$



(k) O_{rt} and X_{rt} at $\triangle = 16$



(l) O_{rt} and X_{rt} at $\triangle = 19$

Figure 4.6: Number of trains waiting and/or moving on routes when $R = 552$ and $b = 5$

A very detailed demonstration of the train movements and delays for different time periods can be seen in Figure 4.6. Finally, gray routes represent the routes that are not being used or there is no delay is being occurred on at that time period.

4.7 Future Work

4.7.1 Congestion: Impacts of capacities

In Section 4.6.1, it is observed that the impact of attacker's budget (b) has a negligible impact on total transportation cost. On the other hand, significant increases are observed in total delay cost as the budget increases in all scenarios even with equal c values as demonstrated in Figures 4.1 and 4.2 (Figures B.1 and B.2 in APPENDIX). Being able to measure the capacity of each arc and node allows one to assess the impacts of interdictions more precisely. Note that arc and node capacity constraints (4.5c) and (4.5d) are already able to assess the capacity of nodes and arcs in terms of unit trains for a given specific amount of time. For this coal case study, it is assumed that once a node is interdicted, then the capacity of incoming and outgoing arcs is also set to zero, as well as the capacity of the node itself. However, in some situations it is possible that same interdiction might affect the capacity of other non-interdicted nodes and arcs as well (huge explosion, flood etc.). Hence, when the impact of interdiction on other non-interdicted nodes or arcs is known, it is easy to make adjustments to their capacity levels with the help of constraints (4.5c) and (4.5d). Moreover, the same capacity reduction technique can be employed when there is another commodity being transported through the same tracks and rail yards. In such circumstances, unit train capacity constraints (4.5c) and (4.5d) should be adjusted so that the impacts of congestion can be reflected in the model.

4.7.2 IP Second Stage Formulation with Empty Trains

The second stage interdiction model in Section 4.3.2 includes the flow of unit trains in only one direction from mines to plants. However, the reverse flow of empty trains from plants to mines can burden on planners since scheduling and routing those trains also requires time and resources. Hence, in this section, another second stage IP formulation is provided as an alternative to model 4.1 which accounts for the flow of empty trains as well. Even though the empty train model formulation has not been implemented, the necessary changes and additions are addressed in this section to incorporate the unit train flows from plants to mines. Let D_{rt} be the number of trains waiting (or being loaded/unloaded) at $h(r)$ of route $r \in R$ in time period $t \in T$.

$$\ell(\hat{T}) = \min \sum_{r\in R}\sum_{t\in T} c_r X_{rt} + \sum_{r\in R}\sum_{t\in T} q l_t (O_{rt} + D_{rt}) \tag{4.9a}$$

$$\text{s.t.} \quad O_{rt} + X_{rt} \le O_{r,t-\triangle} \qquad \forall r \in R\, t = \Delta, \Delta+1, \ldots, T \tag{4.9b}$$

$$D_{r,t-\triangle} + X_{r,t-\tau_r} \ge D_{rt} \qquad \forall r \in R\, t = \Delta, \Delta+1, \ldots, T \tag{4.9c}$$

$$D_{r,t-\triangle} + X_{r,t-\tau_r} \le D_{rt} + \sum_{k|o(k)=h(r)} (X_{kt} + O_{kt}) \quad \forall r \in R\, t = \Delta, \Delta+1, \ldots, T \tag{4.9d}$$

$$\sum_{r|i=o(r)} O_{r,t-\triangle} + \sum_{k|i=h(k)} (D_{k,t-\triangle} + X_{k,t-\tau_k}) = \sum_{r|i=o(r)} (O_{rt} + X_{rt}) + \sum_{k|i=h(k)} D_{kt} \quad \forall i \in N, t \in T \tag{4.9e}$$

$$\sum_{r|i=o(r)} O_{rt} + \sum_{r|i=h(r)} D_{rt} \le H_i \qquad \forall i \in N, t \in T \tag{4.9f}$$

$$\sum_{r|i=h(r)} D_{rt} \le L_i \qquad \forall i \in N, t \in T \tag{4.9g}$$

$$X_{r,t-\tau_r} + \sum_{t'=t+1}^{t+U_{it}} X_{r,t'-\tau_r} \le D_{r,t+U_{it}} \qquad \forall t \in T, \forall r \in R, i = h(r) \tag{4.9h}$$

$$\sum_{t\in T(d)} \sum_{r|i\in RN(r)} X_{r,t-\tau_{ri}} \le TC_{id} \qquad \forall d \in D\, i \in N \tag{4.9i}$$

$$\sum_{t\in T(d)} \sum_{r|a\in RT(r)} X_{r,t-\tau_{ra}} \le TC_{ad} \qquad \forall d \in D\, a \in A \tag{4.9j}$$

$$X_{rt} \le 1 - \hat{z}_i \quad \forall i \in N, r|i \in RN(r), t \in \mathscr{T} \tag{4.9k}$$

$$X_{rt} \le 1 - \hat{z}_a \quad \forall a \in A, r|a \in RT(r), t \in \mathscr{T} \tag{4.9l}$$

$$\sum_{t\in T} \sum_{r|i=h(r)} X_{r,t} = h_i \qquad \forall i \in P \tag{4.9m}$$

$$\sum_{r\in R} (X_{rt} + O_{rt} + D_{rt}) \le n \quad \forall r \in \mathscr{R}, t \in \mathscr{T} \tag{4.9n}$$

$$X_{rt}, O_{rt}, D_{rt} \in \mathbb{Z}^+ \qquad \forall r \in R, t \in \mathscr{T} \tag{4.9o}$$

Objective function (4.9a) is the total cost incurred by the total distance traveled and the total delay incurred when trains have to wait at their origin and destination stations. Constraint set (4.9b) balances flow at the origin station of route r. For each planning period $(\triangle)$ and route (r), number of trains waiting at the origin node and departing the origin node to travel on route r at time period t can not be greater than number of trains available at the origin node Δ time units before, at time period $(t - \triangle)$. Similarly, constraint set (4.9c) balances flow at the destination node of route r. Based on a node wether a origin or detination, constraints (4.9d) assure the assignment of the loaded/unloaded train to a route r. Constraints (4.9e) ensure that the total number of incoming trains, $(O_{r,t-\triangle} + D_{k,t-\triangle} + X_{k,t-\triangle})$, must be equal to the number of outgoing trains, $(O_{rt} + D_{kt} + X_{rt})$. Constraints (4.9f) guarantee that the number of trains waiting at node $i \in N$ can not exceed the corresponding holding capacity. Moreover, constraint set (4.9g) makes sure that loading/unloading operation time can not exceed the available capacity of node i, L_i.

Constraints (4.9h) demonstrates that an loaded (or unloaded) train stays at the destination until it is completely unloaded (or loaded). Constraint sets (4.9i) and (4.9j) guarantee that the total number of trains sent on a track segment and through a node in any given day does not exceed the daily track segment and node capacity, respectively. If a failure has occurred on node $i \in RN(r)$ or arc $a \in RT(r)$ on a route r, constraints (4.9k) and(4.2h) assure that trains are not dispacthed through route r for each time period. Constraints (4.9m) state that the demand of each plant should be satisfied in the planning period. Finally, constraints (4.9n) ensure that number of rains waiting at the origin node and departing that node must be less than or equal to n for each route and time period.

4.8 Summary & Final Remarks

In this study, we explored the vulnerable infrastructure elements in a unit train rail transportation network. We began by describing the problem elements and boundaries in Section 4.1. Section 4.2 discusses the commonly encountered model formulations in the literature where the vulnerability of the network is the point of interest. The steps of developing a dynamic and time-indexed consequence estimation model in case of a disruption are explained in Section 4.3. In Section 4.4, possible solution methodologies are developed and their applicabilities to the coal case supply chain are discussed. The basic properties of coal supply chain are introduced in Section 4.5. In addition, the dataset and further modifications on it are explained as well. Finally, the two stage interdiction model is solved under different circumstances and the results of the computational experiments are reported in Section 4.6.

