TWO-LEVEL SYSTEM INTERDICTION

By

KELLY M. SULLIVAN

A DISSERTATION PRESENTED TO THE GRADUATE SCHOOL
OF THE UNIVERSITY OF FLORIDA IN PARTIAL FULFILLMENT
OF THE REQUIREMENTS FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

UNIVERSITY OF FLORIDA

2012

To my parents, grandparents, and three wonderful sisters.

ACKNOWLEDGMENTS

First, I would like to thank my advisor, Dr. J. Cole Smith, for mentoring me during my studies at the University of Florida. He has taught me much about research and teaching and provided me with a constant source of encouragement. I thank Dr. Smith for answering every single one of my emails, texts, and calls over the last four years, no matter the time of day or night.

I am thankful for the feedback of three very helpful committee members: Dr. Joseph Hartman, Dr. Jean Philippe Richard, and Dr. My Thai. Their comments have greatly improved the research presented in this document. Thanks also to Dave Morton and Feng Pan who made significant contributions to the nuclear interdiction research presented in this dissertation.

I also owe thanks to Dr. C. Richard Cassady for his guidance and friendship as well. He taught my first engineering class, served as my first research advisor, and is the reason I am pursuing a career in the field.

Among the many friends that I have made in Gainesville, I want to express sincerest thanks to Shantih Spanton who was there for many good times and some really tough ones too. I also want to thank my dear friends Siqian Shen and Soheil Hemmati, as well as my best friend, Kim Carlson.

Finally, I am most grateful for the love and support of my parents, Mike and Carol, and the world's best siblings, Susan, Michelle, Staci, and Kevin. Without my family, I would be lost.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Abstract of dissertation Presented to the Graduate School
of the University of Florida in Partial Fulfillment of the
Requirements for the Degree of Doctor of Philosophy

TWO-LEVEL SYSTEM INTERDICTION

By

Kelly M. Sullivan

August 2012

Chair: J. Cole Smith
Major: Industrial and Systems Engineering

We consider a number of interdiction problems with applications in homeland security. First, we study a stochastic interdiction model formulated by Morton et al. (*IIE Transactions*, 39:3–14, 2007) that aims to locate radiation sensors at border crossings in order to detect and help prevent the smuggling of nuclear material. In this model, an interdictor places sensors at customs checkpoints to minimize a potential smuggler's maximum probability of crossing a border undetected. Our contribution focuses on a version of this model in which the interdictor has different, and likely more accurate, perceptions of the system's parameters than the smuggler does. We develop a class of valid inequalities along with a corresponding separation procedure that can be used within a cutting-plane approach to reduce computational effort.

We then extend this work by performing a polyhedral study for the deterministic special case. For this case, we develop the minimal convex hull representation for the polytope linking the interdictor's variables with the smuggler's. In the process, we expose an exponential class of easily-separable inequalities that generalize all of those developed so far for this class of problems. We argue that some instances of the stochastic model have facets corresponding to the solution of NP-hard problems. Our computational results show that the cutting planes developed in this chapter may strengthen the linear programming relaxation of the stochastic model by as much as 25 percent.

Next, we consider the interdiction of a capacitated network that exists in Euclidean space. Nodes in this network exist at a point in space and (directed) arcs connect node pairs in a straight line. An opponent wishes to maximize flow from a source node to a sink node across the network, while an interdictor seeks to minimize the opponent's maximum flow by choosing multiple locations to attack. In this problem, attacks are made at points in space. Damage is inflicted on each arc by reducing its capacity as a function of the distance from the midpoint of the arc to an attack. We provide mathematical programming-based approaches for solving this problem.

CHAPTER 1
INTRODUCTION

Operations research is a broad field that is concerned wtih the development of quantitative techniques to support decision making. Typically, operations research is concerned with the design or operation of a system to achieve maximum functionality. Within operations research, the field of interdiction seeks to identify combinations of disruptions that have the most detrimental effect to a system's functionality. Many systems have the capability to react to disruption by rerouting supplies or reassigning demand, for instance, so identifying most effective interdictions is not trivial. The earliest interdiction studies (see, e.g., [1, 2]) were military in nature, seeking strategical advances that would disrupt an enemy's supply and/or communication infrastructure. Since, interdiction research has expanded to consider the effects of many types of disruptions, both intentional and random, leading to an increased understanding of vulnerability in many types of systems in a variety of industries.

Interdiction problems typically consider a *defender* who attempts to minimize costs via the usage of an underlying system, and an *attacker* who acts to inhibit the defender's use of the system via selecting interdiction actions. Most interdiction literature falls into the category of two-level models. In two-level models, the attacker and defender each make one move: The attacker acts first to damage the defender's system, the defender then develops an operational plan for his/her system that minimizes cost given the attacker's action. The attacker's goal is to maximize the defender's minimum cost, and therefore, two-level models exhibit the structure of a Stackelberg game, which are naturally formulated as min-max mathematical programs. These optimization problems are usually very difficult to solve because evaluating the objective function requires solving the defender's optimization problem. Some attention has also been given to higher-order nested interdiction games, as well. In three-level models, the defender acts first to construct, fortify, or protect components of his/her system before the attacker

solves an two-level problem problem. The ability to solve three-level (or higher) problems could be extremely valuable in designing robust systems, but doing so is impossible until efficient methods are available for two-level problems.

In this dissertation, we focus on (i) advancing state-of-the-art theory relating to solving existing discrete two-level interdiction models, and (ii) discovering a new understanding of system vulnerability through the development of new discrete two-level interdiction models. Motivation for the problems considered in Chapters 2–4 is presented in Section 1.1 followed by a survey of relevant literature in Section 1.2.

1.1 Motivation and Significance

The research in Chapters 2 and 3 of this dissertation focuses on interdiction models with application to the prevention of nuclear smuggling. Chapter 4 develops a new class of interdiction models aimed at examing geographical pressure-points in a supply network. We provide further background and motivation to these application areas below.

Nuclear material is highly regulated worldwide with some 190 nations party to the Nuclear Non-Proliferation Treaty (NPT) [3]. This treaty aims to prevent the spread of nuclear weapons technology and ultimately achieve global nuclear disarmament. Unfortunately, security breaches at nuclear facilities, particularly those in the Former Soviet Union (FSU), have resulted in the theft and trafficking of radioactive material that could be used by a terrorist clan or NPT non-party to construct nuclear weapons. One way to reduce the threat of nuclear proliferation is to install radiation sensors along thoroughfares that might be popular routes for traffickers. Performing such "interdictions" is precisely the aim of the Second Line of Defense program of the United States Department of Energy, who promote and regulate the implementation of nuclear sensors in the FSU and around the world. Sensor installations have dual benefit: (1) deterrence of potential smugglers, and (2) detection and prevention of actual smugglers. However, installation of sensors is costly and typically performed incrementally; thus, a

best implementation strategy is not obvious. Moreover, in many cases, covert installation of sensors is impossible or impractical, and thus, potential smugglers are capable of rerouting to avoid this technology. Chapters 2 and 3 relate to the development of cost-effective sensor-installation strategies minimize a potential smuggler's chance of successfully trafficking nuclear material across a border.

The motivating applications of Chapter 4 are important as well but differ greatly from those of Chapters 2 and 3. Many critical systems (e.g., transportation networks or power grids) well-defined geographical characteristics. Moreover, these systems are subject to a number of disruptions (e.g., earthquakes, explosive weapons, or radiation sensors) that have effect over an area. The research presented in Chapter 4 aims to model these geometric relationships for a simple maximum-flow network (and hopefully pave the way for future geographical interdiction models).

1.2 Literature Survey

Interdiction literature dates back to Wollmer's 1964 paper [1] that poses the following *maximum flow network interdiction problem* (MFNIP): Find the n arcs in a maximum-flow network that, when removed, result in the smallest maximum-flow. A number of extensions to this work were developed in the following years. Lubore et al. [4] develops an efficient algorithm for solving the problem in the case of a single arc removal. Ghare et al. [5] develop an implicit enumeration algorithm for solving the problem for general n. Ratliff et al. [6] develop a network-modification method for solving MFNIP. McMasters and Mustin [2] consider a variant of MFNIP in which partial interdiction of arcs is permitted. Corley [7] shows that MFNIP can be adapted to identify the most critical nodes in a maximum flow network. MFNIP is known to be NP-hard either under the assumption that each arc must be either completely destroyed or unaffected [8] or if arcs can be partially interdicted [9]. Wood [8] provides a single-level mixed-integer programming (MIP) formulation of MFNIP using a dualization-linearization technique that has proven useful in modeling other min-max problems as well, and

which forms the basis of most contemporary interdiction research. Royset et al. [10] give a bi-objective model based on Wood's MIP, and Altner et al. [11] develop valid inequalities that strengthen the model of [8].

Fulkerson and Harding [12] and Golden [13] study the problem of continuously interdicting a shortest path subject to a budget constraint. The discrete shortest path network interdiction problem (SPNIP) (see, e.g., [14–16]) seeks to find the n arcs whose removal maximizes the length of a network's shortest path. This problem is known to be NP-hard in general [17], but Khachiyan et al. [18] give a polynomial algorithm for the case in which a fixed number of outgoing arcs must be removed from each node. Israeli and Wood [19] formulate SPNIP as a mixed integer program (MIP) and detail decomposition algorithms to solve the problem. Interdiction of a more general minimum-cost-flow network [20] and of a multicommodity-flow network [21, 22] has also been studied. While the majority of interdiction research centers around network-related second-stage problems, interdiction of facility location-allocation problems [23–27] has also received a significant amount of attention.

Recent contributions in this area consider problems in which interdictions have random effects on the system [28, 29], and in which the other player's behavior is uncertain [25, 30–33]. Bailey et al. [34] model a problem that has both of these types of uncertainty. While most studies of interdiction under uncertainty use an objective function based on expectation, [35, 36] maximize the probability that the disruption caused is in excess of a threshold. Information asymmetry between attacker and defender has been studied [32, 37–39] as have multiple-objective problems [10, 39, 40]. Applications of interdiction models vary widely, including infection control [41], natural disaster mitigation [42], military tactic development [43], natural resource and infrastructure security [38, 44–46], and prevention of nuclear proliferation [32, 33, 47].

Many approaches have been developed to model and solve two-level interdiction problems, but most are adopted from either (i) the transformation of a min-max

problem into a single stage minimization problem, as demonstrated in [8], or (ii) outer-approximating decomposition approaches such as Benders decomposition [48] (see [19, 33] for examples). Approaches (i) and (ii) are dependent upon the lower-level maximization problem having tractable structure, as is the case with network flow problems as well as other linear programs. Difficulties are associated with both approaches: Approach (i) typically results in a nonlinear formulation which must be linearized by introducing additional variables and constraints. Approach (ii) may prove difficult because the attacker's decision is typically combinatorial, meaning that the Benders master problem, which must be solved and re-solved iteratively, may be an integer program. Three-level interdiction problems are commonly approached (as in [49, 50]) by reformulating the lower two levels via approach (i) and solving the resulting model as a subroutine in an algorithm that implicitly enumerates first-level solutions.

CHAPTER 2
SECURING A BORDER UNDER ASYMMETRIC INFORMATION

2.1 Background

In this chapter, we consider a stochastic interdiction problem formulated by Morton et al. [32]. In this problem, an interdictor attempts to minimize a nuclear smuggler's probability of traversing a border undetected by *monitoring* (i.e., installing radiation sensors at) a limited number of customs checkpoints. In order to cross the border, the smuggler must pass through a single checkpoint. Associated with each checkpoint are evasion probabilities that depend on whether the checkpoint is monitored. The characteristics of the smuggler, including the nature of the material being smuggled, the smuggler's starting location and final destination, etc., are uncertain, but are assumed to follow a known discrete distribution. The realizations, or "scenarios," of this discrete distribution are manifested in the evasion probabilities at each checkpoint. The interdictor's goal is to identify a subset of checkpoints to monitor that minimizes the smuggler's unconditional probability of evasion, i.e., the weighted sum of evasion probabilities over all scenarios.

The timing of events and imbalance of information are key features in the problem we study. The interdictor must choose which checkpoints to monitor before the scenario characterizing the smuggler is revealed. Given the locations of the sensors, the smuggler then traverses the checkpoint having the greatest evasion probability. The smuggler is assumed to have perfect knowledge of where the sensors are placed, and thus, the problem exhibits the structure of a Stackelberg game. However, in contrast to most research on this class of problems, we consider the case in which the smuggler's perception of checkpoint-evasion probabilities, both with and without sensor installation, do not necessarily match the interdictor's perception. A key assumption is that the interdictor knows the smuggler's perception of these probabilities, or, more generally, that the interdictor is willing to assume a probability distribution over the smuggler's

perception. Thus, the interdictor may use deception (e.g., not monitoring a checkpoint for which the smuggler overestimates the evasion probability) to his advantage. In what follows, we refer to the interdictor's perception of the evasion probabilities as being the "true" evasion probabilities.

The problem we study is closely related to a class of network interdiction problems in which the smuggler seeks a maximum-reliability path through a network $G(N, A)$ and sensors may be placed on any arc in the network. Morton et al. [32] describe this stochastic network interdiction problem (SNIP) along with a number of related problems, including the problem we study in this chapter. For the case of SNIP in which sensors may only be installed on a predefined cut-set of arcs (i.e., those arcs $(i, j) \in A$ such that $i \in T$ and $j \in N \setminus T$ for some $T \subseteq N$), SNIP reduces to an interdiction problem defined on a *bipartite* network and is therefore designated as BiSNIP. The problem we study is exactly BiSNIP (where each cut-set arc corresponds to crossing through a checkpoint), but without the restriction that the smuggler must know the true evasion probabilities. Since the smuggler's *perception* of these probabilities need not match the true probabilities, our problem is a generalization of BiSNIP referred to as BiPSNIP. The relationship between SNIP and BiPSNIP is described in [32], and as such, we omit the details of SNIP in favor of defining BiPSNIP directly.

Morton et al. [32] define BiSNIP, BiPSNIP, and their counterparts when sensor installation is not restricted to a cut-set. In addition, [32] formulates both BiSNIP and BiPSNIP as MIPs and details valid inequalities for BiSNIP (but not for BiPSNIP) and a corresponding separation procedure that improves computation times significantly. BiSNIP is proven to be NP-hard in [51], which also establishes the result for the more general BiPSNIP.

In this chapter, we make the following contributions to the analysis of BiPSNIP. We revise the MIP formulation for BiPSNIP given in [32] to obtain a formulation that is tighter while using fewer constraints. We then define "step inequalities" for BiPSNIP that

generalize the inequalities given for BiSNIP in [32]. We devote significant attention to the analysis of these inequalities, leading to (1) characterization of necessary and sufficient conditions for these inequalities to define facets to the convex hull of integer-feasible solutions, and (2) a procedure that identifies a most-violated step inequality in polynomial time. Our computational results suggest that adding the step inequalities in branch-and-cut fashion improves solution times over solving the problem with commercially-available MIP solvers.

We organize the remainder of this chapter as follows. Section 2.2 summarizes our formulation of BiPSNIP and its relation to that in [32]. We define our new step inequalities in Section 2.3, characterize these inequalities, and describe an efficient separation routine. We demonstrate the effectiveness of our techniques on a set of test instances in Section 2.4, and examine characteristics of optimal solutions prescribed by our models in the context of a United States border security problem involving the road network.

2.2 Revised BiPSNIP Formulation

In this section, we define BiPSNIP and develop the notation to formulate BiPSNIP as a MIP. We then revise the BiPSNIP formulation given in [32] with one that is tighter and requires fewer constraints.

In BiPSNIP, we consider an interdictor who installs sensors on a subset of checkpoints with the aim of thwarting a smuggler's attempt to cross a border. Let Ω denote a finite set of scenarios, where each element $\omega \in \Omega$ specifies the true and perceived evasion probabilities for a potential smuggler. In this manner, we model the case in which the interdictor is unsure of the smuggler's characteristics, and hence the smuggler's evasion probabilities—both true and perceived—through each checkpoint. We assume the interdictor is willing to model this uncertainty via a discrete distribution on these evasion probabilities. Let f^ω, $\omega \in \Omega$, be the probability that scenario ω is

realized. Intuitively, we view each scenario ω as a unique smuggler, and hence we will refer to "smuggler ω."

Let $K = \{1, \dots, n\}$ denote the set of checkpoints. Each smuggler is characterized by the following data, defined for each $\omega \in \Omega$ and $k \in K$:

- p_k^ω: smuggler ω's *true* evasion probability through checkpoint k if k *is not monitored*,
- q_k^ω: smuggler ω's *true* evasion probability through checkpoint k if k *is monitored*,
- $\bar{p}_k^\omega$: smuggler ω's *perceived* evasion probability through checkpoint k if k *is not monitored*,
- $\bar{q}_k^\omega$: smuggler ω's *perceived* evasion probability through checkpoint k if k *is monitored*.

The asymmetry of the problem hinges on the assumption that smuggler ω knows only the $\bar{p}^\omega$-and $\bar{q}^\omega$-values, as well as the sensor locations, when selecting a checkpoint to traverse. By contrast, the interdictor has full knowledge of both the true probabilities (p^ω and q^ω) and the perceived probabilities ($\bar{p}^\omega$ and $\bar{q}^\omega$) when deciding which checkpoints to monitor. If the interdictor is unsure of the smuggler's perception, but is willing to posit a probability distribution on that perception, the same model holds with $\omega \in \Omega$ defined to include this uncertainty in the smuggler's characteristics. For simplicity in exposition, we assume that all $\bar{p}$- and $\bar{q}$-values are distinct and positive, and that $\bar{p}_k^\omega > \bar{q}_k^\omega,\ \forall \omega \in \Omega,\ k \in K$.

After sensors have been placed, each smuggler's problem amounts to selecting the checkpoint with the greatest perceived probability. That is, let $\hat{x}_k,\ k \in K$ indicate whether (1) or not (0) checkpoint k is monitored. Each smuggler ω traverses the checkpoint k^ω that maximizes $\bar{p}_k^\omega(1 - \hat{x}_k) + \bar{q}_k^\omega \hat{x}_k$ over all checkpoints $k \in K$. The resulting true evasion probability for smuggler ω is $p_{k^\omega}^\omega$ if $\hat{x}_{k^\omega} = 0$ and $q_{k^\omega}^\omega$ otherwise. The interdictor seeks to minimize the sum of all possible smugglers' true evasion probabilities, weighted by the probability that scenario ω is realized.

For each $\omega \in \Omega$, let $\bar{\bar{q}}^{\omega} = \max_{k \in K}\{\bar{q}_k^{\omega}\}$ and let $k^{\omega\star}$ be the checkpoint for which $\bar{\bar{q}}^{\omega} = \bar{q}_{k^{\omega\star}}^{\omega}$. Note that if $\bar{p}_k^{\omega} < \bar{\bar{q}}^{\omega}$, then smuggler ω would never attempt to traverse checkpoint k, regardless of the sensor placements. Hence, we define $K^{\omega} = \{k \in K | \, \bar{p}_k^{\omega} > \bar{\bar{q}}^{\omega}\}$ to be the set of checkpoints that—without sensors installed—smuggler ω perceives as allowing an evasion probability above that of a fully monitored border. Define an auxiliary checkpoint $n+1$, on which no sensor can be placed, where $\bar{p}_{n+1}^{\omega} = \bar{q}_{k^{\omega\star}}^{\omega}$ and $p_{n+1}^{\omega} = q_{k^{\omega\star}}^{\omega}$. Checkpoint $n+1$ corresponds to the smuggler's choice if all checkpoints in K^{ω} are monitored. Letting $\bar{K}^{\omega} = K^{\omega} \cup \{n+1\}$, we define A as those (ω, k) pairs in $\Omega \times \{1, \ldots, n+1\}$ such that $k \in \bar{K}^{\omega}$.

Smuggler ω's problem can now be stated as follows: Select the checkpoint $k \in \bar{K}^{\omega}$ having the largest $\bar{p}_k^{\omega}$-value among all checkpoints that are not monitored. (Note: Because $\bar{p}_k^{\omega} > \bar{p}_{n+1}^{\omega}$ for all $k \in K^{\omega}$ and checkpoint $n+1$ cannot be monitored, checkpoint $n+1$ is smuggler ω's selection if and only if all of the checkpoints in K^{ω} are monitored.) A consequence of this simplification is that sensor installation now effectively removes a checkpoint and as a result, the problem no longer depends upon q and $\bar{q}$.

In order to represent the BiPSNIP model, we define additional notation, and summarize notation defined previously, as follows.

DATA

- Ω: set of scenarios, i.e., set of potential smugglers
- f^{ω}: probability of realizing scenario $\omega \in \Omega$
- $K = \{1, \ldots, n\}$: set of checkpoints
- B: budget for installing sensors
- b_k: cost to install a sensor at checkpoint k $(0 \leq b_k \leq B, \; \forall k \in K)$
- $K^{\omega} \subseteq K$: set of checkpoints that are candidates for selection by smuggler $\omega \in \Omega$
- $n+1$: auxiliary checkpoint, chosen by smuggler $\omega \in \Omega$ if each checkpoint in K^{ω} receives a sensor
- $\bar{K}^{\omega} = K^{\omega} \cup \{n+1\}$

- $A = \{(\omega, k) : \omega \in \Omega, k \in \bar{K}^\omega\}$
- p_k^ω: true evasion probability for smuggler ω through checkpoint $k \in \bar{K}^\omega$
- $\bar{p}_k^\omega$: perceived evasion probability for smuggler ω through checkpoint $k \in \bar{K}^\omega$ $(\bar{p}_k^\omega > \bar{p}_{n+1}^\omega > 0,\ \forall k \in K^\omega)$

VARIABLES

- $x_k = \begin{cases} 1 & \text{if a sensor is installed at checkpoint } k \\ 0 & \text{otherwise} \end{cases}$
- $\theta^\omega =$ smuggler ω's true evasion probability
- ψ^ω = smuggler ω's perceived evasion probability
- $y_k^\omega = \begin{cases} 1 & \text{if smuggler } \omega \text{ chooses checkpoint } k \\ 0 & \text{otherwise} \end{cases}$

Model (2–1) below is a revised version of the BiPSNIP formulation given by [32], where $X = \left\{x \in \{0,1\}^{|K|} | \sum_{k \in K} b_k x_k \leq B\right\}$.

$$\text{Min} \quad \sum_{\omega \in \Omega} f^\omega \theta^\omega, \tag{2–1a}$$

$$\text{subject to} \quad \theta^\omega \geq p_k^\omega y_k^\omega, \qquad (\omega, k) \in A, \tag{2–1b}$$

$$\psi^\omega \geq \bar{p}_k^\omega (1 - x_k), \qquad (\omega, k) \in A, \tag{2–1c}$$

$$\psi^\omega = \sum_{k:(\omega,k) \in A} \bar{p}_k^\omega y_k^\omega, \qquad \omega \in \Omega, \tag{2–1d}$$

$$\sum_{k:(\omega,k) \in A} y_k^\omega = 1, \qquad \omega \in \Omega, \tag{2–1e}$$

$$y_k^\omega \leq 1 - x_k, \qquad (\omega, k) \in A, \tag{2–1f}$$

$$y_k^\omega \geq 0, \qquad (\omega, k) \in A, \tag{2–1g}$$

$$x \in X. \tag{2–1h}$$

The objective function (2–1a) minimizes the true evasion probability. Constraints (2–1b) bound the conditional true evasion probabilities, conditioned on each smuggler ω, depending on which checkpoints are monitored and which checkpoint is chosen by

smuggler ω. For a given binary $x = \hat{x}$, smuggler ω solves

$$\min\{\psi^\omega : \psi^\omega \geq \bar{p}_k^\omega(1 - \hat{x}_k),\ \forall k \in \bar{K}^\omega\}. \tag{2–2}$$

Constraints (2–1c)–(2–1g) represent optimality conditions for (2–2) in the form of primal feasibility (2–1c), dual feasibility (2–1e)–(2–1g), and strong duality (2–1d). Constraints (2–1h) ensure that sensor installation is discrete and feasible within budget restrictions. We let $x_{n+1} \equiv 0$ and remove it from the model so that constraints (2–1c) and (2–1f) have right-hand side $\bar{p}_{n+1}^\omega$ and 1, respectively, for $(\omega, n+1)$.

Remark 2.1. If X is defined by a cardinality constraint, i.e., $b_k = 1\ \forall k \in K$ and $B \in \mathbb{Z}$, then Model (2–1) can be simplified. Smuggler ω's choice is limited to the $B+1$ checkpoints $k \in K$ with the largest $\bar{p}_k^\omega$-values. Accordingly, constraints (2–1b) and (2–1c) corresponding to checkpoints that will not be chosen can be removed, and corresponding y-variables can be fixed to zero. This notion extends to handle non-unit b_k-values and non-integer B. We traverse smuggler ω's checkpoints k from largest to smallest $\bar{p}_k^\omega$, summing b_k until the sum exceeds B. The remaining checkpoints can be removed from K^ω, and hence A. □

We now reformulate model (2–1) in order to tighten the formulation. Again, let k^ω denote the checkpoint that smuggler ω traverses so that $y_{k^\omega}^\omega = 1$ and $y_k^\omega = 0$ for $k \neq k^\omega$. For fixed ω, this means that (2–1b) reduces to $\theta^\omega \geq 0$ for all $k \neq k^\omega$ such that $(\omega, k) \in A$, and the k^ω-th constraint is a simple nonnegative lower bound on θ^ω. Since we minimize θ^ω, it is optimal to set θ^ω equal to the lower bound given by the k^ω-th constraint. This condition can equivalently be given by

$$\theta^\omega = \sum_{k:(\omega,k)\in A} p_k^\omega y_k^\omega,\ \omega \in \Omega. \tag{2–3}$$

Updating (2–1), we obtain the valid MIP formulation

$$\text{Min} \quad \sum_{\omega \in \Omega} f^{\omega} \theta^{\omega}, \tag{2–4a}$$

$$\text{subject to} \quad x \in X,$$

$$\theta^{\omega} = \sum_{k:(\omega,k)\in A} p_k^{\omega} y_k^{\omega}, \ \omega \in \Omega, \tag{2–4b}$$

(2–1c) through (2–1g).

Replacing θ^{ω} in the objective function (2–4a) with the right-hand side of (2–4b) and eliminating (2–4b) is also valid, but we prefer the form of model (2–4) for ease of exposition in the remainder of this chapter. Model (2–4) can also be simplified according to Remark 2.1.

Model (2–4) has $|A| - |\Omega|$ fewer constraints, as we removed $|A|$ constraints of the form (2–1b), and added $|\Omega|$ constraints of the form (2–4b). Moreover, model (2–4) is tighter than model (2–1) in the sense that solutions that are feasible to the linear programming relaxation of (2–4) are feasible to the linear programming relaxation of (2–1), but the converse does not hold. We now formalize these statements.

Proposition 2.1. Let P_1 denote the linear programming relaxation of (2–1) and P_2 the linear programming relaxation of (2–4), i.e., the relaxations from replacing X in (2–1) and (2–4) with $\bar{X} = \left\{x \in [0,1]^{|K|} | \sum_{k\in K} b_k x_k \leq B\right\}$. Let $(\hat{x}, \hat{\theta}, \hat{\psi}, \hat{y})$ be a feasible solution to P_2. Then $(\hat{x}, \hat{\theta}, \hat{\psi}, \hat{y})$ is also feasible to P_1.

Proof. We need only show that $(\hat{x}, \hat{\theta}, \hat{\psi}, \hat{y})$ satisfies (2–1b) since the remainder of the constraints in (2–1) are in (2–4) as well. By hypothesis,

$$\hat{\theta}^{\omega} = \sum_{k:(\omega,k)\in A} p_k^{\omega} \hat{y}_k^{\omega}, \quad \forall \omega \in \Omega. \tag{2–5}$$

Since all of the terms in (2–5) are nonnegative, we have that $\hat{\theta}^{\omega} \geq p_k^{\omega} \hat{y}_k^{\omega}$ for each $(\omega, k) \in A$. So $(\hat{x}, \hat{\theta}, \hat{\psi}, \hat{y})$ satisfies (2–1b) and must be feasible to P_1. □

Remark 2.2. The converse of Proposition 2.1 is false: Feasibility of $(\hat{x}, \hat{\theta}, \hat{\psi}, \hat{y})$ to problem P_1 does not ensure that $(\hat{x}, \hat{\theta}, \hat{\psi}, \hat{y})$ is feasible to P_2, even in the (nontrivial) case in which one of the constraints (2–1b) is tight for each $\omega \in \Omega$. As an example, suppose that $K^\omega = \{1, 2\}$, $\forall \omega \in \Omega$, and all p-values are strictly positive. Further, consider the vector $\hat{x} = (0.5, 0.5) \in \bar{X}$ and let $\hat{y}_1^\omega = \hat{y}_2^\omega = 0.5$ for each $\omega \in \Omega$. Observe that $\hat{x}$ and $\hat{y}$ satisfy constraints (2–1e) and (2–1f)–(2–1h). Define

$$\hat{\psi}^\omega = \sum_{k:(\omega,k)\in A} \bar{p}_k^\omega \hat{y}_k^\omega = 0.5\bar{p}_1^\omega + 0.5\bar{p}_2^\omega, \tag{2–6}$$

to satisfy (2–1d). Since $\bar{p}_k^\omega \geq 0$ for all $(\omega, k) \in A$, (2–6) implies $\hat{\psi}^\omega \geq 0.5\bar{p}_1^\omega$ and $\hat{\psi}^\omega \geq 0.5\bar{p}_2^\omega$ for each $\omega \in \Omega$, i.e., (2–1c) is satisfied. Under $\hat{y}$, constraints (2–1b) reduce to $\theta^\omega \geq 0.5p_k^\omega,\ k \in \{1, 2\},\ \omega \in \Omega$. For each $\omega \in \Omega$, let $k^\omega \in \arg\max_{k\in\{1,2\}} p_k^\omega$, and set $\hat{\theta}^\omega = 0.5p_{k^\omega}^\omega$ so that this choice of $(\hat{x}, \hat{\theta}, \hat{\psi}, \hat{y})$ satisfies (2–1b), and is feasible to P_1.

To assess whether $(\hat{x}, \hat{\theta}, \hat{\psi}, \hat{y})$ is feasible to P_2 it suffices to examine constraint (2–4b) since all other constraints of P_2 are already verified. Constraint (2–4b) reduces to $\theta^\omega = 0.5p_1^\omega + 0.5p_2^\omega,\ \omega \in \Omega$, in this case. However, $\hat{\theta}^\omega = 0.5p_{k^\omega}^\omega < 0.5p_1^\omega + 0.5p_2^\omega$, given that p_1^ω and p_2^ω are positive. Thus, $(\hat{x}, \hat{\theta}, \hat{\psi}, \hat{y})$ is not feasible to P_2. □

2.3 Step Inequalities and Separation

In this section we derive a class of valid inequalities that bounds the evasion probability θ^ω for a given scenario $\omega \in \Omega$. Following [32], we refer to our inequalities as *step inequalities* because the bound for θ^ω is computed by subtracting "steps"—which may be taken or not, depending on the value of x—from a constant. Our valid inequalities for BiPSNIP generalize the BiSNIP inequalities of [32]. In addition, we characterize necessary and sufficient conditions under which our inequalities are facet-defining, and we identify a procedure that separates a most-violated BiPSNIP step inequality in polynomial time. Our facet result and separation procedure for the BiPSNIP inequalities generalize analogous results, given in [51] and [32] respectively, for the BiSNIP inequalities.

Without loss of generality, assume that the checkpoints in K^ω are uniquely indexed such that $K^\omega = \{1, \dots, n'\}$, where

$$\bar{p}_1^\omega > \cdots > \bar{p}_{n'}^\omega > \bar{\bar{q}}^\omega. \tag{2–7}$$

With these assumptions, smuggler ω will traverse the lowest-indexed checkpoint in $\bar{K}^\omega = K^\omega \cup \{n+1\}$ that does not have a sensor. Recall that auxiliary checkpoint $n+1$, which never has a sensor, corresponds to the smuggler's choice if all checkpoints in K^ω have a sensor.

The inequalities proposed in this section are associated with subsets of $\bar{K}^\omega$. Hence, we introduce the following notation corresponding to these subsets.

- $H = \{h_1, \dots, h_m\} \subseteq \bar{K}^\omega$, where $h_1 < \cdots < h_m$. Set H represents the checkpoint subset from which our proposed step inequality will be generated.
- k: An element of $\bar{K}^\omega$ and/or H.
- $I(H) = \{1, \dots, m\}$. The set $I(H)$ is used to index a checkpoint according to its ordering in the set H, i.e., $i \in I(H)$ corresponds to h_i (the i-th checkpoint in H).
- i: An element of $I(H)$.
- $\pi_i = p_{h_i}^\omega$, $i \in I(H)$. The vector $\pi \in \mathbb{R}_+^m$ contains those p_k^ω-values corresponding to $k \in H$.
- indices $\{\ell_1, \dots, \ell_m\}$ corresponding to $I(H)$, where ℓ_i is the smallest index in $I(H)$ such that $\ell_i \geq i$ and $\pi_{\ell_i} \geq \pi_j,\ \forall j = \ell_i + 1, \dots, m$. Note that π_{ℓ_i} represents the largest possible true evasion probability among checkpoints in H if checkpoints $1, \dots, h_i - 1$ all receive sensors.

Given a subset of checkpoints $H \subseteq \bar{K}^\omega$, the BiPSNIP step inequality can be stated as

$$\theta^\omega \geq \pi_{\ell_1} - \sum_{i \in I(H):\ell_i = i} (\pi_i - \pi_{\ell_{i+1}}) x_{h_i} - \sum_{i \in I(H):\ell_i > i} (\pi_{\ell_i} - \pi_i)(1 - x_{h_i}), \tag{2–8}$$

where $\pi_{\ell_{m+1}} \equiv \pi_m (= \pi_{\ell_m})$ so that the term $(\pi_m - \pi_{\ell_{m+1}}) x_{h_m} = 0$. However, as the following example shows, inequalities (2–8) are only sometimes valid, depending on our choice of H.

Example. *Suppose* $n = 5$ *and, as indicated in* (2–7)*, suppose the smuggler prefers checkpoints in the order* $1, 2, 3, 4, 5$*, provided they do not have sensors. That is, the smuggler chooses the lowest-indexed checkpoint without a sensor or, if checkpoints* 1–5 *all receive sensors, the smuggler chooses the auxiliary checkpoint* 6*. For this scenario, let* $(p_1^\omega, \dots, p_6^\omega) = (0.82, 0.8, 0.9, 0.65, 0.7, 0.55)$*. Smuggler* ω*'s ordering of the checkpoints differs from the true ordering, indicating information asymmetry. For example, smuggler* ω *prefers checkpoint* 1 *to checkpoint* 3 *even though* $p_1^\omega < p_3^\omega$*. Thus, if* $H = \{1, \dots, 6\}$ *we have* $(\pi_1, \dots, \pi_6) = (0.82, 0.8, 0.9, 0.65, 0.7, 0.55)$*,* $(\ell_1, \dots, \ell_6) = (3, 3, 3, 5, 5, 6)$*, and the step inequality* (2–8) *is*

$$\theta^\omega \geq 0.9 - 0.08(1 - x_1) - 0.1(1 - x_2) - 0.2x_3 - 0.05(1 - x_4) - 0.15x_5. \tag{2–9}$$

If $H = \{2, 4, 5, 6\}$, we have $(\pi_1, \dots, \pi_4) = (0.8, 0.65, 0.7, 0.55)$, $(\ell_1, \dots, \ell_4) = (1, 3, 3, 4)$, and the inequality is

$$\theta^\omega \geq 0.8 - 0.1x_2 - 0.05(1 - x_4) - 0.15x_5. \tag{2–10}$$

Both (2–9) and (2–10) are valid inequalities. However, with $H = \{3, 4, 5, 6\}$, we obtain $(\pi_1, \dots, \pi_4) = (0.9, 0.65, 0.7, 0.55)$, $(\ell_1, \dots, \ell_4) = (1, 3, 3, 4)$, and the inequality

$$\theta^\omega \geq 0.9 - 0.2x_3 - 0.05(1 - x_4) - 0.15x_5, \tag{2–11}$$

is not valid because $\theta^\omega = 0.82$ when $x_k = 0$ for all k. □

Motivated by this example, we seek to characterize which choices of H induce valid inequalities. To this end, we define set G_k as the checkpoints in $\bar{K}^\omega$ with both: (i) a larger perceived evasion probability than checkpoint k and (ii) a smaller true evasion probability than k. That is, $G_k = \{1 \leq k' \leq k - 1 : p_{k'}^\omega < p_k^\omega\}$ for all $k \in \bar{K}^\omega$.

Theorem 2.1. A set $H = \{h_1, \dots, h_m\} \subseteq \bar{K}^\omega$ induces a valid inequality of the form (2–8) if (a) for every $k \in H$, each element $k' \in G_k$ belongs to H as well, and (b) $\pi_m \leq p_k^\omega$, $\forall k \in \{h_m + 1, \dots, n', n + 1\}$. Moreover, if B, in the constraint set $X =$

$\left\{x \in \{0,1\}^{|K|} \mid \sum_{k\in K} b_k x_k \leq B\right\}$, is sufficiently large, H induces a valid inequality of the form (2–8) only if conditions (a) and (b) are satisfied.

Proof. First, suppose that conditions (a) and (b) hold. Let $(\hat{x}, \hat{\theta}^\omega)$ be a subvector of a feasible solution $(\hat{x}, \hat{\theta}, \hat{\psi}, \hat{y})$ to model (2–4). Let $\hat{x}_{n+1} \equiv 0$, and let k' be the smallest index in $\bar{K}^\omega$ such that $\hat{x}_{k'} = 0$. We show that the right-hand side of (2–8) is at most $p^\omega_{k'}$. To this end, for each $k \in H$, define $i(k)$ as the element of $I(H)$ satisfying $h_{i(k)} = k$. Since the terms of (2–8) corresponding to $i \in I(H) : h_i > k'$ are all nonpositive, and terms corresponding to $i \in I(H) : \ell_i > i,\ h_i < k'$ are zero, it suffices to show that

$$p^\omega_{k'} \geq \pi_{\ell_1} - \sum_{i\in I(H):\ \ell_i = i,\ h_i < k'} (\pi_i - \pi_{\ell_{i+1}}) - (\pi_{\ell_{i(k')}} - \pi_{i(k')}),\ \text{if } k' \in H \text{ and } \ell_{i(k')} > i(k'), \tag{2–12a}$$

$$p^\omega_{k'} \geq \pi_{\ell_1} - \sum_{i\in I(H):\ \ell_i = i,\ h_i < k'} (\pi_i - \pi_{\ell_{i+1}}),\ \text{else.} \tag{2–12b}$$

Consider the case in which $k' \in H$ and $\ell_{i(k')} > i(k')$, corresponding to (2–12a). Note that the terms $\pi_{\ell_1} - \sum_{i\in I(H):\ell_i = i, h_i < k'} (\pi_i - \pi_{\ell_{i+1}})$ collapse to $\pi_{\ell_{i(k')}}$. Thus, the right-hand side of (2–12a) becomes $\pi_{i(k')}$, which equals $p^\omega_{k'}$, and the inequality is valid. Now suppose that $k' \notin H$ (case 1), or $k' \in H$ with $\ell_{i(k')} = i(k')$ (case 2), where both cases correspond to (2–12b). If $k' \notin H$ in case 1, then the right-hand side of (2–12b) equals $\pi_{i(k'')}$, where $k'' \in H$ is the smallest index greater than k' such that $\ell_{i(k'')} = i(k'')$ if $k' < h_m$, and otherwise equals π_m if $k' > h_m$. If $k' < h_m$, then we have that $p^\omega_{k'} \geq \pi_{i(k'')}$ by hypothesis (a), or else k' would belong to $G_{k''}$, and so (2–12b) holds. If $k' > h_m$, then $p^\omega_{k'} \geq \pi_m$ by hypothesis (b), which satisfies (2–12b). Finally, suppose that $k' \in H$ and $\ell_{i(k')} = i(k')$ in case 2. The right-hand side of (2–12b) collapses to $\pi_{i(k')} = p^\omega_{k'}$, and so (2–12b) holds again.

Next, suppose that condition (a) is violated. In this case, there exists some $k', k'' \in \bar{K}^\omega$ for which $k' \in G_{k''},\ k' \notin H$, and $k'' \in H$. In particular, let $k'' \in H$ again be the smallest index greater than k' such that $\ell_{i(k'')} = i(k'')$. Then, consider the solution in which

- $\hat{x}_k = 1 \quad \forall k = 1, \ldots, k' - 1,$

- $\hat{x}_{k'} = 0$,
- $\hat{x}_{h_i} = 0 \quad \forall i \in I(H) : \ell_i = i,\ h_i > k'$,
- $\hat{x}_{h_i} = 1 \quad \forall i \in I(H) : \ell_i > i,\ h_i > k'$,
- $\hat{x}_k = 0$ (arbitrarily) $\quad \forall k \in K^\omega \setminus H,\ k > k'$.

Note that $\hat{x} \in X$ when B is sufficiently large. For this solution, we should find that $\hat{\theta}^\omega = p^\omega_{k'}$. But the right-hand side terms of (2–8) are zero for each term $i \in I(H)$ for which $h_i > k'$, and for each term $i \in I(H)$ in which both $h_i < k'$ and $\ell_i > i$. Thus, (2–8) becomes

$$\theta^\omega \geq \pi_{\ell_1} - \sum_{i \in I(H):\ \ell_i = i,\ h_i < k'} (\pi_i - \pi_{\ell_{i+1}}) x_{h_i}. \tag{2–13}$$

The right-hand side of (2–13) condenses to $\pi_{i(k'')} = p^\omega_{k''}$. However, $p^\omega_{k'} < p^\omega_{k''}$, noting that $k' \in G_{k''}$. Thus, inequality (2–8) is not valid in this case.

Finally, if condition (b) is violated then we have that $k' \notin H$ for some $k' \in \bar{K}^\omega,\ h_m < k'$, and $\pi_m > p^\omega_{k'}$. The solution in which $\hat{x}_k = 1,\ \forall k = 1, \ldots, k' - 1$, and $\hat{x}_k = 0,\ \forall k = k', \ldots, n', n+1$, causes all terms corresponding to $i \in I(H),\ \ell_i > i$ of (2–8) to equal zero, and the remaining terms to collapse to π_m. Since $\pi_m > p^\omega_{k'}$, (2–8) is not valid. This completes the proof. □

Note that in our example above, the choices of H used to generate (2–9) and (2–10) satisfy the condition for validity from Theorem 2.1, but the choice of H used to generate (2–11) does not: In that case, $3 \in H$, but $1 \notin H$ and $1 \in G_3$. If the smuggler's perceived evasion probabilities match the true evasion probabilities, i.e., $p = \bar{p}$, then (2–8) is valid for any H that includes $n + 1$. This special case takes the form of the step inequalities derived in [32] for the situation with symmetric information. In the following theorem, we establish another important parallel between the symmetric and asymmetric cases. First, we establish the dimensionality of the feasible region linking θ^ω and the x-variables. For the remainder of this section, we assume for simplicity that $|K^\omega| = n' = n$. The case in which $|K^\omega| < n$ is a straightforward extension of these results.