First phase of identifying critical elements in the rail network was handled with the Algorithm 1. All critical components of the network (i.e. tunnels, bridges) included in the $K-th$ shortest path algorithms are added to the reduced network that was used to test the two stage interdiction model. Our model captures the movement of unit trains in time and space over a finite time horizon and identifies the critical nodes in the network whose unavailability causes the largest destruction in terms of total operation and delay costs. The frequency of node interdictions (see Tables 4.1, 4.2 ,B.1 and B.2) and their impacts on objectives (see Figures 4.1 and 4.2 (B.1, B.2) for different scenarios are demonstrated in Section 4.6. It is also shown that our model can be solved very efficiently for a single scenario. Note that it takes at most 6 seconds to produce results for a single scenario by solving our model. Hence, multiple scenarios are tested and solutions are provided via a Web tool called "Web based Network Interdiction and Resilience Visualization Tool" (see Figure 4.5). Therefore, for each scenario, we are able to demonstrate the values of our decision variables on a map: (i) number of trains waiting for departure for route r in time period t, (O_{rt}) (ii) number of trains moving on route r in time period t, X_{rt}.

Chapter 5

Conclusions and Future Work

This report summarizes the findings of the DHS 1101 project titled "Designing Resilient and Sustainable Supply Networks". This project focus on mitigating the effects of disruptions in networks. In particular, the impact of disruptions was measured in terms of the increase in travel distance/cost as a result of the disruption. Three main project deliverables were discussed in this report: a survey paper on networks subject to disruptions (Chapter 2), a series of papers on facilities subject to disruptions (Chapter 3), and a case study on vulnerabilities in the coal supply chain (Chapter 4).

The survey paper in Chapter 2, which was published in the international journal of risk assessment and management, reviewed the literature on mitigating against disruptions in networks. This survey reviewed papers on disruptions in a variety of network types, showing the similarities in models from different types. The survey reviewed papers on both design and risk-reduction decisions and on both random and intentional failures. A classification was produces that organizes a broad collection of literature.

In Chapter 3, a series of papers, which are currently in the peer-review process, were presented that study the problem of locating and protecting facilities subject to disruptions. First, a model called the r-all-neighbor p-center problem (RANPCP) was presented for locating facilities in order to mitigate against the worst-case disruption scenario. An MIP formulation was develop and, rather than solving the MIP formulation with available branch-and-bound codes, we solved the problem using a binary search algorithm. Experimenting with the RANPCP model led to two main insights:

1. It is better to locate facilities in anticipation of a disruption and be wrong than to locate facilities without considering disruptions and be wrong.

2. Large decreases in vulnerability can be obtained with small increases in cost.

Next, the RANPCP model was extended to include hardening decisions, made simultaneously with the location decisions. In this integrated model, termed the minimax facility location and hardening problem (MFLHP), hardened facilities are immune to failures. Again, we presented an MIP formulation for the MFLHP and solved it using a binary search algorithm. Experimenting with the MFLHP model also led generate two main insights:

1. Hardening helps make networks less vulnerable.

2. Integrating location and hardening decisions helps make networks more resilient

Finally, in Chapter 4, a case study of vulnerabilities in the coal supply chain was presented. We developed a model for identifying the most critical infrastructure elements in the portion of the US rail network that transports subituminous coal. In the first stage of our model, an attacker chooses a limited number of infrastructure elements to destroy. In the second stage, trains are routing and scheduled to minimize the total transportation cost, including travel cost and delay cost, through the residual network. This model demonstrated that as the budget of attacker increases, the portion of total cost due to train delays increases as well. Whereas, our two stage interdiction model is able to find same level of total operation cost even

though the number of interdictions increases. In addition to this, with the help of our Web tool, train movements and delays on routes in the network can be easily monitored. Hence, frequency of the attacks to critical elements is demonstrated.

The work presented in the report has also inspired promising areas for future work. First, the models presented in this report could be extended to model network design and protection decisions that takes place over time. This is more representative of reality, where networks are built and protected gradually over time. Also, it would be valuable for researchers to develop models that include both random failures and intentional attacks. In these models, a decision maker could resources to protect against random failures, intentional attackers, and both random failures and intentional attacks (i.e., all-hazards protection). Finally, more models that are able to tradeoff between multiple objectives such as efficiency and risk are needed.

Bibliography

Abril, M., F. Barber, L. Ingolotti, M. Salido, P. Tormos, and A. Lova. An assessment of railway capacity. *Transportation Research Part E: Logistics and Transportation Review*, 44(5):774–806, 2008.

Ahuja, R. K., C. B. Cunha, and G. Sahin. *Network models in railroad planning and scheduling*, volume 1, pages 54–101. 2005.

Ahuja, R. K., T. L. Magnanti, and J. B. Orlin. *Network Flows: Theory, Algorithms, and Applications*. Prentice Hall, 1 edition, 1993.

Aksen, D., N. Aras, and N. Piyade. A Bilevel P-Median Model for the Planning and Protection of Critical Facilities. *Journal of Heuristics*, pages 1–26, 2011.

Aksen, D., N. Piyade, and N. Aras. The budget constrained r-interdiction median problem with capacity expansion. *Central European Journal of Operations Research*, pages 1–23, 2009.

Albert, R., H. Jeong, and A.-L. Barabasi. Error and attack tolerance of complex networks. *Nature*, 406(6794):378–382, 2000.

Altay, N. and W. G. Green. OR/MS research in disaster operations management. *European Journal of Operational Research*, 175(1):475–493, 2006.

Ambs, K., S. Cwilich, M. Deng, D. J. Houck, D. F. Lynch, and D. Yan. Optimizing restoration capacity in the AT&T network. *Interfaces*, 30(1), 2000.

Assad, A. Models for rail transportation. *Transportation Research Part A: General*, 14(3):205–220, 1980.

Azaiez, M. N. and V. M. Bier. Optimal resource allocation for security in reliability systems. *European Journal of Operational Research*, 181(2):773–86, 2007.

Babick, J. P. *Tri-level optimization of critical infrastructure resilience*. Master's thesis, Naval Postgraduate School, 2009.

Bailey, M. D., S. M. Shechter, and A. J. Schaefer. SPAR: stochastic programming with adversarial recourse. *Oper. Res. Lett.*, 34(3):307–315, 2006.

Balakrishnan, A., T. L. Magnanti, and J. S. Sokol. Telecommunication link restoration planning with multiple facility types. *Annals of Operations Research*, 106:127–54, 2001.

Balakrishnan, A., T. L. Magnanti, J. S. Sokol, and Y. Wang. Spare-capacity assignment for line restoration using a single-facility type. *Operations Research*, 50(4):617–635, 2002.

Barabási, A.-L. and R. Albert. Emergence of Scaling in Random Networks. *Science*, 286(5439):509–512, 1999.

Bard, J. F. *Practical bilevel optimization: algorithms and applications*. The Netherlands: Kluwer Academic Publishers, 1998.

Bayrak, H. and M. D. Bailey. Shortest path network interdiction with asymmetric information. *NETWORKS*, 52(3, Sp. Iss. SI), 2008.

Berdica, K. An introduction to road vulnerability: what has been done, is done and should be done. *Transport Policy*, 9(2):117–127, 2002.

Berman, O. and D. Krass. Facility location problems with stochastic demands and congestion. In *Facility Location: Applications and Theory*, chapter 11, pages 329–371. Springer, 2002.

Berman, O., D. Krass, and M. B. C. Menezes. Facility Reliability Issues In Network P-Median Problems: Strategic Centralization And Co-Location Effects. *Operations Research*, 55(2):332–350, 2007.