Lemma 2.1. Let $Y^\omega = \{(\hat{\theta}^\omega, \hat{x}) : \exists\,(x, \theta, \psi, y)$ feasible to model (2–4) with $\hat{\theta}^\omega \geq \theta^\omega$ and $\hat{x} = x\}$. Then, the dimension of Y^ω is $n+1$, $\forall \omega \in \Omega$.

Proof. We fix $\omega \in \Omega$ and reorder the checkpoint indices so that condition (2–7) is satisfied. We generate $c^k \in Y^\omega$, $k = 1, \ldots, n+2$, which denote $n+2$ affinely independent points of the form $(\hat{\theta}^\omega, \hat{x})$ as follows. Let $c^1 = (p_2^\omega, e_1)$ and $c^k = (p_1^\omega, e_k)$, $k = 2, \ldots, n$, where e_k denotes the n-dimensional unit vector with a 1 in the k-th component. Let $c^{n+1} = (p_1^\omega, 0)$ and $c^{n+2} = (p_1^\omega + 1, 0)$. Note that the $\hat{x}$-component of each vector satisfies $\hat{x} \in X$ by the assumption that $b_k \leq B$, $\forall k \in K$. The $\hat{\theta}$-component satisfies constraint (2–4b) by construction and by condition (2–7).

Form the $(n+1) \times (n+1)$ matrix C in which row k corresponds to $(c^k - c^{n+2})$, for $k = 1, \ldots, n+1$. We prove that the only solution to $\lambda^T C = 0$ is the trivial solution $\lambda_k = 0,\ \forall k = 1, \ldots, n+1$. Columns $2, \ldots, n+1$ of C are identity columns, implying that $\lambda_1 = \cdots = \lambda_n = 0$. Column 1 now implies that $\lambda_{n+1} = 0$, as required. □

We now establish conditions under which (2–8) induces a facet to the convex hull of integer-feasible solutions to the budget-relaxed BiPSNIP.

Theorem 2.2. Assume that the hypotheses of Lemma 2.1 hold, and that the checkpoint indices are ordered according to (2–7). Fix $\omega \in \Omega$ and let $\bar{Y}^\omega$ be identical to Y^ω but without the restriction that $\sum_{k \in K} b_k x_k \leq B$. Then, a valid inequality (2–8)satisfying the assumptions of Theorem 2.1 induces a facet to $\bar{Y}^\omega$ if and only if index 1 is an element of H.

Proof. First, suppose that $1 \notin H$. If $p_1^\omega < \pi_{\ell_1}$, then by Theorem 2.1, (2–8) would not be valid. Else, we would have that $p_1^\omega > \pi_{\ell_1}$. Consider lifting (2–8) by adding $\tau(1 - x_1)$ to its right-hand side. If $x_1 = 1$, the inequality is valid for any value of τ. If $x_1 = 0$, then τ can take on a value of $p_1^\omega - \pi_{\ell_1} > 0$. Since this value is positive, (2–8) can be strengthened and thus does not define a facet.

Now suppose that checkpoint 1 is an element of H. We generate affinely independent points, $c^k \in \bar{Y}^\omega$, $k = 1, \dots, n+1$, of the form $(\hat{\theta}^\omega, \hat{x})$ that are binding on (2–8). Define $I^\star = \{i \in I(H) : \ell_i > i\} \setminus \{1\}$. For $k \in H$, let the first component of c^k be p_k^ω, the next $k-1$ components be 1, and all other components be 0, except for those corresponding to variables x_{h_i} for $i \in I^\star$ such that $h_i > k$, which take on values of 1. For $k \notin H,\ k \leq n$, let c^k be given exactly as c^1, except for a 1 in the component corresponding to x_k. If $n+1 \notin H$, then define c^{n+1} exactly as c^1, except for a 1 in the component corresponding to x_{h_m}. (This component must have been zero in all points thus far, and so c^{n+1} is distinct from $c^1, \dots, c^n$.) Observe that $c^1, \dots, c^{n+1}$ are all elements of $\bar{Y}^\omega$, and all are binding on (2–8).

Define C as an $n \times (n+1)$ matrix whose rows are composed of the $(c^k - c^1)$ vectors for $k = 2, \dots, n+1$, such that row k of C corresponds to $(c^{k+1} - c^1)$. We prove that $c^1, \dots, c^{n+1}$ are affinely independent by showing that C has rank n. To this end, we show that columns $2, \dots, n+1$ are linearly independent. Let $\bar{C}$ denote the matrix composed of the last n columns of C. We prove that the only solution to $\bar{C}\lambda = 0$, is the trivial solution $\lambda_k = 0,\ \forall k = 1, \dots, n$. For $k \in \bar{K}^\omega \setminus H,\ k \neq n+1$, $\bar{C}_{k-1\cdot}\lambda = 0$ implies that $\lambda_k = 0$, where $\bar{C}_{k-1\cdot}$ is the $(k-1)$-st row of $\bar{C}$, corresponding to the last n components of $(c^k - c^1)$. We can thus remove the k-th column of $\bar{C}$ along with the row of $\bar{C}$ corresponding to $(c^k - c^1)$. If $n+1 \in \bar{K}^\omega \setminus H$, then we perform the same operations, infer $\lambda_{h_m} = 0$ and eliminate its corresponding row and column in $\bar{C}$. For $k = h_i,\ i \in I^\star$, equations $\bar{C}_{h_i-1\cdot}\lambda = 0$ and $\bar{C}_{h_{\ell_i}-1\cdot}\lambda = 0$ together imply that $\lambda_k = 0$, and we can eliminate from $\bar{C}$ column k and the row corresponding to $(c^k - c^1)$. Let $\tilde{C}$ denote the reduced C-matrix and observe that $\tilde{C}$ is a lower triangular matrix with nonzero diagonals. Thus, $\tilde{C}$ has full rank, i.e., the remaining λ_k must all be zero. This implies that the rank of C is n, as is the dimension of the face intersected by (2–8). □

Note that for the case in which $p = \bar{p}$, the step inequality (2–8) induces a facet to $\bar{Y}^\omega$ whenever H includes checkpoint 1, proving that the step inequalities used in

[32] are indeed facets to the convex hull of integer-feasible solutions for the BiSNIP problem. This special case was proven by Pan [51], a result that is now generalized by Theorem 2.2.

While step inequalities help to tighten the continuous relaxation of model (2–4), they are not sufficient to represent the convex hull of feasible solutions to (2–4), even in the case where there is only one scenario and probabilities are symmetric. As an example, consider a scenario ω with checkpoints $\bar{K}^\omega = \{1, 2, 3\}$ and $(p_1^\omega, p_2^\omega, p_3^\omega) = (\bar{p}_1^\omega, \bar{p}_2^\omega, \bar{p}_3^\omega) = (1, 0.4, 0)$. Suppose $b_1 = b_2 = 1$ and $B = 1$. Consider a fractional interdiction vector, $(\hat{x}_1, \hat{x}_2) = (0.75, 0.25)$, and observe that $(\hat{y}_1^\omega, \hat{y}_2^\omega, \hat{y}_3^\omega) = (0.25, 0.5, 0.25)$ combined with $\hat{\theta}^\omega = \hat{\psi}^\omega = 0.45$ completes a feasible solution to the continuous relaxation of model (2–4). There are only two step inequalities for this scenario, given as $\theta^\omega \geq 1 - x_1$ and $\theta^\omega \geq 1 - 0.6x_1 - 0.4x_2$. Under $\hat{x}$, these reduce to $\theta^\omega \geq 0.25$ and $\theta^\omega \geq 0.45$, respectively, neither of which cut off $(\hat{x}, \hat{y}, \hat{\psi}, \hat{\theta})$. In this problem, the set X consists of only three solutions, $X = \{(0, 0), (0, 1), (1, 0)\}$, which result in $\theta^\omega \geq 1$, $\theta^\omega \geq 1$, $\theta^\omega \geq 0.4$,

respectively, for any feasible solution to (2–4). The only way to obtain $\hat{x}$ as a convex combination of the points in X is $0.75(1, 0) + 0.25(0, 1)$ which would imply $\theta^\omega \geq 0.75(0.4) + 0.25(1) = 0.55$ if $(\hat{x}, \hat{y}, \hat{\psi}, \hat{\theta})$ were in the convex hull of feasible solutions to (2–4). Since $\hat{\theta}^\omega = 0.45$, $(\hat{x}, \hat{y}, \hat{\psi}, \hat{\theta})$ resides outside the convex hull of feasible solutions, yet cannot be cut off by a step inequality.

In this example, the step inequalities were accounting for a solution ($x_1 = x_2 = 1$ and $\theta^\omega = 0$) that is infeasible because of the budget constraint. (This is a case where the assumption in Theorem 2.2 about B being sufficiently large fails.) We now describe a modified step inequality that takes into account the restrictions imposed by the budget constraint. For simplicity, we consider only the case in which X is governed by a cardinality constraint, i.e., $b_k = 1 \; \forall k \in K$, although as discussed in Remark 2.1 this type of analysis extends in a straightforward way to handle a general knapsack constraint.

Remark 2.3. When X is defined by the cardinality constraint $\sum_{k\in K} x_k \leq B$ and checkpoints are ordered according to (2–7), smuggler ω will choose the lowest-indexed checkpoint in the set $\{1, \dots, B+1\}$ that is not monitored. As described in Remark 2.1, model (2–4) can be simplified by removing (ω, k) from A for all $k > B+1$. Step inequality construction is modified as well: Generating (2–8) now entails selecting H from a smaller $\bar{K}^\omega = \{1, \dots, B+1\}$. When step inequalities are generated in this way, the assumption that B is sufficiently large is not necessary for Theorems 2.1 and 2.2 to hold (because all points used in their proofs satisfy $\sum_{k\in K} x_k \leq B$). □

We now describe a separation procedure that takes as input a solution $\hat{x}$ to the linear programming relaxation of model (2–4). In polynomial time, the separation procedure either identifies a most-violated step inequality or shows that there are no violated step inequalities. The procedure accomplishes this by maximizing the right-hand side of (2–8) with respect to all checkpoint subsets H corresponding to valid facet-defining step inequalities. This maximization is carried out by solving a longest-path problem on an acyclic network, $G(V, E)$. Let the node set $V = \bar{K}^\omega \cup \{s, t\}$, where s denotes a dummy starting node and t a dummy termination node. We construct the arc set E as follows: Create arc (s, k) for each $k \in \bar{K}^\omega$ satisfying $p_1^\omega \leq p_k^\omega$, including $k = 1$. Similarly, create arc (k, t) for each $k \in \bar{K}^\omega$ satisfying $p_k^\omega \leq p_{k'}^\omega,\ \forall k' = k+1, \dots, n+1$. Finally, create arc (k, k') for $k, k' \in \bar{K}^\omega$ if $k < k'$ and $k \notin G_{k'}$. Each s-t path in this network corresponds to a unique facet-inducing H-set. Inclusion of $k \in \bar{K}^\omega$ in an s-t path corresponds to the inclusion of $G_k \cup \{k\}$ in H, and by construction, the step inequality derived from H must be valid and facet-defining.

Given a solution $\hat{x}$ to the linear programming relaxation of model (2–4), we define the arc lengths, $c_{kk'}$, $(k, k') \in E$, as follows:

$$c_{sk'} = p^{\omega}_{k'} - \sum_{j \in G_{k'}} (p^{\omega}_{k'} - p^{\omega}_{j})(1 - \hat{x}_j), \ \forall (s, k') \in E, \tag{2–14a}$$

$$c_{kk'} = -(p^{\omega}_{k} - p^{\omega}_{k'})\hat{x}_k - \sum_{j \in G_{k'} \setminus G_k} (p^{\omega}_{k'} - p^{\omega}_{j})(1 - \hat{x}_j), \ \forall (k, k') \in E, k, k' \in \bar{K}^{\omega}, \tag{2–14b}$$

$$c_{kt} = 0, \ \forall (k, t) \in E. \tag{2–14c}$$

Consider the length of an s-t path given the arc lengths (2–14). Choosing (s, k') initializes our construction of the right-hand side of (2–8) via arc length $c_{sk'}$ in (2–14a). An arc (k, k') in the s-t path with $k, k' \in \bar{K}^{\omega}$ contributes $c_{kk'}$ from (2–14b), which accounts for overlap in G_k and $G_{k'}$ with the contribution capturing the addition of checkpoints $(G_{k'} \cup \{k'\}) \setminus G_k$ to H. Finally, there is no additional contribution from the arc that terminates in t.

Recall that Theorem 2.1 specifies two conditions that characterize a set H that yields a valid step inequality and Theorem 2.2 specifies one further condition that ensures H yields a facet. Our separation procedure yields a valid inequality, since the first condition of Theorem 2.1 is satisfied by allowing the selection of checkpoint $k \in \bar{K}^{\omega}$ within H only if all checkpoints in G_k are also in H, and the second condition of Theorem 2.1 is satisfied by our construction of the (k, t) arcs in E. Furthermore, the separation procedure yields a facet-defining inequality due to our construction of the arcs $(s, k) \in E$, which ensure that checkpoint 1 belongs to H. Moreover, any facet-defining valid step inequality induced by a set H can be recovered by choosing a path through nodes h_i for $i \in I(H)$ such that $\ell_i = i$. Because there is a one-to-one correspondence between s-t paths in our network and facet-defining step inequalities, our separation routine yields a most-violated step inequality.

2.4 Computational Results

In this section, we first present computational results obtained from solving randomly generated instances of model (2–4) via a branch-and-cut algorithm under a number of different cut-generation strategies. We then analyze a problem instance derived from the road network in Canada, Mexico, and the United States, and provide insight into our model's solutions. All computations were performed using a Dell Poweredge T610 machine containing two six-core hyperthreading 3.33 GHz Xeon processors. Each MIP problem was solved using CPLEX 12.1 using callback functions from the Concert Technology 2.9 C++ library to implement the separation procedure.

We now describe the procedure used to generate BiPNSIP instances given a set of checkpoints K and scenarios Ω. We first randomly generated p_k^ω (the true evasion probabilities) uniformly on the interval $[0.25, 0.75]$ for each $(\omega, k) \in \Omega \times K$. We assume the effectiveness of a sensor is governed by a parameter $\lambda \in [0, 1)$, so that $q_k^\omega = \lambda p_k^\omega$.

Values p_k^ω (or $\bar{p}_k^\omega$) for each $k \in K$ can be used to deduce a *preference order* of checkpoints — the k with the greatest p_k^ω is preferred first, the second-greatest p_k^ω is preferred second, and so on. If each smuggler ω's preference order (i.e., the order obtained from the $\bar{p}_k^\omega$-values) matches the preference order obtained from the p_k^ω-values, the problem is symmetric and could be solved using the simpler BiSNIP model and step inequalities from [32]. Our hypothesis is that the problem is more difficult to solve when the p_k^ω- and $\bar{p}_k^\omega$-values lead to preference orders that do not match closely.

To control the degree to which each smuggler's preference order resembles the preference order obtained from p_k^ω, we specify $d \geq 0$ and randomly generate $\bar{p}_k^\omega$ uniformly in the interval $[p_k^\omega - d, p_k^\omega + d]$. Naturally, d must be chosen so that all of the $\bar{p}_k^\omega$-values fall inside the range $[0, 1]$. When $d = 0$, $\bar{p}_k^\omega$ equals p_k^ω, resulting in the symmetric model. The parameter λ is assumed to be known to each smuggler, so that $\bar{q}_k^\omega = \lambda \bar{p}_k^\omega$. Under larger λ, fewer checkpoints are relevant to each smuggler's

optimization problem (i.e., $|K^\omega|$ is smaller), and hence, the problem should be easier to solve.

We construct sets $\bar{K}^\omega$ for each $\omega \in \Omega$ as described in Section 2.2, excluding any checkpoints that would never be chosen by smuggler ω, even when all other checkpoints have sensors. We assume that $b_k = 1$ and B is integer-valued so that $x \in X$ states that at most B sensors may be placed, and we generate step inequalities as in Remark 2.3.

To solve model (2–4), we employ a branch-and-cut algorithm with parameters (D, R). In this algorithm, we add step inequalities only in the first D levels of the branch-and-bound tree, and we add at most R cuts per scenario at each node. We test the following algorithms: No step inequality generation $(0, 0)$, aggressive step inequality generation $(100, 100)$, and two hybrid methods $(10, 2)$ and $(20, 1)$. (We also implemented variations of these algorithms in which step inequalities are added exhaustively at the root node. Our experience suggests that this results in longer solution times, and results are thus omitted.) We use callback functions from the Concert Technology library to add step inequalities as needed in algorithms $(100, 100)$, $(10, 2)$, and $(20, 1)$. Doing so disables some features of CPLEX that could be used in solving the $(0, 0)$ implementation. In order to ensure a fair comparison, we disable the same CPLEX features in the $(0, 0)$ implementation by running an empty callback.

We initially set parameters $d = 0.1$ and $\lambda = 0.75$, and generated five instances each having size $|K| = |\Omega| = 30$ with $B = 15$, and $|K| = |\Omega| = 50$ with $B = 25$. We then generated modified instances by varying one of the parameters and holding the rest constant. We considered $d \in \{0, 0.1, 0.25\}$, $\lambda \in \{0.5, 0.75, 0.85\}$, $B \in \{7, 15, 22\}$ (for the $|K| = |\Omega| = 30$ experiments), and $B \in \{12, 25, 37\}$ (for the $|K| = |\Omega| = 50$ experiments). We solved each test case using the branch-and-cut algorithms explained above, and report Spearman's rank correlation (column **S**) for the $B + 1$ greatest p_k^ω- and $\bar{p}_k^\omega$-values (averaged across all scenarios), the average solution time (in seconds, given by $\mathbf{T}(\mathbf{s})$) for each of the four algorithms, the average optimality gap (column $\mathbf{G}(\%)$, which is the

Table 2-1. Comparison of algorithms as asymmetry parameter d is varied.

						(0,0)			(10,2)			(20,1)			(100,100)		
$\|\Omega\|$	$\|K\|$	B	λ	d	S	T(s)	G(%)	U	T(s)	G(%)	U	T(s)	G(%)	U	T(s)	G(%)	U
30	30	15	0.75	0	1.00	20.0	0.0	0	4.1	0.0	0	4.2	0.0	0	4.6	0.0	0
				0.1	0.86	62.6	0.0	0	6.3	0.0	0	5.2	0.0	0	4.0	0.0	0
				0.25	0.78	8.5	0.0	0	1.2	0.0	0	1.4	0.0	0	1.4	0.0	0
50	50	25	0.75	0	1.00	3600	1.9	5	2579	0.1	1	3007	0.2	3	2744	0.4	3
				0.1	0.86	3600	3.4	5	1950	0.1	1	1925	0.1	1	2004	0.2	1
				0.25	0.79	3600	2.3	5	253	0.0	0	293	0.0	0	466	0.0	0

absolute optimality gap divided by the objective lower bound, multiplied by 100) after one hour of computation, and the number of instances unfinished (column **U**) after one hour of computation.

Tables 2-1–2-3 show that the $(D, R) = (10, 2)$, $(20, 1)$, and $(100, 100)$ implementations consistently outperform the pure branch-and-bound implementation, suggesting that the step inequalities and corresponding separation procedure help in solving BiPSNIP instances. Implementation $(100, 100)$ performed the best in only one group of experiments (row two in each of Tables 2-1–2-3) and did so by only a slight margin. Implementations $(10, 2)$ and $(20, 1)$ were usually the two best implementations of the four, with $(10, 2)$ typically outperforming $(20, 1)$ by a slight margin. We therefore use $D = 10$ and $R = 2$ in the remainder of our computations.

Table 2-1 displays the results as d is varied. As seen in this table, algorithms $(10, 2)$ and $(20, 1)$ consistently outperform $(0, 0)$ and usually outperform $(100, 100)$. This suggests that while step inequalities help in solving BiPSNIP, they should not be added too aggressively. Perhaps counter-intuitively, the problems seem to become much easier as d increases, i.e., as the problem's asymmetry increases. One explanation for this relationship arises from a scenario ω in which the smuggler severely overestimates the true evasion probability p_k^ω through a particular checkpoint k. Therefore, we require fewer sensors to reduce the evasion probability for scenario ω to p_k^ω and simpler solutions ensue.

Variation of λ (see Table 2-2) shows a more intuitive result: Less effective sensors (i.e., larger values of λ) means that fewer checkpoints are relevant in each scenario and

Table 2-2. Comparison of algorithms as sensor effectiveness λ is varied.

						(0,0)			(10,2)			(20,1)			(100,100)		
$\|\Omega\|$	$\|K\|$	B	d	λ	S	T(s)	G(%)	U	T(s)	G(%)	U	T(s)	G(%)	U	T(s)	G(%)	U
30	30	15	0.1	0.5	0.78	116.1	0.0	0	20.3	0.0	0	19.2	0.0	0	23.0	0.0	0
				0.75	0.86	62.6	0.0	0	6.3	0.0	0	5.2	0.0	0	4.0	0.0	0
				0.85	0.87	4.1	0.0	0	0.5	0.0	0	0.5	0.0	0	0.6	0.0	0
50	50	25	0.1	0.5	0.77	3600	4.8	5	3600	1.4	5	3600	2.0	5	3600	2.4	5
				0.75	0.86	3600	3.4	5	1950	0.1	1	1925	0.1	1	2004	0.2	1
				0.85	0.87	2524	0.3	1	48	0.0	0	61	0.0	0	52	0.0	0

Table 2-3. Comparison of algorithms as budget B is varied.

						(0,0)			(10,2)			(20,1)			(100,100)		
$\|\Omega\|$	$\|K\|$	λ	d	B	S	T(s)	G(%)	U	T(s)	G(%)	U	T(s)	G(%)	U	T(s)	G(%)	U
30	30	0.75	0.1	7	0.55	6.5	0.0	0	0.3	0.0	0	0.3	0.0	0	0.3	0.0	0
				15	0.86	62.6	0.0	0	6.3	0.0	0	5.2	0.0	0	4.0	0.0	0
				22	0.91	11.0	0.0	0	3.6	0.0	0	4.2	0.0	0	4.7	0.0	0
50	50	0.75	0.1	12	0.56	3458	0.3	3	18	0.0	0	22	0.0	0	24	0.0	0
				25	0.86	3600	3.4	5	1950	0.1	1	1925	0.1	1	2004	0.2	1
				37	0.91	3600	3.1	5	2057	0.0	0	1887	0.0	0	2869	0.8	3

the problem is easier to solve. Again, a less aggressive cut-generation strategy seems to perform favorably in these instances. In particular, when $\lambda = 0.85$, we require an average of 0.5 seconds to solve our problem instances when $|\Omega| = |K| = 30$, and an average of 48 seconds to solve the instances in which $|\Omega| = |K| = 50$. Changes in B (see Table 2-3) also seem to have a pronounced effect on solution time. Instances in which B is about $|K|/2$ required the most computation and the required computational effort is not symmetric about $|K|/2$. Low-budget instances solved in a fraction of the time of high-budget instances, again because of the number of relevant checkpoints in the sets K^ω.

In the remainder of this section, we present computational results based on data derived from the road network in Mexico, Canada, and the contiguous 48 U.S. states. The data include 113 checkpoints corresponding to ports of entry for road crossings from Mexico and Canada into the U.S. There are 140 scenarios arising from each combination of seven origins in Mexico, seven origins in Canada, and 10 destinations in the United States. Evasion probabilities p_k^ω are given for each (ω, k) where an ω with origin in Canada has access to all k on the U.S.-Canada border and an origin in

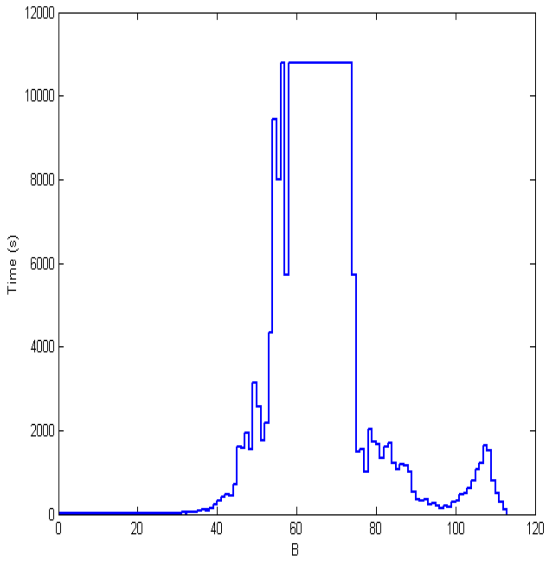

Figure 2-1. Computation time (in seconds) vs. B for Mexico-Canada-U.S. instance.

Mexico similarly has access to all k on the U.S.-Mexico border. We randomly generate probabilities $\bar{p}_k^\omega$ uniformly in the range $[p_k^\omega - p_k^\omega(1-p_k^\omega), p_k^\omega + p_k^\omega(1-p_k^\omega)]$, and we assume sensors to have effectiveness $\lambda = 0.5$.

This instance was solved 114 times, corresponding to each value of $B = 0, 1, \ldots, 113$. All but 17 of the 114 runs finished within three hours, and the remainder were terminated within 2.4% of optimality (see Figures 2-1 and 2-2). Statistics on the number of nodes generated from the branch-and-bound tree and the number of step inequalities generated within the time limit are shown in Figures 2-3 and 2-4. (Because the procedure terminates after three hours, we do not know the actual number of branch-and-bound nodes and step inequalities needed to solve the most difficult instances.) Computation time increases sharply for the medium-level budget values, with the algorithm failing to terminate within the three-hour time limit when $B = 56$ and $B = 58, 59, \ldots, 73$.

As shown in Figure 2-5, sensor installation appears to have diminishing returns: The incremental improvement in optimal objective function value is larger for the first sensor than for, say, the 50-th. Figure 2-6 illustrates the solution to each of the

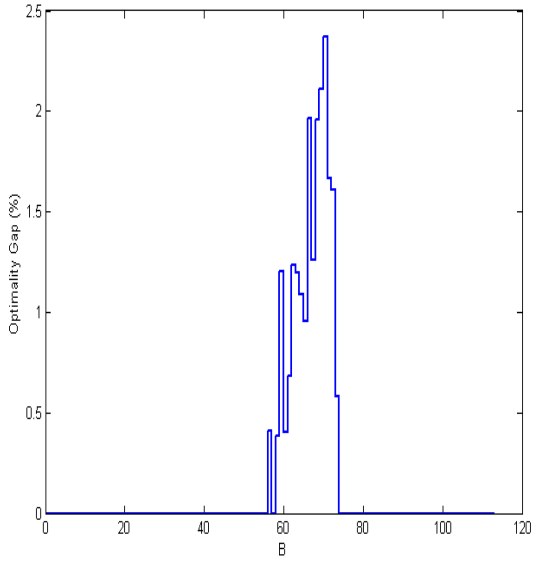

Figure 2-2. Optimality gap (in %) vs. B for Mexico-Canada-U.S. instance.

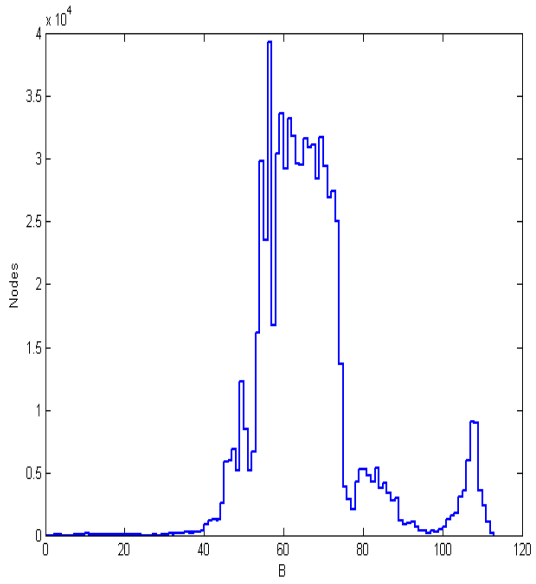

Figure 2-3. Number of branch-and-bound nodes vs. B for Mexico-Canada-U.S. instance.

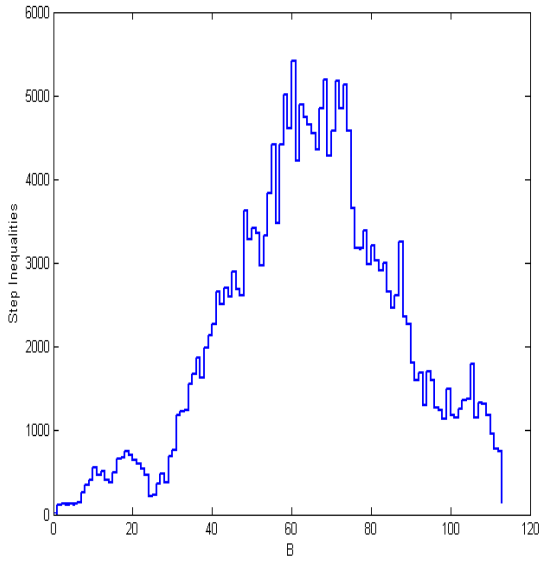

Figure 2-4. Number of step inequalities added vs. B for Mexico-Canada-U.S. instance.

114 problems, with rows corresponding to checkpoints and columns corresponding to the 114 instances. For each instance, shaded pixels indicate which checkpoints received sensors. As illustrated in Figure 2-6, the solutions to instances $B = 1$ through $B = 71$ largely exhibit a nested structure: Each successive solution builds upon the one before, moving only a few sensors to new checkpoints. As B increases from 1 to 71, the Mexican border is covered first, followed by the eastern Canadian border. Then, as B changes from 71 to 72, the solution obtained shifts the bulk of sensors from the eastern Canadian border to the western Canadian border. As B increases beyond 72, the solutions again exhibit the nested structure, gradually restrengthening the eastern Canadian border.

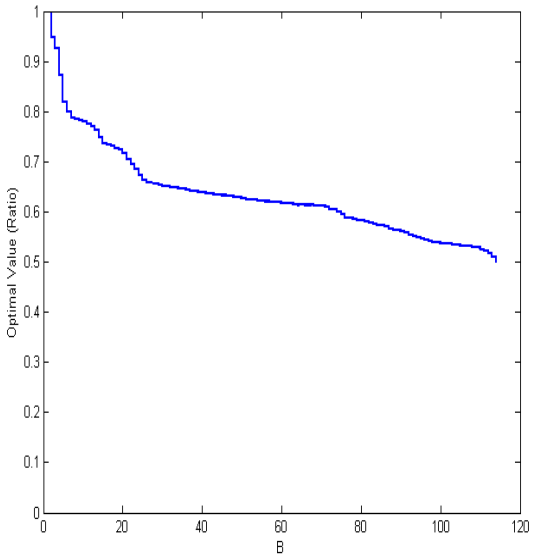

Figure 2-5. Optimal value vs. B for the instance Mexico-Canada-U.S. instance. The optimal value is reported as a ratio of the optimal objective function value for budget level B to that when $B = 0$.

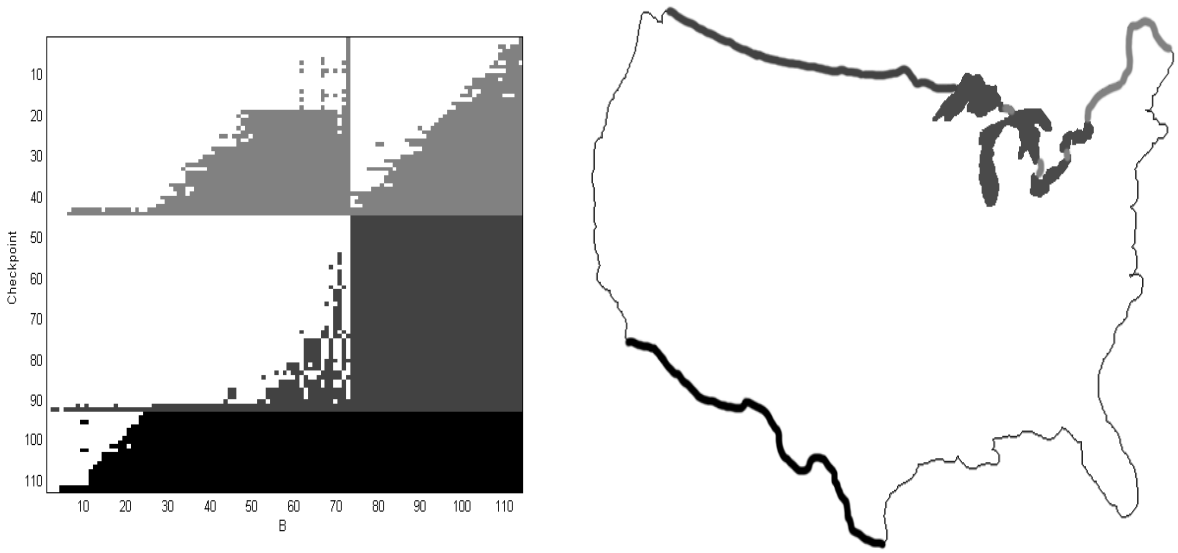

Figure 2-6. Plot of sensor-placement locations (shaded pixels) vs. B for the Mexico-Canada-U.S. instance. Pixel color indicates the geographical region in which each sensor is located, as indexed in the figure to the right.

CHAPTER 3
CONVEX HULL REPRESENTATION FOR THE DETERMINISTIC BORDER MONITORING PROBLEM

3.1 Background

We consider the asymmetric bipartite stochastic network interdiction problem (BiPSNIP), which has applications in countering nuclear smuggling. In this chapter, we briefly summarize BiPSNIP in the context of a game played by a *leader* against a number of *opponents*. The game is played over a set of *alternatives* from which each opponent must select one. The leader may remove a limited number of the alternatives in order to influence each opponent's selection. The leader seeks to select alternatives to remove such that the total cost of the resulting opponent-alternative pairings is minimized. The "opponents" and "alternatives" of this chapter play the role of the "smugglers" and "customs checkpoints," respectively, of Chapter 2. We use more general terminology in this chapter in order to emphasize that (i) BiPSNIP has the structure of a very simple game and (ii) BiPSNIP may have many applications outside the scope of nuclear interdiction. For a more in-depth introduction to this problem, we refer the reader to [32] and Chapter 2 of this dissertation.

Consider a set of alternatives $\{1, \dots, n\}$ simultaneously available to a set of opponents Ω. In BiPSNIP, the leader may remove each alternative $k \in \{1, \dots, n\}$ at a cost of b_k, and the total cost of removals must not exceed B. We consider the special case of BiPSNIP in which $b_k = 1,\ \forall k \in \{1, \dots, n\}$ and $B \in \mathbb{Z}$. We refer to this version as the *cardinality*-constrained BiPSNIP, or C-BiPSNIP.

Each opponent $\omega \in \Omega$ in C-BiPSNIP values the alternatives differently and seeks to choose the alternative that has the greatest perceived value. Opponent $\omega \in \Omega$ perceives the value of an alternative i to be 0 if i is interdicted and $\bar{p}_i^\omega$ otherwise. For $\omega \in \Omega$ define the sequence $\{k_i^\omega\}_{i=1}^{\ell}$ for indices $i \in \{i : \bar{p}_{k_i^\omega}^\omega > 0\}$ such that

$$\bar{p}_{k_1^\omega}^\omega > \bar{p}_{k_2^\omega}^\omega > \cdots > \bar{p}_{k_\ell^\omega}^\omega, \tag{3–1}$$

and define $n' = \min\{\ell, B\}$. (Note: ℓ and n' may take different values for each opponent $\omega \in \Omega$, but we suppress subscripts ω on these values for notational convenience.) Thus, opponent $\omega \in \Omega$ selects the alternative k_i^ω corresponding to the lowest index $i = 1, \dots, n'$ that is not interdicted; if all are interdicted, opponent ω selects a "last resort" alternative. Let $n+1$ be a dummy alternative representing each opponent's last resort, and let $k_{n'+1}^\omega = n+1, \ \forall \omega \in \Omega$. If $\ell \geq B+1$, define $\bar{p}_{n+1}^\omega = \bar{p}_{k_{B+1}^\omega}^\omega$; else, define $\bar{p}_{n+1}^\omega = \epsilon$, where $\epsilon > 0$ is an arbitrarily small constant. We assume that alternative $n+1$ cannot be interdicted.

Remark 3.1. This notation may also be used to model the more general case in which opponent $\omega \in \Omega$ perceives the value of alternative k to be $\hat{q}_k^\omega$ if k is interdicted and $\hat{p}_k^\omega > \hat{q}_k^\omega$ otherwise (as presented in Chapter 2). In this case, define $\bar{\hat{q}}^\omega = \max_{k=1}^n \{\hat{q}_k^\omega\}$ and $\bar{p}_k^\omega = \hat{p}_k^\omega - \bar{\hat{q}}^\omega, \ \forall \omega \in \Omega, k \in \{1, \dots, n\}$ and proceed as before (computing ℓ, n', and the k_i^ω-indices using the modified $\bar{p}$-values). □

Remark 3.2. Although we consider C-BiPSNIP, many of the results contained in this chapter are also valid for the more general BiPSNIP. Notably, all of the valid inequalities in this chapter can also be extended to BiPSNIP. However, the convex hull result (Theorem 3.5) holds only for C-BiPSNIP.□

For $i = 1, \dots, n$, let x_i equal one if alternative i is interdicted and zero otherwise. For $i = 1, \dots, n'+1$, let binary variable $y_i'^\omega$ indicate whether alternative k_i^ω is chosen by opponent ω. Accordingly, for a fixed interdiction vector $\bar{x}$, each opponent $\omega \in \Omega$ seeks a solution to

$$y^\omega \in \operatorname{argmax} \sum_{i=1}^{n'+1} \bar{p}_{k_i^\omega}^\omega (1 - \bar{x}_{k_i^\omega}) y_i'^\omega, \tag{3–2a}$$

$$\text{s.t.} \sum_{i=1}^{n'+1} y_i'^\omega = 1, \tag{3–2b}$$

$$y_i'^\omega \geq 0, \ \forall i = 1, \dots, n'+1, \tag{3–2c}$$

where $\bar{x}_{k_{n'+1}^\omega} = \bar{x}_{n+1} = 0, \ \forall \omega \in \Omega$.

The leader scores each alternative-opponent pairing as well: For $i = 1, \dots, n' + 1$, let $p^{\omega}_{k^{\omega}_i}$ denote the *leader's cost* incurred due to opponent $\omega \in \Omega$ selecting alternative k^{ω}_i. The leader seeks an interdiction vector x that minimizes the sum of p-values corresponding to each opponent's choice, weighted by $f^{\omega} > 0$:

$$\text{Min} \sum_{\omega \in \Omega} f^{\omega} \theta^{\omega}, \tag{3–3a}$$

$$\text{s.t. } \theta^{\omega} = \sum_{i=1}^{n'} p^{\omega}_{k^{\omega}_i} y^{\omega}_i, \ \forall \omega \in \Omega, \tag{3–3b}$$

$$y^{\omega} \text{ is optimal for (3–2)}, \ \forall \omega \in \Omega, \tag{3–3c}$$

$$\sum_{i=1}^{n} x_i \leq B, \tag{3–3d}$$

$$x \in \{0,1\}^n. \tag{3–3e}$$

In the context of the nuclear smuggling interpretation of BiPSNIP described in Chapter 2, $\hat{p}$ ($\hat{q}$) from Remark 3.1 represents each smuggler's assessment of the probability of successful evasion through a customs checkpoint without (with) a sensor. The p-values of (3–3) represent the true-evasion probabilities, known only by the interdictor, associated with the same customs checkpoints. We now illustrate models (3–2) and (3–3) using an example.

Example. *Suppose* $n = 6$, $B = 4$, $\Omega = \{1,2\}$, $f^1 = f^2 = 0.5$, *and (following Remark 3.1)* $\hat{q}^{\omega}_k$ *and* $\hat{p}^{\omega}_k$ *are given for each* $k \in \{1, \dots, 6\}$ *and* $\omega \in \{1,2\}$*:*

- *Opponent 1:* $\hat{p}^1 = [0.3, 0.55, 0.4, 0.6, 0.8, 0.2]$, $\hat{q}^1 = [0.05, 0.28, 0.15, 0.25, 0.2, 0.02]$*;*
- *Opponent 2:* $\hat{p}^2 = [0.65, 0.45, 0.25, 0.72, 0.35, 0.9]$, $\hat{q}^2 = [0.25, 0.3, 0.17, 0.45, 0.07, 0.55]$.

Proceeding as in Remark 3.1, $\bar{\bar{q}}$, $\bar{p}$, ℓ, *and* n' *are computed for each opponent in Table 3.1.*

Given $\bar{x} \in \{0,1\}^6$, *constraint (3–3c) requires that opponent* 1 *choose either (a) the first alternative* k *in* $(5, 4, 2, 3)$, *moving from left to right, such that* $\bar{x}_k = 0$, *or (b)*

Table 3-1. Derived data for example problem.

	$\omega = 1$	$\omega = 2$
$\bar{\bar{q}}^\omega$	0.28	0.55
$\bar{p}^\omega$	$[0.02, 0.27, 0.12, 0.32, 0.52, -0.08]$	$[0.1, -0.1, -0.3, 0.17, -0.2, 0.35]$
ℓ	5	3
n'	4	3
$(k_1^\omega, \dots, k_{n'+1}^\omega)$	$(5, 4, 2, 3, 7)$	$(6, 4, 1, 7)$
$(\bar{p}^\omega_{k_1^\omega}, \dots, \bar{p}^\omega_{k_{n'+1}^\omega})$	$(0.52, 0.32, 0.27, 0.12, 0.02)$	$(0.35, 0.17, 0.1, \epsilon)$

alternative 7 *if* $\bar{x}_5 = \bar{x}_4 = \bar{x}_2 = \bar{x}_3 = 1$. *Likewise, opponent* 2 *must choose the first alternative* k *in* $(6, 4, 1)$ *such that* $\bar{x}_k = 0$ *or alternative* 7 *if* $\bar{x}_6 = \bar{x}_4 = \bar{x}_1 = 1$. *Different* p*-vectors result in different optimal solutions to model* (3–3). *For instance, if* $p^1 = \bar{p}^1$ *and* $p^2 = \bar{p}^2$ *(the symmetric case), the solution* $x = (0, 1, 0, 1, 1, 1)$ *forces* $y_4^1 = y_3^2 = 1$ *(i.e., opponents 1 and 2 select alternatives* $k_4^1 = 3$ *and* $k_3^2 = 1$ *respectively) and is optimal for* (3–3) *with value* $0.5(0.12) + 0.5(0.1) = 0.11$. *Suppose instead* p^1 *and* p^2 *are given exactly as* $\bar{p}^1$ *and* $\bar{p}^2$*, except* $p_4^1 = p_4^2 = 0.09$. *In this case, both opponents have severely overvalued alternative* 4*, and the interdictor can take advantage using, e.g.,* $x = (0, 0, 0, 0, 1, 1)$. *Constraint* (3–3c) *forces each opponent to select alternative* 4*, resulting in an (optimal) objective value of* 0.09 *in model* (3–3).