Berman, O., D. Krass, and M. B. C. Menezes. Locating Facilities in the Presence of Disruptions and Incomplete Information. *Decision Sciences*, 40(4):845–868, 2009.

Beygelzimer, A., G. Grinstein, R. Linsker, and I. Rish. Improving network robustness by edge modification. *Physica A: Statistical Mechanics and its Applications*, 357(3-4):593–612, 2005.

Bier, V., L. Cox, and M. N. Azaiez. Why Both Game Theory and Reliability Theory Are Important in Defending Infrastructure against Intelligent Attacks. In *Game Theoretic Risk Analysis of Security Threats*. Springer, 2009.

Bier, V., S. Oliveros, and L. Samuelson. Choosing what to protect: Strategic defensive allocation against an unknown attacker. *Journal of Public Economic Theory*, 9(4):563, 2007a.

Bier, V. M. Choosing what to protect. *Risk Analysis*, 27(3):607–620, 2007.

Bier, V. M. and V. Abhichandani. Optimal allocation of resources for defense of simple series and parallel systems from determined adversaries. In *Risk-based decisionmaking in water resources X: proceedings of the tenth conference, November 3-8, 2002, Santa Barbara, California*, page 59. American Society of Civil Engineers, 2003.

Bier, V. M., E. R. Gratz, N. J. Haphuriwat, W. Magua, and K. R. Wierzbicki. Methodology for Identifying Near-Optimal Interdiction Strategies for a Power Transmission System. *Reliability Engineering and System Safety*, 92(9):1155–61, 2007b.

Bier, V. M., N. Haphuriwat, J. Menoyo, R. Zimmerman, and A. M. Culpen. Optimal resource allocation for defense of targets based on differing measures of attractiveness. *Risk Analysis*, 28(3):763–770, 2008.

Bier, V. M., A. Nagaraj, and V. Abhichandani. Protection of simple series and parallel systems with components of different values. *Reliability Engineering and System Safety*, 87(3):315–323, 2005.

Birge, J. R. and F. Louveaux. *Introduction to stochastic programming*, volume 9. New York: Springer-Verlag, 1997.

Bleizeffer, D. Coal prices skyrocket. http://trib.com/news/state-and-regional/articlef6e38032-640b-5e76-b784-78f76ad211ad.html, 2006.

Boyd, J. D. Trains in the line of fire. http://www.joc.com/rail-intermodal/trains-line-fire, 2011.

Brown, G. G., W. M. Carlyle, J. O. Royset, and R. K. Wood. On the Complexity of Delaying an Adversary's Project. In R. Sharda, S. Voß, B. Golden, S. Raghavan, and E. Wasil, editors, *The Next Wave in Computing, Optimization, and Decision Technologies*, volume 29 of *Operations Research/Computer Science Interfaces Series*, chapter 1, pages 3–17. Springer US, 2005.

Brown, P. S. *Optimizing the long-term capacity expansion and protection of Iraqi oil infrastructure*. Master's thesis, Naval Postgraduate School, 2005.

Bundschuh, M., D. Klabjan, P. Pei, and D. L. Thurston. Modeling robust and reliable supply chains. *Optimization Online*, 2006.

Burdett, R. and E. Kozan. Techniques for absolute capacity determination in railways. *Transportation Research Part B: Methodological*, 40(8):616–632, 2006.

Byrka, J., A. Srinivasan, and C. Swamy. Fault-tolerant facility location: A randomized dependent LP-rounding algorithm. In F. Eisenbrand and B. Shepherd, editors, *Proceeding of the 14th Conference on Integer Programming and Combinatorial Optimization*, volume 6080 of *Lecture Notes in Computer Science*, pages 244–257. Springer-Verlag, 2010.

Cappanera, P. and M. P. Scaparra. Optimal Allocation of Protective Resources in Shortest-Path Networks. *Transportation Science*, 45(1):64–80, 2011.

Chaudhuri, S., N. Garg, and R. Ravi. The p-neighbor k-center problem. *Information Processing Letters*, 65(3):131–134, 1998.

Church, R. L. COBRA: a new formulation of the classic p-median location problem. *Annals of Operations Research*, 122(1):103–120, 2003.

Church, R. L. and R. A. Gerrard. The Multi-Level Location Set Covering Model. *Geographical Analysis*, 35(4):277–290, 2003.

Church, R. L. and M. P. Scaparra. Analysis of Facility Systems Reliability When Subject to Attack or a Natural Disaster. In A. T. Murray and T. H. Grubesic, editors, *Critical Infrastructure: Reliability and Vulnerability*, chapter 11, pages 221–241. Berlin, Germany: Springer-Verlag, 2006.

Church, R. L. and M. P. Scaparra. Protecting Critical Assets: the R-Interdiction Median Problem With Fortification. *Geographical Analysis*, 39(2):129–146, 2007.

Church, R. L., M. P. Scaparra, and R. S. Middleton. Identifying Critical Infrastructure: The Median and Covering Facility Interdiction Problems. *Annals of the Association of American Geographers*, 94(3):491–502, 2004.

Church, R. L., M. P. Scaparra, and J. R. O'Hanley. Optimizing passive protection in facility systems. Technical report, ISOLDE X, Spain, 2005.

Cooper, L. Location-allocation problems. *Operations Research*, 11(3):331–343, 1963.

Corley and H. Chang. Finding the n most vital nodes in a flow network. *Management Science*, 21(3):362–4, 1974.

Cormican, K., D. P. Morton, and R. K. Wood. Stochastic network interdiction. *Operations Research*, 46(2):184–197, 1998.

Costa, L. D. Reinforcing the resilience of complex networks. *Phys. Rev. E*, 69(6), 2004.

Crainic, T. G. Service network design in freight transportation. *European Journal of Operational Research*, 122(2):272–288, 2000.

Cui, T., Y. Ouyang, and Z.-J. M. Shen. Reliable Facility Location Under the Risk of Disruptions. *Operations Research*, 58(4-Part-1):998–1011, 2011.

Cunningham, W. H. Optimal attack and reinforcement of a network. *Journal of the Association for Computing Machinery*, 32(3):549–61, 1985.

Daganzo, C. F. 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, 1984a.

Daganzo, C. F. The length of tours in zones of different shapes. *Transportation Research Part B*, 18B(2):135–45, 1984b.

Daganzo, C. F. and G. F. Newell. Configuration of physical distribution networks. *Networks*, 16(2):113–22, 1986.

Daskin, M., K. Hogan, and C. Revelle. Integration of Multiple, Excess, Backup, and Expected Covering Models. *Environment And Planning B*, 15(1), 1988.

Daskin, M. S. Application of an expected covering model to emergency medical service system design. *Decision Sciences*, 13(3):416–39, 1982.

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

Daskin, M. S. *Network and Discrete Location: Models, Algorithms, and Applications*. Hoboken, New Jersey: John Wiley and Sons, 1995.

Daskin, M. S., C. R. Coullard, and Z. J. M. Shen. An inventory-location model: Formulation, solution algorithm and computational results. *Annals of Operations Research*, 110(1):83–106, 2002.

Dighe, N., J. Zhuang, and V. M. Bier. Secrecy in defensive allocations as a strategy for achieving more cost-effective attacker deterrence. *International Journal of Performability Engineering*, 5(1):31–43, 2009.

Dong, L., K. Xing-hua, and Y. Xiang-tao. 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*, volume 3, pages 903–906. IEEE Computer Society, 2009.

Drezner, Z. Heuristic Solution Methods For Two Location Problems With Unreliable Facilities. *Journal of The Operational Research Society*, 38(6):509–514, 1987.