□

In [32] (and Chapter 2 of this dissertation), (3–3c) is enforced by imposing the Karush-Kuhn-Tucker (KKT) optimality conditions of (3–2), which are necessary and sufficient since (3–2) is a linear program. The complementary slackness conditions, along with the fact that all $\bar{p}$-values are positive, imply that $x_{k_i^\omega} y_i^\omega = 0$ for all $\omega \in \Omega$, $i = 1, \dots, n' + 1$. The resulting problem is thus a mixed-integer linear programming problem, albeit one whose continuous relaxation tends to be fairly weak. The goal in this chapter is to develop a model that explicitly captures the relationships between x- and y-variables, making use of the k_i^ω-orderings rather than the $\bar{p}_i^\omega$-values, in order to obtain a tighter formulation. Ultimately, we are interested in studying the set

$$P^{\Omega} = \left\{ (x, y) \in X \times \{0,1\}^{(n'+1)\times|\Omega|} \;\middle|\; y_i^{\omega} = x_{k_1^{\omega}} \cdots x_{k_{i-1}^{\omega}} (1 - x_{k_i^{\omega}}),\ \forall \omega \in \Omega,\ i = 1, \dots, n'+1 \right\}, \tag{3–4}$$

where $x_{k_{n'+1}^{\omega}} = x_{n+1} = 0$ is fixed as a constant and $X = \{x \in \{0,1\}^n : \sum_{i=1}^{n} x_i \leq B\}$. Due to (3–1), if $(x, y) \in P^{\Omega}$, then y represents a solution to (3–2) under interdiction x.

Model (3–3) is easy to solve when $\Omega = \{1\}$ is a singleton set, because the interdictor need only compare the $n'+1$ solutions resulting from interdicting all of checkpoints $k_1^1, \dots, k_{h-1}^1$ and not interdicting checkpoint k_h^1, for each $h = 1, \dots, n'+1$. This suggests that we can likely find a tight formulation for $P^{\{1\}}$. A similar formulation exists for the case in which $p_{k_1^1}^1 > \cdots > p_{k_{n'+1}^1}^1$ (i.e., the interdictor and opponent agree upon a ranking of alternatives) and the $\sum_i x_i \leq B$ constraint is removed. For this case, Miller and Wolsey [52] prove that the step inequalities of [32] convexify the set

$$\left\{ (\theta^1, x) \;\middle|\; \begin{array}{l} \exists y^1 \text{ such that } (\theta^1, x, y^1) \\ \text{satisfies (3–3b), (3–3c), and (3–3e)} \end{array} \right\}, \tag{3–5}$$

corresponding to $\Omega = \{1\}$. However, inclusion of $\sum_i x_i \leq B$ significantly changes the feasible region and in fact leads to an exponential class of facets to $\text{conv}[P^{\{1\}}]$. We contribute in Section 3.3 a full characterization of $\text{conv}[P^{\{1\}}]$, and show that our representation is minimal. The motivation for obtaining this convex hull is in tightening the linear programming relaxations for (stochastic, multi-scenario) BiPSNIP problems. Partial convex hull representations, such as those generated via the intersection of subproblem convex hulls, have been used successfully in previous research (see, e.g., the general discussions in [53, 54]) to develop strong cutting planes for difficult problems. We thus test the effectiveness of valid inequalities that are facet-defining to these one-scenario convex hulls in solving C-BiPSNIP instances in this chapter.

The remainder of this chapter is organized as follows. In Section 3.2, we simplify notation for the one-opponent problem and characterize its solutions using an intuitive

linear formulation with binary variables. We develop inequalities and a corresponding separation procedure to tighten this formulation in Section 3.3, resulting in a minimally-represented ideal formulation. In Section 3.4, we discuss the extension of our results to the multiple-opponent problem. We report computational results from a number of multiple-opponent instances in Section 3.5 solved with and without our inequalities.

3.2 Basic Formulation

For the majority of this chapter, we consider a single opponent (and thus remove the ω-index) who, without loss of generality, prefers the alternatives in increasing order, i.e., $k_i = i,\ \forall i = 1, \dots, n'$, and

$$\bar{p}_1 > \bar{p}_2 > \cdots > \bar{p}_{n'} > \bar{p}_{n'+1}. \tag{3–6}$$

Accordingly, the opponent will select the lowest-indexed alternative among $\{1, \dots, n'\}$ that is not interdicted, or the dummy alternative $n'+1$ if all of $\{1, \dots, n'\}$ are interdicted. Define P, the one-opponent feasibility set analogous to (3–4), as

$$P = \left\{ (x, y) \in X \times \{0,1\}^{n'+1} \;\middle|\; \begin{array}{l} y_i = x_1 \cdots x_{i-1}(1 - x_i),\ \forall i = 1, \dots, n' \\ y_{n'+1} = x_1 \cdots x_{n'} \end{array} \right\}. \tag{3–7}$$

To characterize the polyhedron $\text{conv}(P)$, we first develop a linear representation of P. Let $\bar{P}$ denote the set of all $(x, y) \in \mathbb{R}^n \times \mathbb{R}^{n'+1}$ satisfying the following constraints:

$$\sum_{j=1}^{n} x_j \leq B, \tag{3–8a}$$

$$\sum_{i=1}^{n'+1} y_i = 1, \tag{3–8b}$$

$$y_1 = 1 - x_1, \tag{3–8c}$$

$$y_i \leq 1 - x_i, \quad \forall i = 2, \dots, n', \tag{3–8d}$$

$$y_i \leq x_j, \quad \forall i = 2, \dots, n'+1,\ \forall j = 1, \dots, i-1, \tag{3–8e}$$

$$y_i \geq 0, \quad \forall i = 1, \dots, n'+1, \tag{3–8f}$$

$$0 \leq x_j \leq 1, \quad \forall j = n'+1, \dots, n. \tag{3–8g}$$

Constraint (3–8a) is valid for P since $x \in X,\ \forall (x, y) \in P$. Following from (3–7),

$$\sum_{i=1}^{n'} y_i = 1 - x_1 x_2 \cdots x_{n'} = 1 - y_{n'+1}, \tag{3–9}$$

and hence, (3–8b) is valid as well. Constraint (3–8c) is simply the definition (3–7) of y_1. The remaining y-variables are linearized in the standard way using (3–8d) and (3–8e). (The typical lower-bounding linearization constraints, e.g., $y_i \geq (1 - x_i) - \sum_{j=1}^{i-1}(1 - x_j)$ are implied by (3–8b)–(3–8e) and are thus omitted.) Constraints (3–8f)–(3–8g) are valid because x and y are binary vectors. Constraints $0 \leq x_i \leq 1$ for each $i = 1, \ldots, n'$ are implied by (3–8b)–(3–8e) and (3–8f), and are therefore omitted.

We now argue that (3–8) is a valid formulation of P, i.e., $\bar{P} \cap (\mathbb{Z}^n \times \mathbb{Z}^{n'+1}) = P$. Let $(\bar{x}, \bar{y})$ be any binary solution in $\bar{P}$. If $\bar{x}_i = 1,\ \forall i = 1, \ldots, n'$, then (3–8b)–(3–8d) and (3–8f) imply that $\bar{y}_{n'+1} = 1$, i.e., $(\bar{x}, \bar{y}) \in P$. Otherwise, let $k^\star$ denote the smallest index $i \in \{1, \ldots, n'\}$ such that $\bar{x}_i = 0$. From (3–8c), (3–8d), and (3–8f), $\bar{y}_i = 0$ for all $i \leq k^\star - 1$. From (3–8e), $\bar{y}_i = 0$ for all $i \geq k^\star + 1$ and thus (3–8b) implies $\bar{y}_{k^\star} = 1$. Therefore, $(\bar{x}, \bar{y}) \in P$, proving that (3–8) is a valid formulation of P.

Although formulation (3–8) is correct, it is not ideal in the sense that some of its vertices contain fractional variable values. For instance, when $n = 4$ and $B = n' = 2$, the solution $x = (1/2, 1/2, 0, 1)$ together with $y = (1/2, 0, 1/2)$ is an extreme point of $\bar{P}$. In the following section, we develop inequalities that are valid for P, prove that they define facets to conv(P), and show that adding them to (3–8) results in an ideal formulation. Towards this end, we first establish the dimensionality of P.

Lemma 3.1. The set P has dimension $n + n' - 1$.

Proof. Because $P \subseteq \mathbb{R}^n \times \mathbb{R}^{n'+1}$ and each $(x, y) \in P$ must satisfy the linearly independent equalities (3–8b) and (3–8c), it follows that

$$\dim(P) \leq (n + n' + 1) - 2 = n + n' - 1. \tag{3–10}$$

We show that $\dim(P) \geq n + n' - 1$ by providing $n + n'$ affinely independent points $c^1, \ldots, c^{n+n'} \in P$ of the form $(x_1, \ldots, x_n, y_1, \ldots, y_{n'+1})$. Let c^1 equal zero in all components except for the y_1-component, which equals one. For $i = 2, \ldots, n$, let c^i be given exactly as c^1 except for the x_i-component, which equals one. For $i = 1, \ldots, n'$, let c^{n+i} equal one in components $x_1, \ldots, x_i$, and y_{i+1}, and zero elsewhere. Noting that $n' \leq B$, each of the points $c^1, \ldots, c^{n+n'}$ are elements of P. Let C be the $(n + n' - 1) \times (n + n' + 1)$ matrix such that row $i = 1, \ldots, n + n' - 1$ is formed by $c^{i+1} - c^1$. Define $\widetilde{C}$ as the $(n + n' - 1) \times (n + n' - 1)$ matrix obtained by deleting the x_1- and y_1-columns from C. Because $\widetilde{C}$ is lower-triangular with nonzero diagonals, it follows that the rows of $\widetilde{C}$ and C are linearly independent. Therefore, the points $c^1, \ldots c^{n+n'}$ are affinely independent, and $\dim(P) \geq n + n' - 1$. □

3.3 Reformulation

In this section, we begin by finding all classes of facet-defining inequalities to the deterministic (one-scenario) C-BiPSNIP formulation in Section 3.3.1. We then show how this convex hull can be captured via the Special Structures Reformulation-Linearization Technique (SSRLT) [55] in Section 3.3.2. Finally, in Section 3.3.3, we show that the facet-defining inequalities derived in Section 3.3.1 actually generalize the more traditional step inequalities that have been proposed for BiPSNIP.

3.3.1 Convex Hull Formulation

Beginning with model (3–8) from the previous section, we now develop inequalities that lead to an ideal formulation. Our first class of inequalities exploit the relationship between y-variables to strengthen the linearization constraints (3–8e). We now describe these inequalities and show that they are both valid and facet-defining for $\text{conv}(P)$.

Theorem 3.1. For $j = 2, \ldots, n'$, the inequality

$$\sum_{i=j+1}^{n'+1} y_i \leq x_j, \tag{3–11}$$

is valid and facet-defining for P.

Proof. We first show that (3–11) is valid. Note that (3–11) is implied by (3–8b) when $x_j = 1$, and is a consequence of (3–8e) for $i = j+1, \dots, n'+1$ when $x_j = 0$.

To show that (3–11) is facet-defining, we demonstrate that the dimension of the face of P induced by (3–11) is $n + n' - 2$. We thus construct affinely independent points $c^1, \dots, c^{n+n'-1} \in P$ of the form $(x_1, \dots, x_n, y_1, \dots, y_{n'+1})$ that are binding on (3–11). Let c^1 equal one in the y_1-component and zero elsewhere. For $i = 2, \dots, j-1$, let c^i be given exactly as c^1 with the x_i-component changed from zero to one. For $i = j, \dots, n-1$, define c^i exactly as c^1, changing the x_{i+1}-component from zero to one. For $i = 1, \dots, n'$, let c^{n+i-1} equal one in components $x_1, \dots, x_i$, and y_{i+1}, and zero elsewhere. Let C denote the $(n + n' - 2) \times (n + n' + 1)$ matrix constructed such that row i of C is given by $c^{i+1} - c^1$. Let $\widetilde{C}$ be the matrix obtained by removing the x_1-column, the x_j-column, and the y_1-column from C, and observe that $\widetilde{C}$ is lower-triangular with nonzero diagonals. Thus, $\widetilde{C}$ is nonsingular and $c^1, \dots, c^{n+n'-1}$ are affinely independent, as required. □

Remark 3.3. Inequality (3–11) is valid for $j = 1$ as well; however, it is implied (in the continuous sense) by (3–8b) and (3–8c), and is therefore not facet-defining. □

We now characterize some inequalities that capture subtle relationships among the x- and y-variables caused by inclusion of the cardinality constraint (3–8a). Let $\Phi = \{\phi_1, \dots, \phi_{|\Phi|}\}$ and $\Psi = \{\psi_1, \dots, \psi_{|\Psi|}\}$ be subsets of $\{2, \dots, n'\}$ and $\{n'+1, \dots, n\}$, respectively, such that

$$\phi_1 < \cdots < \phi_{|\Phi|} < \psi_1 < \cdots < \psi_{|\Psi|}. \tag{3–12}$$

Define $I(\Phi) = \{1, \dots, |\Phi|\}$ and $I(\Psi) = \{1, \dots, |\Psi|\}$, and for each $i = 1, \dots, n'+1$, define $\Phi_i = \{j = 2, \dots, i-1 : j \in \Phi\}$. Using this notation, we define the following inequality:

$$\sum_{i \in I(\Phi)} x_{\phi_i} + \sum_{i \in I(\Psi)} x_{\psi_i} \leq \sum_{i=1}^{n'+1} \min\{|\Phi| + |\Psi|,\ B - i + 1 + |\Phi_i|\} y_i. \tag{3–13}$$

Similar inequalities could be obtained by allowing the inclusion of 1 in Φ, but these can be obtained from (3–13) by substituting x_1 for $y_2 + \cdots + y_{n'+1}$. We now prove conditions

under which (3–13) is valid and facet-defining, and then show how to separate (3–13) in polynomial time.

Theorem 3.2. Let $\Phi \subseteq \{2, \dots, n'\}$ and $\Psi \subseteq \{n'+1, \dots, n\}$ satisfy (3–12). Then (3–13) is valid for P.

Proof. Let $(\bar{x}, \bar{y}) \in P$, and let $k^\star$ be the index for which $\bar{y}_{k^\star} = 1$. Note that $k^\star$ exists, and $\bar{y}_i = 0$ for all $i \neq k^\star$, by (3–8b). If $|\Phi| + |\Psi| < B - k^\star + 1 + |\Phi_{k^\star}|$, then the right-hand side of (3–13) becomes $|\Phi| + |\Psi|$, and (3–13) is valid because each x-variable is no more than 1.

Otherwise, suppose that $|\Phi| + |\Psi| \geq B - k^\star + 1 + |\Phi_{k^\star}|$. Note that because $\bar{y}_{k^\star} = 1$, (3–8b)–(3–8e) imply that $\bar{x}_1 = \cdots = \bar{x}_{k^\star - 1} = 1$ and either $\bar{x}_{k^\star} = 0$ or $k^\star = n' + 1$. Then (3–13) simplifies to the following inequality:

$$\sum_{i=|\Phi_{k^\star}|+1}^{|\Phi|} \bar{x}_{\phi_i} + \sum_{i=1}^{|\Psi|} \bar{x}_{\psi_i} \leq B - k^\star + 1. \tag{3–14}$$

However, (3–14) is implied by (3–8a) and the fact that $\bar{x}_1 = \cdots = \bar{x}_{k^\star - 1} = 1$. □

Theorem 3.3. Let Φ and Ψ be given as in Theorem 3.2. Then (3–13) is facet-defining for conv(P) if and only if (a) $B - n' + 1 \leq |\Phi| + |\Psi| \leq B - 1$ and (b) if $\Phi \neq \emptyset$ then $\phi_1 \geq B - |\Phi| - |\Psi| + 3$.

Proof. We first prove that (3–13) defines a face of conv(P) of dimension at least $n + n' - 2$ if conditions (a) and (b) hold. We specify $n + n' - 1$ affinely independent points $c^1, \dots, c^{n+n'-1} \in P$ of the form $(x_1, \dots, x_n, y_1, \dots, y_{n'+1})$ that are binding on (3–13). In the following, we show that the given points are feasible by demonstrating that the y components are feasible given values for the x-components (i.e., that (3–8b)–(3–8g) is satisfied), and that the sum of one-valued x-components does not exceed B. We then prove that the specified classes of points are indeed binding on (3–13).

Vector c^1 has a value of one in components y_1, x_{ϕ_j} for all $j \in I(\Phi)$, and x_{ψ_j} for all $j \in I(\Psi)$, and zeros elsewhere. Note that neither Φ nor Ψ may contain index 1, so c^1

must satisfy $x_1 = 0$. Because c^1 has $|\Phi| + |\Psi| \leq B - 1$ x-variables equal to 1, c^1 is an element of P. Furthermore, because $|\Phi_1| = 0$ and $|\Phi| + |\Psi| \leq B - 1$, the right-hand side of (3–13) is $|\Phi| + |\Psi|$, and the left-hand side of (3–13) matches this value because all $|\Phi| + |\Psi|$ variables on the left-hand side of (3–13) equal one.

For $i = 2, \ldots, n$ such that $i \notin \Phi \cup \Psi$, let c^i be defined exactly as c^1 except for a one in component x_i. By the same argument as before, each of these points is in P and each is on the face defined by (3–13).

For $i = 2, \ldots, n$ such that $i \in \Phi \cup \Psi$, we again define c^i exactly as c^1, except that component x_i equals zero instead of one, components $x_1, \ldots, x_{B-|\Phi|-|\Psi|+1}$ equal one, component y_1 is zero, and component $y_{B-|\Phi|-|\Psi|+2}$ equals one. (Note that $B - |\Phi| - |\Psi| + 1 \leq n'$ by condition (a).) To see that each such point is feasible, observe that $\{1, \ldots, B - |\Phi| - |\Psi| + 2\} \cap (\Phi \cup \Psi) = \emptyset$ by assumption (b). Because the first $B - |\Phi| - |\Psi| + 1$ x-variables equal one in this point, while $x_{B-|\Phi|-|\Psi|+2}$ equals zero, we have that (3–8b)–(3–8g) implies $y_{B-|\Phi|-|\Psi|+2} = 1$ as required. Also, all but one of the variables corresponding to indices in $\Phi \cup \Psi$ equal one, and so a total of $(B - |\Phi| - |\Psi| + 1) + (|\Phi| + |\Psi| - 1) = B$ x-variables equal one. Hence, c^i represents a point in P. To see that (3–13) is binding on this point, note that $B - (B - |\Phi| - |\Psi| + 2) + 1 + |\Phi_{B-|\Phi|-|\Psi|+2}| = |\Phi| + |\Psi| - 1$. The right-hand side of (3–13) thus reduces to $|\Phi| + |\Psi| - 1$, and because all but one x-variable on the left-hand side of (3–13) equals one in point c^i, the inequality is binding at this point.

For $i = 1, \ldots, B - |\Phi| - |\Psi|$, define c^{n+i} exactly as c^1, except that components $x_1, \ldots, x_i$ and y_{i+1} all equal one, while component y_1 equals zero. The feasibility of these points is established using the same argument as above, with the observation that the total number of one-valued x-variables is equal to $i + |\Phi| + |\Psi| \leq (B - |\Phi| - |\Psi|) + |\Phi| + |\Psi| = B$. Noting that $y_{i+1} = 1$ for these points, the right-hand side of (3–13) is $\min\{|\Phi| + |\Psi|, B - (i + 1) + 1 + |\Phi_{i+1}|\}$. Noting that $\Phi_{i+1} = \emptyset$ and $B - i \geq B - (B - |\Phi| - |\Psi|) = |\Phi| + |\Psi|$, the right-hand side of (3–13) equals $|\Phi| + |\Psi|$.

Therefore, (3–13) is binding on c^{n+i} because all x_i components equal one corresponding to $i \in \Phi \cup \Psi$.

For each $i = B - |\Phi| - |\Psi| + 1, \ldots, n' - 1$, define c^{n+i} such that components $x_1, \ldots, x_{i+1}$ and y_{i+2} equal one, along with a few others specified below. Thus far, the points clearly belong to P, noting that $i + 1 \leq n' \leq B$. We now show how to modify these points so that they become binding on (3–13) while remaining in P. The right-hand side of (3–13) equals $\min\{|\Phi| + |\Psi|, B - (i + 2) + 1 + |\Phi_{i+2}|\}$, but noting that the second term is no more than $B - (B - |\Phi| - |\Psi| + 3) + 1 = |\Phi| + |\Psi| - 2$, the right-hand side evaluates to $B - i - 1 + |\Phi_{i+2}|$. The contribution to the left-hand side of (3–13) from the x-variables thus far equals $|\Phi_{i+2}|$, and we therefore seek to set $B - i - 1$ more x-variables equal to one. (In doing so, the total number of x-components set to one becomes $(i + 1) + (B - i - 1) = B$, and (3–8a) remains satisfied.) We therefore set the x-components corresponding to the $B - i - 1$ *highest-indexed* elements of $\Phi \cup \Psi$ equal to 1. Note that if the $(B - i - 1)$st highest element of $\Phi \cup \Psi$ is at least as large as $i + 3$, then the x_{i+2} component remains equal to zero (and so setting the y_{i+2} component equal to one satisfies (3–8b)–(3–8g)), and the additional components set to one are all distinct from indices $\{1, \ldots, i + 1\}$, thus ensuring that (3–13) becomes binding on this point. To establish the condition that the $(B - i - 1)$st highest element of $\Phi \cup \Psi$ is at least as large as $i + 3$, choose $\delta \geq 1$ so that $i = B - |\Phi| - |\Psi| + \delta$. By assumption (b), the $(|\Phi| + |\Psi|)$th highest element of $\Phi \cup \Psi$ is not less than $B - |\Phi| - |\Psi| + 3$ if $\Phi \neq \emptyset$. (If $\Phi = \emptyset$, the claim holds because all elements of $\Phi \cup \Psi$ are greater than n'.) Noting that $B - i - 1 = |\Phi| + |\Psi| - (1 + \delta)$, the $(B - i - 1)$st highest element of $\Phi \cup \Psi$ is at least $B - |\Phi| - |\Psi| + 4 + \delta = i + 4$.