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

Elloumi, S., M. Labbé, and Y. Pochet. A New Formulation and Resolution Method for the P-Center Problem. *INFORMS Journal on Computing*, 16(1):84–94, 2004.

Erdos, P. and A. Renyi. On random graphs. *Publicationes Mathematicae Debrecen*, 6(290-297):156, 1959.

Ergun, O., G. Karakus, P. Keskinocak, J. Swann, and M. Villarreal. Operations Research to Improve Disaster Supply Chain Management. In J. J. Cochran, editor, *Encyclopedia of Operations Research and Management Science*. Wiley, 2010.

Francis, R. L., L. F. McGinnis, and J. A. White. *Facility layout and location: an analytical approach*. Englewood Cliffs, NJ: Prentice Hall, 1992.

Fulkerson, D. R. and G. C. Harding. Maximizing Minimum Source-Sink Path Subject To A Budget Constraint. *Mathematical Programming*, 13(1), 1977.

Gade, D. and E. A. Pohl. Sample average approximation applied to the capacitated-facilities location problem with unreliable facilities. *Journal of Risk and Reliability*, 223:259–269, 2009.

Golany, B., E. H. Kaplan, A. Marmur, and U. G. Rothblum. Nature plays with dice–terrorists do not: Allocating resources to counter strategic versus probabilistic risks. *European Journal of Operational Research*, 192(1):198–208, 2009.

Grigg, C., P. Wong, P. Albrecht, R. Allan, M. Bhavaraju, R. Billinton, Q. Chen, C. Fong, S. Haddad, S. Kuruganty, and Others. 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, 1999.

Grotschel, M., C. L. Monma, and M. Stoer. Design of survivable networks. *Handbooks in Operations Research and Management Science*, 7:617–672, 1995.

Grubesic, T. H., T. C. Matisziw, A. T. Murray, and D. Snediker. Comparative Approaches for Assessing Network Vulnerability. *International Regional Science Review*, 31(1):88–112, 2008.

Guha, S., A. Meyerson, and K. Munagala. Improved algorithms for fault tolerant facility location. pages 636–641. 2001.

Guha, S., A. Meyerson, and K. Munagala. A constant factor approximation algorithm for the fault-tolerant facility location problem. *Journal of Algorithms*, 48(2):429–440, 2003.

Haefner, L. E. The great flood of 1993: Impacts on waterborne commodity flow, rail transportation, and surrounding region. http://www.ctre.iastate.edu/pubs/semisesq/session4/haefner/index.htm, 1996.

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

Hakimi, S. L. Optimum Distribution of Switching Centers in a Communication Network and Some Related Graph Theoretic Problems. *Operations Research*, 13(3):462–475, 1965.

Hausken, K. Probabilistic Risk Analysis and game theory. *Risk Analysis*, 22(1):17–27, 2002.

Hausken, K. Protecting infrastructures from strategic attackers. In *Proceedings of the 2007 European Safety and Reliability Conference*, volume 1, pages 881–887. Stavanger, Norway, 2007.

Hausken, K. Strategic defense and attack for series and parallel reliability systems. *European Journal of Operational Research*, 186(2), 2008a.

Hausken, K. Strategic defense and attack for reliability systems. *Reliability Engineering and System Safety*, 93(11):1740–1750, 2008b.

Hausken, K., V. Bier, and J. Zhuang. Defending against Terrorism, Natural Disaster, and All Hazards. *Game Theoretic Risk Analysis of Security Threats, Springer, New York*, pages 65–97, 2009.

Hausken, K. and G. Levitin. Protection vs. false targets in series systems. *Reliability Engineering and System Safety*, 94(5):973–81, 2009a.

Hemmecke, R., R. Schultz, and D. L. Woodruff. Interdicting Stochastic Networks With Binary Interdiction Effort. In D. L. Woodruff, editor, *Network Interdiction and Stochastic Integer Programming*, volume 22, pages 69–84. Boston: Kluwer, 2003.

Hochbaum, D. S. and D. B. Shmoys. A Unified Approach to Approximation Algorithms for Bottleneck Problems. *Journal of the ACM*, 33(3):533–550, 1986.

Holmgren, A. J., E. Jenelius, and J. Westin. Evaluating Strategies for Defending Electric Power Networks Against Antagonistic Attacks. *IEEE Transactions on Power Systems*, 22(1), 2007.

Horst, R., P. M. Pardalos, and N. Van Thoai. *Introduction to global optimization*. Springer, 2000.

Israeli, E. and R. K. Wood. Shortest-Path Network Interdiction. *Networks*, 40(2):97–111, 2002.

Janjarassuk, U. and J. Linderoth. Reformulation and sampling to solve a stochastic network interdiction problem. *Networks*, 52(3):120–32, 2008.

Jenelius, E., J. Westin, and Holmgren. Critical infrastructure protection under imperfect attacker perception. *International Journal of Critical Infrastructure Protection*, 3(1):16–26, 2010.

Jeon, H. M. *Location-inventory models with supply disruptions*. Ph.D. thesis, Lehigh University, 2008.

Jespersen-Groth, J., D. Potthoff, J. Clausen, D. Huisman, L. Kroon, G. Maróti, and M. Nielsen. Disruption Management in Passenger Railway Transportation. In R. Ahuja, R. Möhring, and C. Zaroliagis, editors, *Robust and Online Large-Scale Optimization*, volume 5868 of *Lecture Notes in Computer Science*, chapter 18, pages 399–421. Berlin, Heidelberg: Springer Berlin / Heidelberg, 2009.

Jodlbauer, H. and K. Altendorfer. Trade-off between capacity invested and inventory needed. *european Journal of Operational Research*, 203(1), 2010.

Kerivin, H. and A. R. Mahjoub. Design of Survivable Networks: A survey. *Networks*, 46(1):1–21, 2005.

Khuller, S., R. Pless, and Y. J. Sussmann. Fault Tolerant K-Center Problems. *Theoretical Computer Science*, 242(1-2):237–245, 2000.

Kohl, N., A. Larsen, J. Larsen, A. Ross, and S. Tiourine. Airline disruption management: Perspectives, experiences and outlook. *Journal of Air Transport Management*, 13(3):149–162, 2007.

Kozan, E. and R. Burdett. A railway capacity determination model and rail access charging methodologies. *Transportation Planning and Technology*, 28(1):27–45, 2005.

Krumke, S. O. On a Generalization of the P-Center Problem. *Information Processing Letters*, 56(2):67–71, 1995.

Kuo, W., V. R. Prasad, F. A. Tillman, and C. L. Hwang. *Optimal reliability design: fundamentals and applications*. Cambridge Univ Pr, 2000.

Laporte, G., J. A. Mesa, and F. Perea. A game theoretic framework for the robust railway transit network design problem. *Transportation Research Part B: Methodological*, 44(4):447–459, 2010.

Larson, R. C. and A. R. Odoni. *Urban operations research*. Prentice Hall, 1981.

Lawley, M., V. Parmeshwaran, J. Richard, A. Turkcan, M. Dalal, and D. Ramcharan. A time-space scheduling model for optimizing recurring bulk railcar deliveries. *Transportation Research Part B: Methodological*, 42(5):438–454, 2008.

Lazoff, D. M. and A. B. Stephens. Fault-tolerant replication in networks with asynchronouscommunication link failures. In *Proceedings of the IEEE International Performance, Computing, and Communications Conference*, pages 131–136. 1997.

Lee, S.-D. On solving unreliable planar location problems. *Computers and Operations Research*, 28(4):329–44, 2001.

Levitin, G. and K. Hausken. Protection vs. redundancy in homogeneous parallel systems. *Reliability Engineering and System Safety*, 93(10):1444–51, 2008.

Levitin, G. and K. Hausken. False targets efficiency in defense strategy. *European Journal of Operational Research*, 194(1):155–162, 2009.

Levitin, G. and K. Hausken. Meeting a demand vs. enhancing protections in homogeneous parallel systems. *Reliability Engineering and System Safety*, 94(11), 2009a.

Levitin, G. and K. Hausken. False targets vs. redundancy in homogeneous parallel systems. *Reliability Engineering and System Safety*, 94(2):588–95, 2009b.

Li, X. and Y. Ouyang. A continuum approximation approach to reliable facility location design under correlated probabilistic disruptions. *Transportation Research Part B*, 44(4):535–548, 2009.

Liberatore, F., M. P. Scaparra, and M. Daskin. 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, 2011.

Lim, C. and J. C. Smith. Algorithms for discrete and continuous multicommodity flow network interdiction problems. *IIE Transactions*, 39(1):15–26, 2007.

Lim, M., M. Daskin, S. Chopra, and A. Bassamboo. A Facility Reliability Problem: Formulation, Properties, and Algorithm. *Naval Research Logistics*, 57(1):58–70, 2010a.

Lim, M., M. Daskin, S. Chopra, and A. Bassamboo. Solving Interdictor-Defender Bilevel Optimization: A Two Population Genetic Algorithm Approach. Technical report, University of Illinois, 2010b.

Liu, C., Y. Fan, and F. Ordóñez. A Two-Stage Stochastic Programming Model for Transportation Network Protection. *Computers and Operations Research*, 36(5):1582–90, 2009.

Ma, Y. and H. Wu. Definitions and curve fitting of time satisfaction functions in facility location problems. In *Proceedings of the 2006 International Conference on Management Science and Engineering*, volume 1, pages 429–33. Piscataway, NJ, USA, 2006.

Mar'in, Á., J. Mesa, and F. Perea. Integrating Robust Railway Network Design and Line Planning under Failures. In R. K. Ahuja, editor, *Robust and Online Large-Scale Optimization*, pages 273–292. Springer, 2009.

Mattsson, L.-G. Railway Capacity and Train Delay Relationships. In A. T. Murray and T. H. Grubesic, editors, *Critical Infrastructure*, Advances in Spatial Science, pages 129–150. Springer Berlin Heidelberg, 2007.

Medal, H., C. Rainwater, E. Pohl, and M. Rossetti. On the R-All-Neighbor P-Center Problem. *Working paper*, 2011a.

Medal, H. R., E. A. Pohl, and M. D. Rossetti. An Integrated Model for Facility Location and Hardening. *submitted to IIE Transactions (June 2011); currently under revision; invited to resubmit to IIE Transactions*, 2011b.

Mehndiratta, S. R., D. Brand, and T. E. Parody. How transportation planners and decision makers address risk and uncertainty. *Transportation Research Record*, (1706):46–53, 2000.

Melachrinoudis, E. and M. E. Helander. A single facility location problem on a tree with unreliable edges. *Networks*, 27(3):219–237, 1996.

Morehead, R. and A. Noore. 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, 2007.

Murray, A. T., T. C. Matisziw, and T. H. Grubesic. A Methodological Overview of Network Vulnerability Analysis. *Growth and Change*, 39(4):573–592, 2008.

Nel, L. D. and C. J. Colbourn. Combining monte carlo estimates and bounds for network reliability. *Networks*, 20(3):277–298, 1990.

Nemani, A. K. and R. K. Ahuja. *OR models in freight railroad industry*. 2011.

O'Hanley, J. R., R. Church, and J. K. Gilless. Locating and Protecting Critical Reserve Sites to Minimize Expected and Worst-Case Losses. *Biological Conservation*, 134(1):130–141, 2007a.

O'Hanley, J. R. and R. L. Church. Designing Robust Coverage Networks to Hedge Against Worst-Case Facility Losses. *European Journal of Operational Research*, 209(1):23–36, 2011.

O'Hanley, J. R., R. L. Church, and Gilless. The importance of *In Situ* site loss in nature reserve selection: Balancing notions of complementarity and robustness. *Biological Conservation*, 2(1):170–180, 2007b.

Pan, F. and D. P. Morton. Minimizing a stochastic maximum-reliability path. *Networks*, 52(3):111–119, 2008.

Paul, G., T. Tanizawa, S. Havlin, and H. E. Stanley. Optimization of robustness of complex networks. *European Physical Journal B*, 38(2):187–191, 2004.

Peeta, S., F. S. Salman, D. Gunnec, and K. Viswanath. Pre-Disaster Investment Decisions for Strengthening a Highway Network. *Computers and Operations Research*, 37(10):1708–1719, 2010.

Peng, P., L. V. Snyder, Z. Liu, and A. Lim. Reliable Logistics Networks Design with Facility Disruptions. *Transportation Research-Part B*, 2011.

Peng, R., G. Levitin, M. Xie, and S. H. Ng. Defending simple series and parallel systems with imperfect false targets. *Reliability Engineering and System Safety*, 95(6):679–688, 2010.

Peterson, S. K. and R. L. Church. A Framework for Modeling Rail Transport Vulnerability. *Growth and Change*, 39(4):617–641, 2008.

Pirkul. The uncapacitated facility location problem with primary and secondary facility requirements. *IIE Transactions*, 21(4), 1989.

Powell, R. Allocating defensive resources with private information about vulnerability. *American Political Science Review*, 101(04):799–809, 2007a.

Powell, R. Defending against terrorist attacks with limited resources. *American Political Science Review*, 101(03):527–541, 2007b.

Prasad, T. D. and N. Park. Multiobjective genetic algorithms for design of water distribution networks. *Journal of Water Resources Planning and Management*, 130(1):73–82, 2004.

Qi, L., Z. J. M. Shen, and L. V. Snyder. The Effect of Supply Disruptions on Supply Chain Design Decisions. *Transportation Science*, 44(2):274–289, 2010.

Qi, X., J. F. Bard, and G. Yu. Supply chain coordination with demand disruptions. *Omega*, 32(4):301–312, 2004.

Qi, X., J. F. Bard, and G. Yu. Disruption management for machine scheduling: The case of SPT schedules. *International Journal of Production Economics*, 103(1):166–184, 2006.

Qiao, J., D. Jeong, M. Lawley, J. P. P. Richard, D. M. Abraham, and Y. Yih. Allocating Security Resources to a Water Supply Network. *IIE Transactions*, 39(1):95–109, 2007.

Quillen, E. Rail merger brings delays, derailments. http://www.hcn.org/issues/118/3771, 1997.

Rail Report: Rail Customer News and Information. Rails Cause Utility Fuel Shortages, Electricity Rate Hikes. 2005.

Ramirez-Marquez, J. E., C. M. Rocco, and G. Levitin. Optimal protection of general source-sink networks via evolutionary techniques. *Reliability Engineering and System Safety*, 94(10), 2009.

Rao, S. and T. J. Goldsby. Supply chain risks: a review and typology. *The International Journal Of Logistics Management*, 20(1):97–123, 2009.

Rocco, C. M., J. E. Ramirez-Marquez, and D. E. Salazar. Bi and tri-objective optimization in the deterministic network interdiction problem. *Reliability Engineering and System Safety*, In Press, Accepted Manuscript, 2010.

Rocco, C. M. S., D. E. A. Salazar, and J. E. Ramirez-Marquez. Multi-objective network interdiction using evolutionary algorithms. Piscataway, NJ, USA, 2009.

Rodrigues, V. S., D. Stantchev, A. Potter, M. Naim, and A. Whiteing. Establishing a transport operation focused uncertainty model for the supply chain. *International Journal of Physical Distribution & Logistics Management*, 38(5):388–411, 2008.

Royset, J. O. and R. K. Wood. Solving the bi-objective maximum-flow network-interdiction problem. *INFORMS JOURNAL ON COMPUTING*, 19(2), 2007.