Let C be the $(n + n' - 2) \times (n + n' + 1)$ matrix such that each row $i = 1, \ldots, n + n' - 2$ is defined by $c^{i+1} - c^1$. Let $\widetilde{C}$ be the matrix obtained by eliminating columns x_1, y_1, and $y_{B-|\Phi|-|\Psi|+2}$ from C, and observe that $\widetilde{C}$ is a lower-triangular matrix with nonzero

diagonals. Thus, the rows of $\widetilde{C}$ and C are linearly independent and $c^1, \dots, c^{n+n'-1}$ are affinely independent, proving that (3–13) is facet-defining.

We now prove that (3–13) defines a facet to $\text{conv}(P)$ only if the assumptions of the theorem are satisfied. First suppose that $\Phi \neq \emptyset$ yet $\phi_1 \leq B - |\Phi| - |\Psi| + 2$, so that assumption (b) fails. Since $B - i + 1 \geq |\Phi| + |\Psi|$ for all $i \leq \phi_1 - 1$ and $|\Phi_i| = 0$ for all $i \leq \phi_1$, inequality (3–13) can be rewritten as

$$\sum_{i=1}^{|\Phi|} x_{\phi_i} + \sum_{i=1}^{|\Psi|} x_{\psi_i} \leq \sum_{i=1}^{\phi_1 - 1} (|\Phi| + |\Psi|) y_i + \min\{|\Phi| + |\Psi|, B - \phi_1 + 1\} y_{\phi_1} + \sum_{i=\phi_1+1}^{n'+1} \min\{|\Phi| + |\Psi|, B - i + 1 + |\Phi_i|\} y_i. \tag{3–15}$$

If $\phi_1 \leq B - |\Phi| - |\Psi| + 1$ the coefficient on y_{ϕ_1} becomes $|\Phi| + |\Psi|$. However, changing the coefficient on y_{ϕ_1} to $|\Phi| + |\Psi| - 1$ results in a strengthened valid inequality since $y_{\phi_1} = 1$ forces $x_{\phi_1} = 0$, meaning the left-hand side can be at most $|\Phi| + |\Psi| - 1$. Thus, (3–13) is not facet-defining in this case. Suppose instead that $\phi_1 = B - |\Phi| - |\Psi| + 2$. In this case, (3–15) further simplifies to

$$\sum_{i=1}^{|\Phi|} x_{\phi_i} + \sum_{i=1}^{|\Psi|} x_{\psi_i} \leq \sum_{i=1}^{\phi_1 - 1} (|\Phi| + |\Psi|) y_i + (|\Phi| + |\Psi| - 1) y_{\phi_1} + \sum_{i=\phi_1+1}^{n'+1} (B - i + 1 + |\Phi_i|) y_i. \tag{3–16}$$

We show that (3–16) is implied by a combination of other valid inequalities. Consider generating (3–13) using $\bar{\Phi} = \{\phi_2, \dots, \phi_{|\Phi|}\}$ and $\bar{\Psi} = \Psi$, and defining $\bar{\Phi}_i = \{j \in \bar{\Phi} : j < i\}$. The resulting inequality,

$$\sum_{i=2}^{|\Phi|} x_{\phi_i} + \sum_{i=1}^{|\Psi|} x_{\psi_i} \leq \sum_{i=1}^{n'+1} \min\{|\Phi| + |\Psi| - 1,\ B - i + 1 + |\bar{\Phi}_i|\} y_i, \tag{3–17}$$

is valid by Theorem 3.2. Noting that $\bar{\Phi}_i = \Phi_i = \emptyset$ for all $i = 1, \dots, \phi_1$ and $\bar{\Phi}_i = \Phi_i \setminus \{\phi_1\}$ for all $i = \phi_1 + 1, \dots, n' + 1$ and recalling that $\phi_1 = B - |\Phi| - |\Psi| + 2$, (3–17) becomes

$$\sum_{i=2}^{|\Phi|} x_{\phi_i} + \sum_{i=1}^{|\Psi|} x_{\psi_i} \leq \sum_{i=1}^{\phi_1} (|\Phi| + |\Psi| - 1) y_i + \sum_{i=\phi_1+1}^{n'+1} (B - i + |\Phi_i|) y_i. \tag{3–18}$$

Adding (3–18) to

$$x_{\phi_1} \leq \sum_{i=1}^{\phi_1-1} y_i + \sum_{i=\phi_1+1}^{n'+1} y_i, \tag{3–19}$$

which is valid by (3–8b) and (3–8d), produces (3–16); therefore, (3–13) is not facet defining.

Now suppose that condition (b) holds but condition (a) fails. If $|\Phi| + |\Psi| < B - n' + 1$, then $\Phi = \emptyset$ by condition (b), the minimum term in (3–13) always evaluates to $|\Psi|$, and the inequality becomes $\sum_{i=1}^{|\Psi|} x_{\phi_i} \leq |\Psi| \sum_{i=1}^{n'+1} y_i$, which is implied by (3–8b) and $x_i \leq 1, \ \forall i = 1, \dots, n$. Suppose instead $|\Phi| + |\Psi| \geq B$. Then $|\Phi| + |\Psi| \geq B - i + 1$ for all $i = 1, \dots, n' + 1$, and (3–13) reduces to

$$\sum_{i=1}^{|\Phi|} x_{\phi_i} + \sum_{i=1}^{|\Psi|} x_{\psi_i} \leq \sum_{i=1}^{n'+1} (B - i + 1 + |\Phi_i|)\, y_i. \tag{3–20}$$

We show that (3–20) is implied by a combination of other inequalities already proven to be valid for P. Substituting (3–8b) and (3–8c) produces the following form of the budget constraint (3–8a):

$$\sum_{i=2}^{n} x_i \leq B y_1 + (B-1) \sum_{i=2}^{n'+1} y_i. \tag{3–21}$$

For $i = 1, \dots, n' + 1$, define $\Phi_i^c = \{2, \dots, i-1\} \setminus \Phi_i$. Multiplying by -1 each inequality from (3–11) corresponding to $j \in \{2, \dots, n'\}$ such that $j \notin \Phi$ and summing with (3–21) yields

$$\begin{aligned} \sum_{i=1}^{|\Phi|} x_{\phi_i} + \sum_{i=n'+1}^{n} x_i &\leq B y_1 + \sum_{i=2}^{n'+1} (B - 1 - |\Phi_i^c|)\, y_i \\ &= B y_1 + \sum_{i=2}^{n'+1} (B - i + 1 + |\Phi_i|)\, y_i, \end{aligned} \tag{3–22}$$

where (3–22) follows because $|\Phi_i^c| = (i-2) - |\Phi_i|$. Noting that $\Phi_1 = \emptyset$, (3–22) implies (3–20); thus, (3–13) cannot define a facet. □

Formulation (3–8) can now be tightened by adding inequalities (3–11) and (3–13). Doing so allows some of the other constraints in (3–8) to be removed, and so we now

summarize the updated model. Define the updated polytope

$$P' = \left\{ (x, y) \in \mathbb{R}^n \times \mathbb{R}^{n'+1} \;\middle|\; \begin{array}{l} (3\text{–}8a)\text{–}(3\text{–}8d),\ (3\text{–}8f),\ (3\text{–}8g) \\ (3\text{–}11),\ \forall j = 2, \dots, n' \\ (3\text{–}13),\ \forall (\Phi, \Psi) \in \Gamma \end{array} \right\}, \tag{3–23}$$

where Γ is the set of all (Φ, Ψ) satisfying the assumptions of Theorem 3.3. We now prove that this is a valid and ideal formulation.

Theorem 3.4. System (3–23) is a valid formulation of P, i.e., $P' \cap (\mathbb{Z}^n \times \mathbb{Z}^{n'+1}) = P$.

Proof. Let $(\bar{x}, \bar{y}) \in P$. Constraints (3–8a)–(3–8d), (3–8f), and (3–8g) are obviously satisfied, and constraints (3–11) and (3–13) must also be satisfied due to Theorems 3.1 and 3.2. Thus, $(\bar{x}, \bar{y}) \in P'$ and $P \subseteq P' \cap (\mathbb{Z}^n \times \mathbb{Z}^{n'+1})$.

Let $(\bar{x}, \bar{y}) \in P' \cap (\mathbb{Z}^n \times \mathbb{Z}^{n'+1})$. In order to show that $(\bar{x}, \bar{y}) \in P$, we need only show that constraints (3–8e) are satisfied, as the remainder of the constraints of (3–8) are part of the definition of P'. Constraints (3–8e) are implied by (3–11) because the y-variables are nonnegative. □

Theorem 3.5. System (3–23) is an ideal formulation of P, i.e., $P' = \text{conv}(P)$.

Proof. Let c be any n-vector and d be any $(n' + 1)$-vector such that at least one component of c or d is nonzero, and define

$$f(x, y) = \sum_{j=1}^{n} c_j x_j + \sum_{i=1}^{n'+1} d_i y_i. \tag{3–24}$$

We show that $F \equiv \text{argmax}_{(x,y)} \{f(x, y) : (x, y) \in P\}$ is contained in one of the faces of P', thus proving the result.

Without loss of generality, we may assume that $c_1 = d_1 = 0$, using (3–8b) and (3–8c) to eliminate x_1 and y_1 from the objective if necessary. Let $(\bar{x}, \bar{y})$ be any element of F and suppose that $c_j < 0$ for some $j \in \{2, \dots, n\}$. If $j \leq n'$, it follows that $\bar{x}_j = 1$ only if $\bar{y}_i = 1$ for some $i > j$ (or else a better solution could be obtained by changing $\bar{x}_j$ to zero); hence, (3–11) corresponding to j is binding at $(\bar{x}, \bar{y})$. If $j \geq n' + 1$, any solution with $x_j = 1$

is dominated by the feasible solution obtained by changing x_j to zero; thus, $(\bar{x}, \bar{y})$ must be in the face of P' corresponding to $x_j \geq 0$. Therefore, if $c_j < 0$ for some $j \in \{2, \dots, n\}$, it follows that F is contained in a face of P'.

Next, suppose that $F \cap \{(x, y) \in P : y_i = 1\} = \emptyset$ for some $i \in \{1, \dots, n' + 1\}$. In this case, each point in F is binding on one of the faces (3–8f). Thus, for the remainder of this proof, we assume that $c_j \geq 0, \ \forall j = 2, \dots, n$, and that c and d are such that $F \cap \{(x, y) \in P : y_i = 1\} \neq \emptyset$ for each $i = 1, \dots, n' + 1$. Let $\Lambda = \{\lambda_1, \dots, \lambda_N\}$ denote the set of indices $j \in \{2, \dots, n\}$ such that $c_j > 0$, with $\lambda_1 < \lambda_2 < \cdots < \lambda_N$. If $\Lambda = \emptyset$ then $d_i \neq 0$ for some $i = 2, \dots, n' + 1$ or else $(c, d) = 0$; however, if $d_i > 0$ then $F \cap \{(x, y) \in P : y_1 = 1\} = \emptyset$, and if $d_i < 0$ then $F \cap \{(x, y) \in P : y_i = 1\} = \emptyset$. Therefore, we now restrict our analysis to the case in which $\Lambda \neq \emptyset$.

Define $\Lambda_i = \Lambda \cap \{2, \dots, i - 1\}$ and $\bar{\Lambda}_i = \Lambda \cap \{i + 1, \dots, n\}$, for all $i = 1, \dots, n'$, and define $\Lambda_{n'+1} = \Lambda \cap \{2, \dots, n'\}$ and $\bar{\Lambda}_{n'+1} = \bar{\Lambda}_{n'}$. Accordingly, if $(\bar{x}, \bar{y}) \in F$ with $\bar{y}_{k^\star} = 1$, then $\bar{x}_\lambda = 1, \ \forall \lambda \in \Lambda_{k^\star}$, and as many $\bar{x}_\lambda, \ \lambda \in \bar{\Lambda}_{k^\star}$, as allowed by (3–8a) should be set to one, i.e.,

$$\sum_{\lambda \in \bar{\Lambda}_{k^\star}} \bar{x}_\lambda = \min\{|\bar{\Lambda}_{k^\star}|, B - k^\star + 1\}. \tag{3–25}$$

Let $(\bar{x}, \bar{y}) \in F$ and let $k^\star \in \{1, \dots, n' + 1\}$ denote the index such that $\bar{y}_{k^\star} = 1$. If $1 \leq N \leq B - n'$, then we prove this claim in two cases: (i) $\lambda_1 \in \Lambda_{n'+1}$, in which we show that (3–8d) is binding on all points in F, and (ii) $\Lambda_{n'+1} = \emptyset$, in which we show that (3–8g) is binding on all points in F. In case (i), if $k^\star \geq \lambda_1$ then (3–11) and (3–8d) imply that $\bar{x}_{\lambda_1} + \bar{y}_{\lambda_1} = 1$. If instead $k^\star \leq \lambda_1 - 1$, then $B - k^\star + 1 \geq B - n' \geq N \geq |\bar{\Lambda}_{k^\star}|$. From (3–25), $\bar{x}_{\lambda'} = 1$ for all $\lambda' \in \bar{\Lambda}_{k^\star}$; thus, $\bar{x}_{\lambda_1} = 1$ and $\bar{x}_{\lambda_1} + \bar{y}_{\lambda_1} = 1$, proving that (3–8d) is binding on F. In case (ii), $\lambda_1 \in \bar{\Lambda}_{n'+1}$. Noting that $k^\star \leq n' + 1$, observe that $B - k^\star + 1 \geq B - n' \geq N \geq |\bar{\Lambda}_{k^\star}|$, and it follows from (3–25) that $\bar{x}_{\lambda'} = 1$ for all $\lambda' \in \bar{\Lambda}_{k^\star}$. This implies that $x_{\lambda_1} \leq 1$ is binding on all of F. Thus, we may now assume that $N \geq B - n' + 1$.

Now suppose that $N \geq B$. We show this implies that (3–8a) is binding on every element of F. Let $(\bar{x}, \bar{y}) \in F$, and let $k^\star \in \{1, \dots, n' + 1\}$ be the index such that $\bar{y}_{k^\star} = 1$.

This implies that $\bar{y}_i = 0$ for all $i \neq k^\star$, and $\bar{x}_1 = \cdots = \bar{x}_{k^\star-1} = 1$. With only $(k^\star - 1)$ x-variables fixed to one thus far, the cardinality constraint (3–8a) permits as many as $B - k^\star + 1$ more x-variables to take value one. Noting that $|\bar{\Lambda}_i| \geq |\Lambda| - |\{2, \dots, i\}| = N - i + 1 \geq B - i + 1$ for $i = 1, \dots, n'$, and $\bar{\Lambda}_{n'+1} = \bar{\Lambda}_{n'}$, it follows from (3–25) that $\sum_{\lambda \in \bar{\Lambda}_{k^\star}} \bar{x}_\lambda = B - k^\star + 1$; thus $\sum_{i=1}^{n} \bar{x}_i = B$ and (3–8a) must be binding. We henceforth assume that $N \leq B - 1$.

Next, suppose that $\lambda_1 \leq \min\{n' + 1, B - N + 2\}$. Let $(\bar{x}, \bar{y}) \in F$, and let $k^\star \in \{1, \dots, n' + 1\}$ denote the index such that $\bar{y}_{k^\star} = 1$. If $k^\star \geq \lambda_1$, then (3–8d) (corresponding to $\lambda_1 \leq n'$) and (3–11) imply that $\bar{x}_{\lambda_1} + \bar{y}_{\lambda_1} = 1$. If instead $k^\star \leq \lambda_1 - 1$, observe that $\bar{y}_{\lambda_1} = 0$ and $B - k^\star + 1 \geq B - \lambda_1 + 2 \geq N \geq |\bar{\Lambda}_{k^\star}|$, and from (3–25), $\bar{x}_{\lambda_1} = 1$ and (3–8d) is binding on all of F.

Thus far, we have eliminated all cases except those for which (a) $B - n' + 1 \leq N \leq B - 1$ and (b) either $\lambda_1 \geq n' + 1$ or $\lambda_1 \geq B - N + 3$. Using $\Phi = \Lambda_{n'+1}$ and $\Psi = \bar{\Lambda}_{n'+1}$, conditions (a) and (b) above identically match the assumptions of Theorem 3.3, therefore indicating that $(\Lambda_{n'+1}, \bar{\Lambda}_{n'+1}) \in \Gamma$ and the corresponding constraint (3–13) defines a facet of P'. This inequality is given by

$$\sum_{i=1}^{N} x_{\lambda_i} \leq \sum_{i=1}^{n'+1} \min\{N, B - i + 1 + |\Lambda_i|\} y_i. \tag{3–26}$$

We now show that F is contained in the face of P' associated with (3–26). Let $(\bar{x}, \bar{y}) \in F$ and let $k^\star \in \{1, \dots, n'\}$ be the index such that $\bar{y}_{k^\star} = 1$. We prove this claim in two cases: (i) $k^\star \leq B - N + 1$ and (ii) $k^\star \geq B - N + 2$. In case (i), observe that $k^\star \leq B - N + 1 \leq n'$ and either $\lambda_1 \geq n' + 1 \geq k^\star + 1$ or $\lambda_1 \geq B - N + 3 \geq k^\star + 2$. In either case, we find that $\Lambda_{k^\star} = \emptyset$, and (3–26) reduces to

$$\sum_{\lambda \in \bar{\Lambda}_{k^\star}} \bar{x}_\lambda \leq N. \tag{3–27}$$

Moreover, $B - k^\star + 1 \geq N \geq |\bar{\Lambda}_{k^\star}|$ combined with (3–25) implies that the left-hand side of (3–27) is equal to N; thus, (3–27) is binding on F. In case (ii), $B - k^\star + 2 \leq N$, and after

canceling $|\Lambda_{k^\star}|$ from each side, (3–26) becomes

$$\sum_{\lambda \in \bar{\Lambda}_{k^\star}} \bar{x}_\lambda \leq B - k^\star + 1. \tag{3–28}$$

We prove (3–28) is satisfied as an equality in subcases: (ii.a) $\lambda_1 \geq n' + 1$ and (ii.b) $\lambda_1 \geq B - N + 3$. In case (ii.a), $|\bar{\Lambda}_{k^\star}| = N \geq B - k^\star + 2$; thus, $(\bar{x}, \bar{y}) \in F$ combined with (3–25) implies that the left-hand side of (3–28) equals $B - k^\star + 1$. In case (ii.b), if $k^\star = n' + 1$, we find that $|\bar{\Lambda}_{k^\star}| \geq |\Lambda \cap \{B - N + 3, \dots, n\}| - |\{B - N + 3, \dots, n'\}| = N - [n' - (B - N + 3) + 1] = B - n' + 2 = B - k^\star + 3$. If instead $k^\star \leq n'$, we find that $|\bar{\Lambda}_{k^\star}| \geq |\Lambda \cap \{B - N + 3, \dots, n\}| - |\{B - N + 3, \dots, k^\star\}| = N - [k^\star - (B - N + 3) + 1] = B - k^\star + 2$. In either case, $|\bar{\Lambda}_{k^\star}| \geq B - k^\star + 2$, proving that $\sum_{\lambda \in \bar{\Lambda}_{k^\star}} \bar{x}_\lambda = B - k^\star + 1$ and (3–28) is binding on F. □

Remark 3.4. Each inequality in (3–23) is facet-defining for $\text{conv}(P)$, and thus, Theorem 3.5 implies (3–23) is a minimal formulation of $\text{conv}(P)$. Proofs that the inequalities in (3–23) carried over from (3–8) define facets are straightforward and therefore omitted. □

Because there are an exponential number of Φ- and Ψ-sets that induce facets of the form (3–13), enumerating these inequalities is impractical. In the following paragraphs, we show how to generate, in polynomial time, a most-violated inequality (3–13) given $(\bar{x}, \bar{y}) \in \bar{P} \setminus P'$.

Let $k \in \{B - n' + 1, \dots, B - 1\}$ and define $\Gamma_k = \{(\Phi, \Psi) \in \Gamma : |\Phi| + |\Psi| = k\}$. We describe a process for identifying $(\Phi^\star, \Psi^\star)$ that maximizes the violation of (3–13) at $(\bar{x}, \bar{y})$ over all sets $(\Phi, \Psi) \in \Gamma_k$; thus, a most-violated (3–13) can be found by repeating this process for each $k = B - n' + 1, \dots, B - 1$.