Ryoo, H.-S. and N. V. Sahinidis. Global optimization of multiplicative programs. *Journal of Global Optimization*, 26(4):387–418, 2003.

San Martin, P. A. *Tri-level optimization models to defend critical infrastructure*. Master's thesis, Naval Postgraduate School, 2007.

Santiváñez, J. and E. Melachrinoudis. Location of a reliable center on a tree network. *Operational Research*, 7(3):419–445, 2008.

Santiváñez, J., E. Melachrinoudis, and M. E. Helander. Network location of a reliable center using the most reliable route policy. *Computers and Operations Research*, 36(5):1437–1460, 2009.

Santos, J. R. Inoperability input-output modeling of disruptions to interdependent economic systems. *Systems Engineering*, 9(1):20–34, 2006.

Scaparra, M. P. Optimal resource allocation for the facility protection in median systems. Working paper, University of Kent, 2006.

Scaparra, M. P. and P. Cappanera. Optimizing security investments in transportation and telecommunication networks. In *INFORMS Annual Meeting, San Francisco, CA*. 2005.

Scaparra, M. P. and R. L. Church. A Bilevel Mixed-Integer Program for Critical Infrastructure Protection Planning. *Computers and Operations Research*, 35(6):1905–1923, 2008a.

Scaparra, M. P. and R. L. Church. An Exact Solution Approach for the Interdiction Median Problem with Fortification. *European Journal of Operational Research*, 189(1):76–92, 2008b.

Schavland, J., Y. Chan, and R. A. Raines. Information security: Designing a stochastic-network for throughput and reliability. *Naval Research Logistics*, 56:625–641, 2009.

Schmitt, A. J., L. V. Snyder, and Z. J. M. Shen. Inventory systems with stochastic demand and supply: Properties and approximations. *European Journal of Operational Research*, 2008.

Shen, Z. J. M., C. Coullard, and M. S. Daskin. A joint location-inventory model. *Transportation Science*, 37(1):40–55, 2003.

Shen, Z. J. M., R. Zhan, and J. Zhang. The Reliable Facility Location Problem: Formulations, Heuristics, and Approximation Algorithms. *INFORMS Journal on Computing*, 2011.

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

Skaperdas, S. Contest success functions. *Econom. Theory*, 7(2):283–290, 1996.

Smith, J. C. Basic Interdiction Models. In J. Cochran, L. A. Cox, P. Keskinocak, J. P. Kharoufeh, and J. C. Smith, editors, *Wiley Encyclopedia of Operations Research and Management Science*. Wiley, 2011.

Smith, J. C. and C. Lim. Algorithms for Network Interdiction and Fortification Games. In A. Chinchuluun, P. M. Pardalos, A. Migdalas, and L. Pitsoulis, editors, *Pareto Optimality, Game Theory And Equilibria*, volume 17 of *Springer Optimization and Its Applications*, chapter 24, pages 609–644. New York, NY: Springer New York, 2008.

Snyder, L. V., Z. Atan, P. Peng, Y. Rong, A. J. Schmitt, and B. Sinsoysalk. OR/MS models for supply chain disruptions: A review. Technical report, Lehigh University, 2010.

Snyder, L. V. and M. S. Daskin. Reliability Models for Facility Location: The Expected Failure Cost Case. *Transportation Science*, 39(3):400–416, 2005.

Snyder, L. V. and M. S. Daskin. Models for Reliable Supply Chain Network Design. In A. T. Murray and T. H. Grubesic, editors, *Critical Infrastructure: Reliability and Vulnerability*, chapter 13, pages 257–289. Berlin, Germany: Springer-Verlag, 2007.

Snyder, L. V., M. P. Scaparra, M. S. Daskin, and R. L. Church. Planning for Disruptions in Supply Chain Networks. In M. P. Johnson, B. Norman, and N. Secomandi, editors, *TutORials in Operations Research: Models, Methods, and Applications for Innovative Decision Making*, TutORials in Operations Research, chapter 9, pages 234–257. Baltimore, MD: INFORMS, 2006.

Sullivan, J. L., L. Aultman-Hall, and D. C. Novak. 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*, 1:271–280, 2009.

Tanizawa, T., G. Paul, R. Cohen, S. Havlin, and H. E. Stanley. Optimization of network robustness to waves of targeted and random attacks. *Physical Review E*, 71(4), 2005.

The Department of Homeland Security Risk Steering Committee. DHS Risk Lexicon. Technical report, Department of Homeland Security, Washington, D.C., 2008.

Ulfarsson, G. and E. Unger. Impacts and Responses of Icelandic Aviation to the 2010 Eyjafjallajökull Volcanic Eruption. *Transportation Research Record: Journal of the Transportation Research Board*, 2214(-1):144–151, 2011.

Veerasamy, J., S. Venkatesan, and J. C. Shah. Spare capacity assignment in telecom networks using path restoration and further improvement using traffic splitting. *Journal of Systems and Software*, 47(1):27–33, 1999.

Von Stackelberg, H. and A. T. Peacock. *The theory of the market economy*. Oxford University Press, 1952.

Wallace, S. W. A piecewise linear upper bound on the network recourse function. *Math. Program. 38, 133-146*, 17(1):87–103, 1987.

Watts, D. J. and S. H. Strogatz. Collective dynamics of 'small-world' networks. *Nature*, 393(6684):440–2, 1998.

Weaver, J. R. and R. L. Church. A median location model with nonclosest facility service. *Transportation Science*, 19(1):58–74, 1985.

Wikipedia. 2010 eruptions of Eyjafjallajökull. Accessed 8 October, 2010, 2010.

Wollmer, R. Removing arcs from a network. *Operations Research*, 12(6):934–940, 1964.

Wood, R. K. Deterministic Network Interdiction. *Mathematical and Computer Modelling*, 17(2):1–18, 1993.

Xia, Y., m.-H. Yang, B. Golany, S. M. Gilbert, and G. Yu. Real-time disruption management in a two-stage production and inventory system. *IIE Transactions*, 36(2):111–125, 2004.

Xu, C. and I. C. Goulter. Reliability-based optimal design of water distribution networks. *Journal of Water Resources Planning and Management*, 125:352, 1999.

Xue, G. Linear time algorithms for computing the most reliable source on an unreliable tree network. *Networks*, 30(1):37–45, 1997.

Yao, Y., T. Edmunds, D. Papageorgiou, and R. Alvarez. Trilevel Optimization in Power Network Defense. *IEEE Transactions On Systems Man And Cybernetics Part C*, 37(4):712–718, 2007.

Yu, G. and X. Qi. *Disruption Management: Framework, Models And Applications*. World Scientific Pub Co Inc, 2004.

Zhan, R. L. *Models and Algorithms for Reliable Facility Location Problems and System Reliability Optimization*. Ph.D. thesis, University Of Florida, 2007.

Zhao, J. and K. Xu. Enhancing the robustness of scale-free networks. *Journal of Physics A: Mathematical and Theoretical*, 42:195003, 2009.

Zhu, Z., J. F. Bard, and G. Yu. Disruption management for resource-constrained project scheduling. *Journal of the Operational Research Society*, 56(4):365–81, 2005.

Zhuang, J. and V. Bier. Secrecy and Deception at Equilibrium, with Applications to Anti-Terrorism Resource Allocation. *Defence and Peace Economics*, 22(1):43–61, 2011.

Zhuang, J. and V. M. Bier. Balancing terrorism and natural disasters-defensive strategy with endogenous attacker effort. *Operations Research*, 55(5):976–991, 2007.

Zhuang, J. and V. M. Bier. Katrina vs. 9/11 How should we optimally protect against both? In H. Richardson, P. Gordon, and J. Moore, editors, *Post-Katrina: Economics, Social Aspects, and Risk*, chapter 4, pages 71–83. Edward Elgar Publishing, 2008.

Zhuang, J. and V. M. Bier. Reasons for Secrecy and Deception in Homeland-Security Resource Allocation. *Risk Analysis*, 30(12):1737–1743, 2010.

List of Deliverables

Publications

- Medal, H., Rainwater, C., Pohl, E. and Rossetti, M., *On the R-All-Neighbor P-Center Problem*, Working paper.
- Medal, H., Pohl, E., Rossetti, M., *An Integrated Model for Facility Location and Hardening*, Working paper.
- Medal, H., Sharp, S., Pohl, E., Mason, S., Rainwater, C., *Analysis of Networked Infrastructure Subject to Disruptions: A survey*, International Journal of Risk Assessment and Management. 15(2/3), 99-127.
- Ertem, M., Buyurgan, N., and Pohl, E., *Using Announcement Options in the Bid Construction Phase for Disaster Relief Procurement*, submitted to the Journal of Socio-Economic Planning Sciences special issue on Disaster Planning and Logistics. (In review)
- Ramirez-Marquez, J., Murynets, I., Medal, H., Rainwater, C., and Pohl, E., *Facility Location with Interdiction: A Multi-Objective Analysis*, Working paper.
- Medal, H., Rainwater, C., Pohl, E. and Rossetti, M., *On the R-All-Neighbor P-Center Problem*, Working paper.
- Salgado, M., Menezes, B., and Pohl, E.A., *Developing Expert Opinion Based Models for Critical Infrastructure Risk Assessment and Vulnerability Analysis*,Proceedings of the 2010 Industrial Engineering Research Conference, Cancun, Mexico, June, 2010.
- Malaviya A.K., Rainwater C., and Sharkey T.C. , *Multi-period network interdiction models with applications to city-level drug enforcement*. IIE Annual Conference (IERC) Proceedings, Cancun, Mexico, June 2010.

Presentation

- Gedik, R., Medal, H., Pohl, E., Rainwater, C., Mason, S., *Assessing Vulnerabilities in the Coal Supply Chain*, INFORMS Annual Meeting, Charlotte, NC, November 13-18, 2011.
- Medal, H., Pohl, E., Rossettie, M., *An Integrated Model for Facility Location and Hardening*, Poster Presentation, INFORMS Annual Meeting, Charlotte, NC, November 13-18, 2011.
- Medal, H., Rainwater, C., Pohl, E. and Wang, C., *Prepositioning of Supplies and Hardening for Disaster Relief*, panelist at U.S. Department of Homeland Security Fifth Annual University Network Summit Ð Student Day, March 2011.
- Medal, H., Rossetti, M. D., and Pohl, E. , *Locating Facilities for the Strategic National Stockpile*, Health and Humanitarian Logistics Conference, Atlanta, GA, February 2011.

- Medal, H., Pohl, E. A., Rainwater, C. and Rossetti, M. D., *Locating Facilities Subject to Interdiction* INFORMS Computing Society Conference on Operations Research, Computing and Homeland Defense, Monterey, CA, January 2011.
- Medal, H., *Fortifying Facilities in the Public Sector: A Risk Equitable Approach*, Institute for Operations Research and Management Science (INFORMS) Conference, Austin, TX, November 2010.
- Medal, H., *The Unreliable Hub Location Problem*, Institute for Operations Research and Management Science (INFORMS) Conference, Austin, TX, November 2010.
- Medal, H., Pohl, E. A., Rainwater, C. and Rossetti, M., *A Mathematical Model for Locating and Protecting Unreliable Facilities*, Industrial Engineering Research Conference (IERC), Reno, NV, May 2011.
- Medal, H., Sharp, S., Nguyen, H., Pohl, E., and Mason, S., *Multi-Modal Supply Chain Network Analysis Under Disruptions*, Industrial Engineering Research Conference 2010, Cancun, Mexico, June 2010.
- Medal H., Industrial Engineering Research Conference, Ph.D. Colloquium, Protecting Networks Using Branch and Price with Stochastic Cuts. Cancun, Mexico, June 2010.
- Medal, H., Rossetti, M., and Pohl, E., *Donations Management in the Humanitarian Supply Chain* Industrial Engineering Research Conference 2010, Cancun, Mexico, June 2010.
- Nguyen H.N., Salgado M.F.P., Mason S., Pohl E., *Risk and Vulnerability Analysis Techniques for Supply Chain Network Infrastructure*, IIE Annual Conference, Cancun, Mexico, June 2010.
- Madadi, A., Kurz, M., Taaffe, K., Mason, S., Pohl, E., Root, S., and Sir, M., *Managing Disruptions in Healthcare Supply Chain Networks*, Industrial Engineering Research Conference 2010, Cancun, Mexico, June 2010.
- DHS S &T Summit, Panel 22 - Transportation Resiliency; Edward A. Pohl, Modeling Transportation Resiliency, Washington D.C., 12 March 2010.
- DHS S &T Summit, Graduate Student Poster Session, Hugh Medal, Designing Resilient Supply Chain Networks, Washington D.C.,10-12 March 2010.
- Medal H., *Routing and Resource Allocation for Disaster Relief*, 2010 Health and Humanitarian Logistics Conference, Georgia Institute of Technology Atlanta, GA, March 2010.

Transportation Security Workshops / Conferences Attended

- Pohl E., Attended the Military Operations Research Society (MORS) conference on Risk Analysis, 2011.
- Medal, H., Gedik, R., Designing Resilient and Sustainable Supply Chain Networks, Poster presentation at the Mack Blackwell Rural Transportation Center Advisory Board meet, 18 November 2011.
- Rainwater C., Military Operations Research Society Workshop on Critical Infrastructure, November 15-18, 2010, Arlington, VA. .

Research Products Developed

- Web based Network Interdiction and Resilience Visualization Tool

Appendix A

Algorithms

A.1 Binary Search Algorithm

The binary search algorithm is as follows.

1. **Initialize:** Let UB and LB be initial upper and lower bounds for the integrated MFLHP. Let $D = \{\phi_{ij} : i \in \mathscr{I}, j \in \mathscr{J}, \phi_{ij} \geq LB, \phi_{ij} \leq UB\}$ 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 $index = lbIndex + \left\lceil \frac{ubIndex - lbIndex}{2} \right\rceil$.

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 $\chi(\tilde{X}^*, \tilde{Z}^*) = \sum_{i \in \mathscr{I}} f_i \tilde{X}^* + \sum_{i \in \mathscr{I}} g_i \tilde{Z}^*$ be the cost of the solution. IF $\chi(\tilde{X}^*, \tilde{Z}^*) \leq b$, set $ubIndex = index$ and return to Step 2.

5. Let $U(\tilde{X}^*, \tilde{Z}^*)$ be the post-interdiction radius for solution $(\tilde{X}^*, \tilde{Z}^*)$. IF $U(\tilde{X}^*, \tilde{Z}^*) < D_{ubIndex}$ 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.