We first define the quantity

$$\eta = \sum_{j=1}^{n'+1} \min\{k, B - j + 1\} \bar{y}_j, \tag{3–29}$$

which represents a lower bound on the right-hand side of (3–13) evaluated at $(\bar{x}, \bar{y})$. This is a lower bound because $B - j + 1 \leq B - j + 1 + |\Phi_j|$ for all $j = 1, \dots, n' + 1$. Note that the lower bound is tight when $\Phi = \emptyset$. We now describe how inclusion of different indices in $\Phi \cup \Psi$ causes the right-hand side of (3–13) to increase. For $j = 2, \dots, n$, we define β_j as the increase to the right-hand side of (3–13) due to including j in $\Phi \cup \Psi$. Since inclusion of j in $\Phi \cup \Psi$ is prevented by condition (b) of Theorem 3.3 for each $j = 2, \dots, \min\{n', B - k + 2\}$, define $\beta_j = \infty$ for these indices.

For $j = B-k+3, \dots, n'$ and $i = j+1, \dots, n'+1$, it follows that $\Phi_i \subseteq \{B-k+3, \dots, i-1\}$; thus, $B - i + 1 + |\Phi_i| \leq k - 2$ for all such i, i.e., the minimum in the i-th term of (3–13) is given by $B - i + 1 + |\Phi_i|$. Noting that for these values of i, the minimum term in (3–29) is given by $B - i + 1$, including j in $\Phi \cup \Psi$ increases the coefficient on $\bar{y}_i$ in (3–13) by exactly one for each $i = j + 1, \dots, n' + 1$; accordingly, define $\beta_j = \sum_{i=j+1}^{n'+1} \bar{y}_i$.

For $j = n' + 1, \dots, n$ including j in $\Phi \cup \Psi$ changes only the left-hand side of (3–13), so define $\beta_j = 0$. As such, inequality (3–13) evaluated at $(\bar{x}, \bar{y})$ reduces to

$$\sum_{j \in \Phi \cup \Psi} (\bar{x}_j - \beta_j) \leq \eta. \tag{3–30}$$

From (3–30), it is apparent that the effect on the violation of including each alternative in $\Phi \cup \Psi$ is independent of all the other alternatives; hence, (3–30) can be maximized by choosing the k alternatives in $\{2, \dots, n\}$ with the greatest $(\bar{x}_j - \beta_j)$-values. This can be accomplished using a polynomial-time sorting algorithm.

3.3.2 Obtaining conv(P) through SSRLT

In the previous sections, we derived a minimal representation of $P' = \text{conv}(P)$. Alternatively, conv(P) could have been obtained from $\bar{P}$ using the Reformulation-Linearization Technique (RLT) of Sherali and Adams [56, 57] or similar methods (e.g., Balas et al. [58] and Lovász and Schrijver [59]) that would lift $\bar{P}$ into higher dimension in order to construct a tighter representation of conv(P) and then project out the added variables.

We now prove a result that allows the use of a specialized RLT to achieve the same result.

Lemma 3.2. Suppose $\bar{y} \in \{0,1\}^{n'+1}$, and define $Q(\bar{y})$ as the set of all x such that $(x, \bar{y}) \in \bar{P}$. Then all extreme points of $Q(\bar{y})$ are integer-valued.

Proof. For fixed $\bar{y} \in \{0,1\}^{n'+1}$, (3–8) reduces to a linear system of the form

$$\begin{bmatrix} e^T \\ I \\ -I \end{bmatrix} x \leq \begin{bmatrix} B \\ U \\ L \end{bmatrix}, \tag{3–31}$$

where e is the n-vector consisting of all ones, and U and L are n-vectors containing integral upper and lower bounds, respectively, on the x-variables. Since each element of e is equal to one, e^T is totally unimodular. Appending I and $-I$ to e^T preserves total unimodularity, resulting in the constraint matrix of (3–31). Since B, U, and L are integral, the desired result holds. □

From Lemma 3.2 it follows that $\text{conv}(P) = \text{conv}[\bar{P} \cap (\mathbb{R}^n \times \mathbb{Z}^{n'+1})]$. Thus, convexifying $\bar{P}$ via RLT only requires us to ensure that the y-variables are binary-valued at all extreme points. This convexification step can be performed using the Special Structures RLT (SSRLT) of Sherali et al. [55], by exploiting the structure of constraint (3–8b). We now describe how inequalities (3–11) and (3–13) can be derived through SSRLT.

Following the techniques described in [55], we: (a) multiply the constraints of (3–8) by y_k for each $k = 1, \ldots, n'+1$ and $x_j \geq 0$ by $1 - \sum_{k=1}^{n'+1} y_k$ for all $j = 1, \ldots, n'+1$ (where the latter set of operations yield equality constraints); (b) apply the identities $y_k^2 = y_k$ for all $k = 1, \ldots, n'+1$, $y_k x_k = 0$ for all $k = 1, \ldots, n'$, $y_j y_k = 0$ for all $j \neq k$, and $y_k x_j = y_k$ for $j < k$; and, (c) define the variable z_{kj} to linearize the products $y_k x_j$, for all $k = 1, \ldots, n'+1$

and $j = \min\{n'+1, k+1\}, \dots, n$. This procedure results in the following constraints:

$$\sum_{j=k+1}^{n} z_{kj} \leq (B-k+1)y_k, \ \forall k = 1, \dots, n'+1, \tag{3–32a}$$

$$x_j = \sum_{k=1}^{j-1} z_{kj} + \sum_{k=j+1}^{n'+1} y_k, \ \forall j = 2, \dots, n', \tag{3–32b}$$

$$x_j = \sum_{k=1}^{n'+1} z_{kj}, \ \forall j = n'+1, \dots, n, \tag{3–32c}$$

$$0 \leq z_{kj} \leq y_k, \ \forall k = 1, \dots, n'+1, \ \forall j = \min\{n'+1, k+1\}, \dots, n. \tag{3–32d}$$

Constraints (3–32a) are obtained by multiplying (3–8a) by y_k, (3–32b) and (3–32c) are derived by multiplying $1 - \sum_{k=1}^{n'+1} y_k$ by $x_j \geq 0$ for all $j = 2, \dots, n$, and (3–32d) result from the multiplication of y_k with (3–8d), (3–8e), and (3–8g). The remaining SSRLT inequalities are implied by constraints in (3–8) or (3–32) and are therefore omitted.

Define $\bar{P}^Z$ as the set of all (x, y, z) satisfying (3–32) such that $(x, y) \in \bar{P}$. The convergence result of [55] implies that $\text{conv}(P) = \{(x, y) : (x, y, z) \in \bar{P}^Z\}$. That is, projecting the z-variables out of $\bar{P}^Z$ produces all of the inequalities in (3–23). In particular, since $z_{kj} \geq 0$ the z-variables can be removed from (3–32b) if the equality is changed to an inequality, directly resulting in (3–11).

To obtain inequalities (3–13), let Φ and Ψ satisfy the assumptions of Theorem 3.3. Aggregate the equalities (3–32b) and (3–32c) corresponding to $j \in \Phi \cup \Psi$ to obtain (3–33a), which produces (3–33b) upon reordering summations:

$$\sum_{i \in I(\Phi)} x_{\phi_i} + \sum_{i \in I(\Psi)} x_{\psi_i} = \sum_{j \in \Phi} \left(\sum_{k=1}^{j-1} z_{kj} + \sum_{k=j+1}^{n'+1} y_k \right) + \sum_{j \in \Psi} \left(\sum_{k=1}^{n'+1} z_{kj} \right) \tag{3–33a}$$

$$= \sum_{k=1}^{n'+1} \left(\sum_{j \in \Phi : j > k} z_{kj} + \sum_{j \in \Psi} z_{kj} + |\Phi_k| y_k \right) \tag{3–33b}$$

$$\leq \sum_{k=1}^{n'+1} \left(\min\left\{ |\{j \in \Phi : j > k\}| + |\Psi|, B-k+1 \right\} + |\Phi_k| \right) y_k. \tag{3–33c}$$

The quantity $\sum_{j\in\Phi:j>k} z_{kj} + \sum_{j\in\Psi} z_{kj}$ must be bounded above by both $(B-k+1)y_k$ from (3–32a), and $(|\{j\in\Phi : j>k\}|+|\Psi|)y_k$ from (3–32d), leading to (3–33c).

However, for $k = 1,\dots,B-|\Phi|-|\Psi|+1$, $|\{j\in\Phi : j>k\}|+|\Psi| = |\Phi|+|\Psi| \le B-k+1$ (with equality when $k = B-|\Phi|-|\Psi|+1$) because $\phi_1 \ge B-|\Phi|-|\Psi|+3$ or $\Phi=\emptyset$. For each $k \ge B-|\Phi|-|\Psi|+1$ increasing k by one (a) decreases $B-k+1$ by exactly one and (b) decreases $|\{j\in\Phi : j>k\}|+|\Psi|$ by at most one, meaning that $B-k+1 \le |\{j\in\Phi : j>k\}|+|\Psi|$ for all such k; thus, (3–33c) becomes:

$$\sum_{i\in I(\Phi)} x_{\phi_i} + \sum_{i\in I(\Psi)} x_{\psi_i} \le \sum_{k=1}^{n'+1} \left(\min\left\{|\Phi|+|\Psi|, B-k+1\right\} + |\Phi_k|\right) y_k. \tag{3–34}$$

We now argue that (3–34) is equivalent to (3–13) if $(\Phi,\Psi)\in\Gamma$. From condition of (b) of Theorem 3.3, $\Phi_k \subseteq \{B-|\Phi|-|\Psi|+3,\dots,k-1\}$ for all $k \ge B-|\Phi|-|\Psi|+4$; therefore, $B-k+1+|\Phi_k| \le |\Phi|+|\Psi|-2$ for all such k. Moreover, for all $k \le B-|\Phi|-|\Psi|+3$, condition (b) of Theorem 3.3 necessitates that $\Phi_k=\emptyset$, proving that (3–34) is equivalent to (3–13).

3.3.3 Generalized Step Inequalities

In previous studies of the multiple-opponent problem (e.g., [32] and Chapter 2 of this dissertation), a variable θ^ω represents the value p_i^ω associated with the alternative i that maximizes $\bar{p}_i^\omega$ over all alternatives that have not been interdicted. In Chapter 2, "step inequalities" are developed and shown to be facets for

$$Y^\omega = \{(\hat{\theta}^\omega, \hat{x}) : \exists (x,\theta,y) \text{ feasible to model (3–3) with } \hat{\theta}\ge\theta \text{ and } \hat{x}=x\}, \tag{3–35}$$

the polyhedron linking θ^ω directly with x. These inequalities are shown in Chapter 2 to be generalizations of the step inequalities of [32], which assume each opponent knows the true p-values. We now show how these inequalities may be obtained by appropriately projecting out y-variables. In fact, this analysis actually leads to a further-generalized step inequality. To this end, we first summarize the step inequalities of Chapter 2.

For the purposes of this analysis, we again drop the superscripts from all data and variables and assume that for the scenario of interest, alternatives are ordered as in (3–6). We use the following notation, consistent with Chapter 2:

- $H = \{h_1, \dots, h_m\} \subseteq \{1, \dots, n'+1\}$, where $h_1 < \cdots < h_m$,
- $I(H) = \{1, \dots, m-1\}$,
- $\bar{I}(H) = \{1, \dots, m\}$,
- $\pi_i = p^{\omega}_{h_i}$, $i \in \bar{I}(H)$,
- $k(i)$ = the element of $\bar{I}(H)$ satisfying $h_{k(i)} = i$, and
- indices $\{\ell_1, \dots, \ell_m\}$ corresponding to $\bar{I}(H)$, where ℓ_i is the smallest index in $\bar{I}(H)$ such that: $\ell_i \geq i$ and $\pi_i \geq \pi_j \ \forall j = i+1, \dots, m$.

For $j = 1, \dots, n'+1$ define the set $G_j = \{i < j : p_i < p_j\}$. Let the variable θ represent the p_j-value corresponding to the chosen alternative in this scenario. From Chapter 2,

$$\theta \geq \pi_{\ell_1} - \sum_{i \in I(H): \ell_i = i} (\pi_i - \pi_{\ell_{i+1}}) x_{h_i} - \sum_{i \in I(H): \ell_i > i} (\pi_{\ell_i} - \pi_i)(1 - x_{h_i}). \tag{3–36}$$

is valid as long as $i \in H$ implies $G_i \subseteq H$ and $p_j > \pi_m$ for all $j = h_m + 1, \dots, n'+1$, and facet-defining when $1 \in H$. We now show how to obtain a generalized version of (3–36) by appropriately projecting out y-variables from (3–23). For this purpose, we define the set $H^+ = \{j > h_m : p_j < \pi_m\}$ and assume only that $G_i \subseteq H$ for all $i \in H$. Let $\bar{H} = \{1, \dots, h_m\} \setminus H$.

Since $p_j \geq 0$ for all $j \in \{1, \dots, n'+1\} \setminus [H \cup H^+ \cup \bar{H}]$, θ can be bounded as

$$\theta \geq \sum_{j \in \bar{H}} p_j y_j + \sum_{j \in H} p_j y_j + \sum_{j \in H^+} p_j y_j, \tag{3–37}$$

where $n'+1$ refers to the alternative selected if each of alternatives $1, \dots, n'$ is interdicted. Using (3–8b), we may eliminate y_{h_m} from (3–37) to obtain

$$\theta \geq \pi_m + \sum_{j \in \bar{H}} p'_j y_j + \sum_{i \in I(H)} \pi'_i y_{h_i} + \sum_{j \in H^+} p'_j y_j, \tag{3–38}$$

where $\pi'_i = \pi_i - \pi_m$ for all $i \in I(H)$ and $p'_j = p_j - \pi_m$ for all $j = 1, \dots, n'$. For $j \in \bar{H}$, $j < h_m$, define $\bar{\ell}_j$ as the smallest value $j' \in \bar{I}(H)$ such that $h_{j'} > j$ and $\ell_{j'} = j'$. Since $i \in H$ implies $j \in H$ for all $j < i$ with $p_j < p_i$, it follows that

$$\theta \geq \pi_m + \sum_{j \in \bar{H}} \pi'_{\bar{\ell}_j} y_j + \sum_{i \in I(H)} \pi'_i y_{h_i} + \sum_{j \in H^+} p'_j y_j. \tag{3–39}$$

Constraints (3–8c) and (3–8d) together with (3–39) and $\pi_{\ell_i} > \pi_i$ for all $i \in I(H)$ such that $\ell_i > i$ imply that

$$\theta \geq \pi_m + \sum_{j \in \bar{H}} \pi'_{\bar{\ell}_j} y_j + \sum_{i \in I(H)} \pi'_{\ell_i} y_{h_i} - \sum_{i \in I(H): \ell_i > i} (\pi_{\ell_i} - \pi_i)(1 - x_{h_i}) + \sum_{j \in H^+} p'_j y_j. \tag{3–40}$$

From the definition of $\bar{\ell}_j$, the values of $\pi_{\bar{\ell}_j}$ and $\pi_{k(j)}$ corresponding to $j \in \bar{H}$ and $j \in H$ form a non-increasing sequence and the first two summations can be combined as

$$\theta \geq \pi_m + \sum_{i \in I(H): \ell_i = i} \sum_{j=1}^{h_i} (\pi_i - \pi_{\ell_{i+1}}) y_{h_i} - \sum_{i \in I(H): \ell_i > i} (\pi_{\ell_i} - \pi_i)(1 - x_{h_i}) + \sum_{j \in H^+} p'_j y_j. \tag{3–41}$$

Substituting (3–11) for all y_{h_i} such that $i \in I(H)$ and $\ell_i = i$, and observing that $\sum_{j=1}^{h_i} (\pi_i - \pi_{\ell_{i+1}}) = \pi'_{\ell_1}$ produces

$$\theta \geq \pi_{\ell_1} - \sum_{i \in I(H): \ell_i = i} (\pi_i - \pi_{\ell_{i+1}}) x_{h_i} - \sum_{i \in I(H): \ell_i > i} (\pi_{\ell_i} - \pi_i)(1 - x_{h_i}) + \sum_{j \in H^+} p'_j y_j. \tag{3–42}$$

If $H^+ = \emptyset$, (3–42) is identically (3–36). Since $G_i \subseteq H$ for all $i \in H$, the results from Chapter 2 imply that (3–42) is valid; further, if $1 \in H$, the inequality is facet-defining for Y^ω. This generalizes to the case where $H^+ \neq \emptyset$ as well: The generalized step inequality (3–42) is facet-defining if and only if $G_i \subseteq H$ for all $i \in H$ and $1 \in H$. Proof of this result follows closely to the proof of the corresponding result for (3–36) in Chapter 2, and is therefore omitted.

Other classes of facets for Y^ω can be obtained via alternative projections of the y-variables. Consider, as an example, the instance with $n = 9$, $B = n' = 6$, and

$[p_1, \ldots, p_7] = [1,\ 0.75,\ 0.6,\ 0.42,\ 0.26,\ 0.19,\ 0.06]$, so that

$$\theta = y_1 + 0.75y_2 + 0.6y_3 + 0.42y_4 + 0.26y_5 + 0.19y_6 + 0.06y_7. \tag{3–43}$$

The terms in (3–43) can be grouped in a number of ways that lead to different projections. Consider for example, rearranging (3–43) as

$$\begin{aligned}
\theta =& 0.06(y_1 + y_2 + y_3 + y_4 + y_5 + y_6 + y_7) && \text{(3–44a)}\\
&+ 0.1(3y_1 + 3y_2 + 3y_3 + 3y_4 + 2y_5 + y_6) && \text{(3–44b)}\\
&+ 0.03y_6 && \text{(3–44c)}\\
&+ 0.06(y_1 + y_2 + y_3 + y_4) + 0.18(y_1 + y_2 + y_3) + 0.15(y_1 + y_2) + 0.25y_1. && \text{(3–44d)}
\end{aligned}$$

Replacing the y-terms in (3–44a), (3–44b), (3–44c), and (3–44d) respectively with (3–8b), (3–13) corresponding to $\Phi = \emptyset$ and $\Psi = \{7, 8, 9\}$, (3–8f), and (3–11) results in the following inequality,

$$\theta \geq 1 - 0.25x_1 - 0.15x_2 - 0.18x_3 - 0.1(1 - x_7) - 0.1(1 - x_8) - 0.1(1 - x_9), \tag{3–45}$$

which is facet-defining for Y^{ω}, yet not of the form (3–42). (The proof of this result is straightforward and therefore omitted.)

Alternatively, (3–43) may be factored as

$$\begin{aligned}
\theta =& 0.06(4y_1 + 4y_2 + 4y_3 + 3y_4 + 2y_5 + 2y_6 + y_7) && \text{(3–46a)}\\
&+ 0.07(2y_1 + 2y_2 + 2y_3 + 2y_4 + 2y_5 + y_6) && \text{(3–46b)}\\
&+ 0.1(y_1 + y_2 + y_3 + y_4) + 0.12(y_1 + y_2 + y_3) + 0.15(y_1 + y_2) + 0.25y_1. && \text{(3–46c)}
\end{aligned}$$

Projecting out the y-variables using (3–13) with $\Phi = \{5\}$ and $\Psi = \{7, 8, 9\}$, (3–13) with $\Phi = \emptyset$ and $\Psi = \{7, 8\}$, and (3–11) respectively on (3–46a), (3–46b), and (3–46c)

produces

$$\theta \geq 1-0.25x_1-0.15x_2-0.12x_3-0.1x_4-0.06(1-x_5)-0.13(1-x_7)-0.13(1-x_3)-0.06(1-x_9), \tag{3–47}$$

which is another facet-defining inequality for Y^ω.

3.4 Extensions to the Multiple-Opponent Problem

The previous section focuses on developing a full polyhedral representation of the one-opponent problem. Noting that the multiple-opponent problem is strongly NP-hard, there must be additional facet-defining inequalities that involve variables representing actions in multiple scenarios. Moreover, noting the polynomial equivalence of separation and optimization (see, e.g., [53]), some of these facets must be difficult to obtain. That is, one would expect some classes of facets to correspond to solutions for an NP-complete problem. We now describe an instance of the multiple-opponent problem (which includes such a class of facets) having alternatives $\{1, \dots, n\}$ and scenarios $\Omega = \{1, \dots, M\}$, where $n \gg M$. To this end, we first partition the alternatives subset $\{M+1, \dots, n-4\}$ into $H_2 = \{M+1, \dots, h'\}$, $H_3 = \{h'+1, \dots, h''\}$, and $H_4 = \{h''+1, \dots, n-4\}$. Define $B = 7$, and for $\omega = 1, \dots, M$ and $j = 1, \dots, 8$, let k_j^ω denote the j-th preferred alternative in scenario ω (as in Section 3.1). For each scenario, suppose $n' = B$. For this instance, we are interested in the polytope conv$[P^\Omega]$, defined in (3–4). We assume the following for each $\omega = 1, \dots, M$: $k_1^\omega = \omega$; $k_j^\omega \in H_j$ for $j = 2, 3, 4$; and $k_j^\omega = j + n - 8$ for $j = 5, 6, 7, 8$. We now establish that P^Ω has dimension $n + 6M$.

Lemma 3.3. The set P^Ω has dimension $n + 6M$.

Proof. Each $(x, y) \in P^\Omega$ satisfies

$$\sum_{j=1}^{8} y_j^\omega = 1, \text{ and} \tag{3–48}$$

$$y_1^\omega = 1 - x_{k_1^\omega}, \tag{3–49}$$

for each $\omega = 1, \dots, M$, analogous to (3–8b) and (3–8c). Equalities (3–49) are all linearly independent of each other because each y_1^ω appears exactly once. Equalities (3–48) must be linearly independent of each other, and of (3–49) because y_2^ω appears only once for each $\omega = 1, \dots, M$. Thus, the dimension of P^Ω is at most the number of variables less the number of linearly independent equalities, $n + 8M - 2M = n + 6M$.