Appendix B

Data

Figure B.1: Cost components with number of interdictions when R=1288

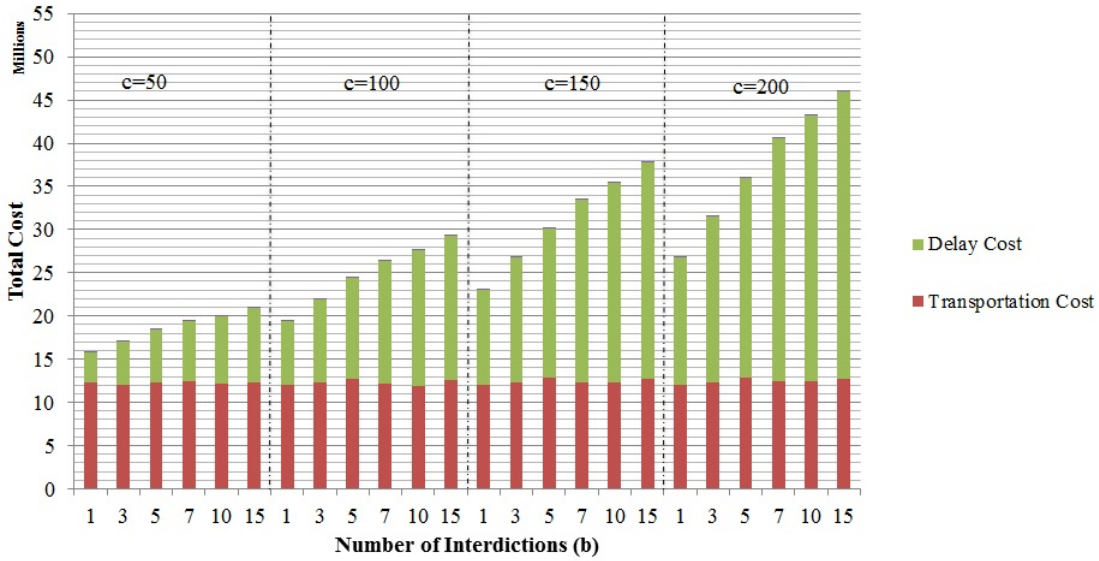


Table B.1: Interdicted nodes with varying *c* when R=1288

B	c=50	c=100
1	74	80
3	72-80-187	72-81-187
5	70-72-74-80-187	70-72-74-81-187
7	70-72-74-79-80-187-313	68-70-72-74-79-80-187
10	68-70-72-74-78-80-86-88-187-312	69-70-72-74-78-80-103-187-311-313
15	50-51-69-70-72-74-78-80-84-86-88-90-92-187-312	50-51-69-70-72-74-78-80-84-86-88-91-92-187-312

B	c=150	c=200
1	80	80
3	73-80-187	73-80-187
5	70-72-74-81-187	70-72-74-81-187
7	68-70-72-74-78-81-187	68-70-72-74-79-81-187
10	68-70-72-74-78-80-103-187-310-312	68-70-72-74-78-80-103-187-310-312
15	47-51-68-70-72-74-78-80-86-88-91-93-187-311-312	47-51-68-70-72-74-78-80-86-88-91-92-187-311-313

Figure B.2: Cost components with number of interdictions when R=1840

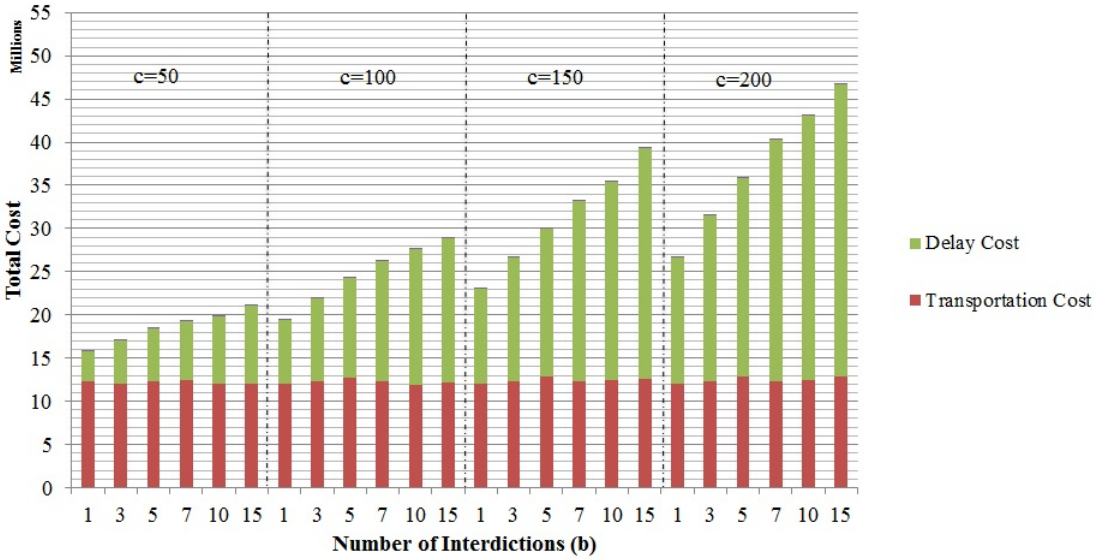


Figure B.3: Solution times of model (4.8) when R=1288

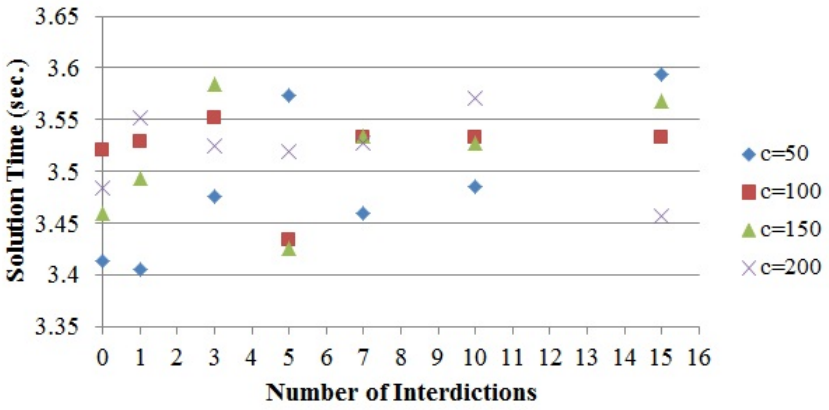


Figure B.4: Solution times of model (4.8) when R=1840

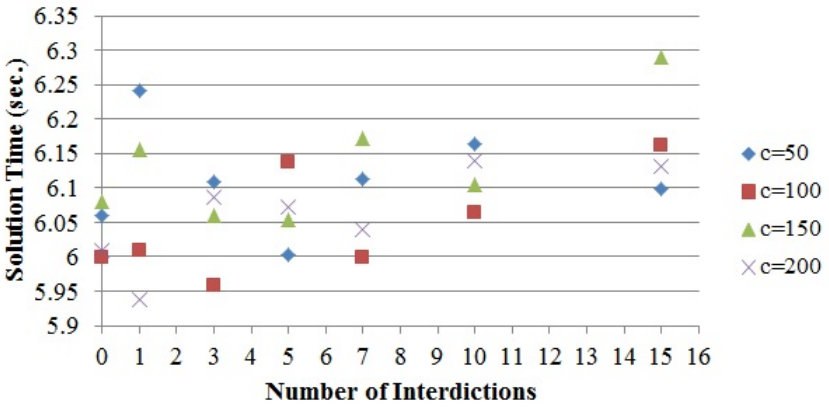


Table B.2: Interdicted nodes with varying c when R=1840

B	c=50	c=100
1	74	80
3	72-80-187	72-81-187
5	70-72-74-81-187	70-72-74-81-187
7	70-72-74-78-80-187-312	68-70-72-74-78-81-187
10	47-69-70-72-74-79-80-187-311-312	47-69-70-72-74-78-80-103-187-312
15	47-51-69-70-72-74-78-80-86-88-91-92-96-187-312	47-51-69-70-72-74-78-80-86-88-91-93-96-187-312

B	c=150	c=200
1	81	80
3	72-80-187	72-80-187
5	70-72-74-80-187	70-72-74-81-187
7	69-70-72-74-78-80-187	68-70-72-74-79-80-187
10	47-68-70-72-74-78-80-103-187-312	47-68-70-72-74-78-80-103-187-312
15	47-51-68-70-72-74-78-80-86-88-91-93-187-311-312	47-51-68-70-72-74-78-80-86-88-91-92-187-311-313