We show that the dimension of P^Ω is at least $n+6M$ by providing $n+6M+1$ affinely independent points $c^0, \dots, c^{n+6M}$ in P^Ω of the form $(x_1, \dots, x_n, y_1^1, \dots, y_8^1, y_1^2, \dots, y_8^M)$. Define c^0 such that components y_1^ω, $\forall \omega = 1, \dots, M$ are equal to one and all others are equal to zero. For $i = 1, \dots, M$, define c^i exactly as c^0, but changing the x_i component from zero to one and swapping the values of y_1^i and y_2^i. For $i = M+1, \dots, n$, define c^i exactly as c^0 except for a one in the x_i-component. For $i = 1, \dots, M$ and $j = 1, \dots, 6$, define $c^{n+6(i-1)+j}$ such that: components x_i and $x_{k_2^i}, \dots, x_{k_{j+1}^i}$ are equal to one; components y_{j+2}^i and $y_1^{i'}$, $\forall i' \neq i$ are equal to one; and all other components are equal to zero.

Let C be the $(n+6M) \times (n+8M)$ matrix such that for each $i = 1, \dots, n+6M$, row i of C is given by $c^i - c^0$. Let $\tilde{C}$ be the square submatrix of C consisting of columns $x_1, \dots, x_n$ and $y_3^\omega, \dots, y_8^\omega$ for all $\omega = 1, \dots, M$. Because $\tilde{C}$ is a lower-triangular matrix with nonzero diagonals, the rows of $\tilde{C}$, and thus of C as well, must be linearly independent, proving that $c^0, \dots, c^{n+6M}$ are affinely independent. □

With k_1^ω and $k_5^\omega, \dots, k_8^\omega$ fixed, the entire problem instance can be described by the ordered triplets $(k_2^\omega, k_3^\omega, k_4^\omega) \in H_2 \times H_3 \times H_4$ for each $\omega = 1, \dots, M$. Noting that the multiple-opponent problem is NP-hard (see [51]) and having a complete and polynomially separable representation of the one-opponent convex hull, we would anticipate that some of the facets of conv$[P^\Omega]$ contain y-variables from several different scenarios; indeed, different combinations of the alternative $(k_2^\omega, k_3^\omega, k_4^\omega)$ lead directly to facet-defining inequalities.

A *matching* is a collection of scenarios $A \subseteq \Omega$ satisfying the following property: For any $i,\ i' \in A,\ i \neq i'$, it follows that $k_2^i \neq k_2^{i'}$, $k_3^i \neq k_3^{i'}$, and $k_4^i \neq k_4^{i'}$. For any matching $A \subseteq \Omega$, we may state the following inequality:

$$x_{n-3} + \sum_{i \in A} y_5^i \leq 1. \tag{3–50}$$

Validity of (3–50) is intuitive: If $x_{n-3} = 1$, then one-scenario constraints analogous to (3–8d) prevent the assignment of any y_5^i to one. Furthermore, it is not possible to have $y_5^i = y_5^{i'} = 1$ for $i, i' \in A,\ i \neq i'$ since each of $k_1^i, \ldots, k_4^i$ must be distinct from each of $k_1^{i'}, \ldots, k_4^{i'}$ and $B = 7$. A matching A is *maximal* if there does not exist $A',\ A \subset A' \subseteq \Omega$ such that A' is a matching. We now show that with each maximal matching of Ω is associated with a unique facet of conv$[P^\Omega]$.

Theorem 3.6. Suppose A is a maximal matching of Ω. Then (3–50) is facet-defining for conv$[P^\Omega]$.

Proof. As argued above, (3–50) is valid for conv$[P^\Omega]$, and thus, we need only establish that the dimension of the face of conv$[P^\Omega]$ defined by this inequality is $n + 6M - 1$ (following from Lemma 3.3). We construct $n + 6M$ affinely independent points $c^1, \ldots, c^{n+6M} \in P^\Omega$ of the form $(x_1, \ldots, x_n, y_1^1, \ldots, y_8^1, y_1^2, \ldots, y_8^M)$ that are binding on (3–50).

For $i = 1, \ldots, M$, define c^i such that components x_i, x_{n-3}, and y_2^i equal one as well as $y_1^{i'}$ for all $i' \neq i$ (and all other components equal zero). For $i = M+1, \ldots, n$ define c^i such that components x_i, x_{n-3}, and $y_1^i,\ \forall i = 1, \ldots, M$, are equal to one and all others are equal to zero. (Note that in c^{n-3}, only one x-variable, x_{n-3}, is set to one.)

For $i = 1, \ldots, M$ such that $i \in A$, define $c^{n+6(i-1)+j}$ such that:

- for $j = 1, 2$, components x_i, $x_{k_2^i}, \ldots, x_{k_{j+1}^i}$, x_{n-3}, and y_{j+2}^i equal one (and all others equal zero);
- for $j = 3$, components x_i, $x_{k_2^i}, \ldots, x_{k_4^i}$, and y_5^i equal one (and all others equal zero); and,

- for $j = 4, 5, 6$, components x_i, $x_{k_2^i}, \ldots, x_{k_{j+1}^i}$, and y_{j+2}^i equal one (and all others equal to zero).

Among these six points, only the one corresponding to $j = 3$ has $x_{n-3} = 0$ and in this point $y_5^i = 1$, indicating that all are binding on (3–50).

For $i = 1, \ldots, M$ such that $i \notin A$, there must exist $i' \in A$ such that $k_2^i = k_2^{i'}$, $k_3^i = k_3^{i'}$, or $k_4^i = k_4^{i'}$. For this i, define $c^{n+6(i-1)+j}$ exactly as if i were an element of A, but for $j = 3$, $x_{i'}$, $x_{k_2^{i'}}, \ldots, x_{k_4^{i'}}$, and $y_5^{i'}$ should also equal one. (Note: this is possible within the $B = 7$ restriction because scenarios i and i' share in common either their second, third, or fourth preference.)

Let C denote the $(n + 6M) \times (n + 8M)$ matrix such that row i is given by c^i for all $i = 1, \ldots, n + 6M$. We prove the only solution to $\lambda^T C = 0$ is $\lambda_i = 0, \ \forall i = 1, \ldots, n + 6M$. Columns y_{j+2}^i for $i \in \Omega \backslash A$ and $j = 1, \ldots, 6$ are identity columns, implying that $\lambda_{n+6(i-1)+j} = 0$. Columns y_{j+2}^i for $i \in A$ and $j = 1, \ldots, 6$ now contain only one λ-variable not already shown to equal zero, implying that $\lambda_{n+6(i-1)+j} = 0$. Having already shown that $\lambda_i = 0$ for all $i \geq n + 1$, columns x_i for $i = 1, \ldots, n - 4$ and $i = n - 2, n - 1, n$ now imply that $\lambda_i = 0$. From column x_{n-3}, we now find that $\lambda_{n-3} = 0$ as well; thus, $\lambda = 0$ is the only solution and $c^1, \ldots, c^{N+6M}$ must be linearly independent (and therefore affinely independent). That is, the face defined by (3–50) has dimension $n + 6M - 1$, as required. □

Given a matching A, it is easy to find a maximal matching A' by augmenting the scenarios in A using a simple greedy algorithm. The decision problem of finding a three-dimensional matching (3DM) of cardinality greater than or equal to an arbitrary constant k is NP-complete (by a trivial reduction from 3DM as defined in [60]). Using the simple conversion algorithm described above, 3DM is equivalent to the problem of finding a maximal matching of cardinality at least k, and therefore, the facets defined by (3–50) correspond to solutions of NP-complete problems.

3.5 Computational Results

We now present a brief computational study, which is designed to illustrate the effectiveness of our proposed inequalities in improving the lower bounds yielded by the linear programming relaxations to multiple-opponent C-BiPSNIP instances. To randomly generate C-BiPSNIP instances, we set $|\Omega| = 25$ and select $n \in \{50, 100, 150\}$ and $B \in \{0.5n, 0.8n\}$. For each combination of n and B, we generate five instances of (3–23) by randomizing the p- and $\bar{p}$-values. For $\omega = 1, \ldots, |\Omega|$ and $j = 1, \ldots, n$, we draw independent realizations of p_j^ω from a continuous uniform distribution on $(0.25, 0.75)$ with probability 0.6 and set $p_j^\omega = 0$ otherwise. For all ω and j such that $p_j^\omega > 0$, we independently generate $\bar{p}_j^\omega$ uniformly from the set $(p_j^\omega - 0.25, p_j^\omega + 0.25)$.

In our implementation, constraints (3–8a)–(3–8d), (3–8f)–(3–8g), and (3–11) are added a priori for each scenario and constraints (3–13) are added as needed. For each random instance, we minimize $\sum_{\omega\in\Omega} \sum_{j=1}^{n'+1} p_j^\omega y_j^\omega$; i.e., we assume equal weights f^ω, for all $\omega \in \Omega$, in model (3–3). We solve each instance using, in turn, three implementations: **No Cuts**, which simply solves the formulation using the default CPLEX settings (and does not make use of the valid inequalities (3–13)); **All Cuts**, in which all violated inequalities (3–13) are added at the root node in a cut-and-branch fashion; and **Significant Cuts**, in which cuts are only added to the root node if the violation exceeds 1, i.e., if the left-hand side of (3–30) less the right-hand side is at least 1.

Table 3-2 shows the objective value of the best integer solution (column **IP**) obtained by any of the three algorithms, and the corresponding optimality gap (column **G(%)**), computed as the absolute optimality gap divided by the objective lower bound. Also shown are the objective values for the root node continuous relaxation after adding cuts as prescribed in each algorithm (column **LP**). For each cut-adding algorithm, we derive gap reduction (column **R(%)**) to measure the effect of (3–13) on tightening the multiple-opponent formulation. This column is computed as $100(\text{LP}^c - \text{LP})/(\text{IP} - \text{LP})$, where LP^c is the root node continuous relaxation value with cuts and LP is the

Table 3-2. Comparison of continuous relaxations under different cut-adding strategies.

				No Cuts	All Cuts			Significant Cuts		
n	B	IP	G(%)	LP	LP	R(%)	Cuts	LP	R(%)	Cuts
50	25	30.35	0.00	23.42	23.42	0.00	0	23.42	0.00	0
50	40	26.88	0.00	6.88	11.80	24.60	228	11.11	21.17	130
100	50	31.20	1.59	24.34	24.34	0.02	1	24.34	0.00	0
100	80	28.40	17.32	7.11	12.13	23.59	289	11.80	22.01	184
150	75	31.73	6.80	24.42	24.42	0.02	1	24.42	0.00	0
150	120	29.19	28.86	7.17	12.08	22.30	317	11.83	21.15	210

continuous relaxation value without cuts. For each cut-adding algorithm, we also indicate the number of cuts added at the root-node.

Table 3-2 shows that varying n seems to have little effect on the gap reduction, but it also suggests that inequalities (3–13) are more important to the multiple-scenario formulation when B is large relative to n. When $B = 0.5n$, cuts provide no improvement to the LP relaxation in most instances and modest improvement in others. However, when $B = 0.8n$, adding (3–13) consistently reduces the gap by 20 to 25%. Interestingly, adding only significantly violated cuts provides improvement to the root node lower-bound nearly equivalent to that obtained by adding all violated cuts.

CHAPTER 4
GEOGRAPHICAL INTERDICTION OF A MAXIMUM FLOW NETWORK

We consider a network that resides in Euclidean space and a defender whose goal is to maximize flow from a source node to a sink node in the network. The problem we examine takes the perspective of an interdictor, who seeks to minimize the defender's maximum flow by locating attacks in the region on which the network is located. Attack locations are not restricted to node or arc locations, and serve to diminish arc capacities in accordance with the distance from the arc to the attack. The problem we consider is known to be NP-hard as it generalizes the maximum flow interdiction problem studied by Wood [8]. We develop two approaches to solving this problem based on solving a sequence of lower-bounding integer programs, and compare the efficacy of our approaches on a set of randomly generated test instances.

4.1 Background

We consider the interdiction of a capacitated network that exists in Euclidean space. Nodes in this network exist at a point in space and (directed) arcs connect node pairs in a straight line. An opponent wishes to maximize flow from a source node to a sink node across the network, while an interdictor seeks to minimize the opponent's maximum flow by choosing multiple locations to attack. In this problem, attacks are made at points in space. Damage is inflicted on each arc by reducing its capacity as a function of the distance from the midpoint of the arc to an attack. We refer to this problem as the *Euclidean maximum flow network interdiction problem* (E-MFNIP). In this chapter, we provide mathematical programming-based approaches for solving E-MFNIP.

The motivation for studying this type of problem is due to the prevalence of networks in complex systems and the vulnerability of those systems to attack. For instance, networks that represent real-world transportation, logistics, and power grid systems have well-defined geographical characteristics. These systems are subject to disruptions (e.g., an earthquake) that may damage multiple components in the same geographical

region. Most network interdiction research neglects any geographical characteristics, focusing instead on identifying network components that are most critical to sustaining functionality. In contrast, our methodology explicitly accounts for the simultaneous failure of network components correlated by the physical location of an attack. As a result, this model provides valuable insights into identifying physical locations of vulnerability.

Over the last half-decade, several studies have examined network resilience under geographically correlated failures. Neumayer et al. [61] use an integer programming model to identify a single "vertical line cut" in a bipartite network that maximizes the total capacity of all intersected arcs. This work is extended in [62], which considers a general flow network, two types of cuts (circular and line segment), and several measures of network performance (including the worst-case maximum flow between a pair of nodes). In a related work, Neumayer and Modiano [63] examine the resilience of a network to a single probabilistic geographical failure. Agarwal et al. [64, 65] consider the problem of locating multiple circle-shaped attacks around telecommunications network in which the probability of a component failure depends on its distance from the epicenter of an attack. The major contribution of these papers is a greedy algorithm that approximates the network's resilience (defined as its ability to transmit a fixed pattern of traffic). Bernstein et al. [66] model geographically correlated failures and ensuing cascading failures in a power grid, and use a method based on [65] to identify attack locations to which the grid is vulnerable. Survivable design of networks subject to geographical failures [67] has also received some attention in the context of undersea cable networks.

Our problem is most similar to that of Neumayer et al. [62], who also characterize worst-case disruptions using the maximum flow metric. To the best of our knowledge, however, our research is the first attempt to model the worst-case effects of multiple geographical disruptions in a maximum flow network. Moreover, our problem is more general than the maximum flow model of [62] in that arc capacities may be defined

generally as a (possibly continuous) function of distance between arc and attack; in [62], an attack either has no effect on an arc or removes it altogether.

The remainder of this chapter is organized as follows. In Section 4.2, we develop a general model for maximum flow interdiction that yields a MIP formulation for E-MFNIP when capacity functions are piecewise-convex. In Section 4.3, we extend this model to develop a MIP model that provides a lower bound for E-MFNIP under more general capacity functions. We then describe how the lower bound can be dynamically improved within an implicit enumeration algorithm. Anticipating the difficulty of solving this model, Section 4.4 describes a problem that restricts the interdictor to select its attacks from a finite list of candidate locations. This model can be employed to yield both a lower and upper bound for the optimal E-MFNIP objective value, giving rise to a discretize-and-refine method for solving E-MFNIP. We summarize computational results from solving E-MFNIP on randomly generated test networks in Section 4.5.

4.2 Maximum Flow Interdiction Models

In this section, we develop a MIP model that may be used in solving E-MFNIP. We begin by recapping a generic single-stage model for maximum flow interdiction in Section 4.2.1. In Section 4.2.2, we specialize this model to obtain a model for E-MFNIP under piecewise-convex capacity functions. Section 4.3 describes how this model may be used in solving E-MFNIP instances with more general capacity functions.

4.2.1 Generic Single-Stage Formulation

Let $G(N, A)$ be a maximum flow network having node set $N = \{1, \dots, n\}$ and directed arc set $A \subseteq N \times N$, where node 1 is the source node and node n is the sink node. Let c_{ij} denote the capacity of arc $(i, j) \in A$, and define variables x_{ij} as the amount of flow on arc $(i, j) \in A$. Also, let arc $(n, 1)$ be an uncapacitated "return arc" from node n to node 1, and let $\bar{A} = A \cup \{(n, 1)\}$. The maximum flow (denoted ν) from node 1 to node

n can be obtained by solving the model

$$\nu = \max\ x_{n1}, \tag{4–1a}$$

$$\text{s.t.} \sum_{j \in FS(i)} x_{ij} - \sum_{j \in RS(i)} x_{ji} = 0, \quad \forall i \in N, \qquad : u_i \tag{4–1b}$$

$$0 \leq x_{ij} \leq c_{ij}, \quad \forall (i,j) \in A, \qquad : v_{ij} \tag{4–1c}$$

where $FS(i) = \{j \in N : (i,j) \in \bar{A}\}$ and $RS(i) = \{j \in N : (j,i) \in \bar{A}\}$. Because the objective maximizes x_{n1}, it is not necessary to require $x_{n1} \geq 0$. Note that aggregating Constraints (4–1b) corresponding to nodes $i = 2, \ldots, n$ implies Constraint (4–1b) corresponding to node 1, and so we omit this constraint in our subsequent analysis.

Model (4–1) is a linear program (LP) and thus, ν could equally be obtained by solving the LP dual of (4–1). Introducing dual variables u associated with Constraints (4–1b) (with $u_1 \equiv 0$) and v associated with upper-bounding Constraints (4–1c), the constraints are:

$$u_i - u_j + v_{ij} \geq 0, \quad \forall (i,j) \in A, \tag{4–2a}$$

$$u_n = 1, \tag{4–2b}$$

$$v_{ij} \geq 0, \quad \forall (i,j) \in A. \tag{4–2c}$$

Letting $\mathcal{U}$ be the feasibility set that encompasses the constraints of (4–2), the dual maximum flow formulation is

$$\nu = \min \sum_{(i,j) \in A} c_{ij} v_{ij}, \ \text{s.t.}\ (u, v) \in \mathcal{U}. \tag{4–3}$$

In our problem, the interdictor chooses an attack $\mathbf{z}$ from some feasible set Z, and arc capacities are computed according to a function $\delta_{ij} : Z \to \mathbb{R}$ for all $(i,j) \in A$. After the leader selects an interdiction $\hat{\mathbf{z}} \in Z$, the follower maximizes flow on the resulting network, solving (4–1) with c_{ij} replaced by $\delta_{ij}(\hat{\mathbf{z}})$. The interdictor's problem is then to choose an interdiction that minimizes the follower's maximum flow, which can be stated

as:

$$\nu^\star = \min \sum_{(i,j)\in A} \delta_{ij}(\mathbf{z}) v_{ij}, \text{ s.t. } (u,v) \in \mathcal{U}, \ \mathbf{z} \in Z. \tag{4–4}$$

When δ_{ij} is defined nontrivially (i.e., not constant) for some $(i,j) \in A$, this model is nonlinear in the objective. However, as shown in [8], for any $\hat{\mathbf{z}} \in Z$ there exists a solution to (4–3) in which $v_{ij} \in \{0,1\}$ for all $(i,j) \in A$. Imposing these restrictions, we can linearize the objective function by defining variables w_{ij} for each $(i,j) \in A$ to represent the product $\delta_{ij}(\mathbf{z}) v_{ij}$, and adding the inequalities

$$w_{ij} \geq \delta_{ij}(\mathbf{z}) - M_{ij}(1 - v_{ij}), \quad \forall (i,j) \in A, \text{ and} \tag{4–5a}$$

$$w_{ij} \geq 0, \quad \forall (i,j) \in A, \tag{4–5b}$$

where M_{ij} is an upper bound on $\delta_{ij}(\mathbf{z})$ over all $\mathbf{z} \in Z$. The resulting formulation is a MIP that is linearly constrained when δ_{ij} is a linear function:

$$\nu^\star = \min \sum_{(i,j)\in A} w_{ij}, \tag{4–6}$$

$$\text{s.t. Constraints (4–5a) and (4–5b)}, \ (u,v) \in \mathcal{U}, \ \mathbf{z} \in Z, \ v_{ij} \in \{0,1\}, \ \forall (i,j) \in A.$$

Remark 4.1. Model (4–6) generalizes several versions of maximum flow interdiction. In the basic *maximum flow network interdiction problem* (MFNIP), the leader has a budget of K and can remove any arc $(i,j) \in A$ at a cost of b_{ij}. Removing an arc (i,j) effectively reduces its capacity from c_{ij} (a constant) to zero. Setting $Z = \{\mathbf{z} \in \{0,1\}^{|A|} : \sum_{(i,j)\in A} b_{ij} z_{ij} \leq K\}$, $\delta_{ij}(\mathbf{z}) = c_{ij}(1 - z_{ij})$, and $M_{ij} = c_{ij}$ in (4–6) results in a model equivalent to the MFNIP model given in [8]. (Hence, the optimization problem given by (4–6) is NP-hard.) When $b_{ij} = 1, \ \forall (i,j) \in A$, in the above definition of Z, the *cardinality* constrained maximum flow network interdiction problem (C-MFNIP), or K most-vital arcs problem, results. □

4.2.2 Specification to E-MFNIP

Typically $\mathbf{z} \in Z$ consists of $|A|$ decisions, each representing the amount of interdiction imparted on an arc. The novelty of our research lies in our definition of the set Z: We assume that G has physical structure, residing in q-dimensional Euclidean space, and that at most K interdictions are made at points on and around G. In what follows, we use the notation $[s]$, defined for any positive integer s, to refer to the set $\{1, \dots, s\}$. Thus, an element $\mathbf{z} \in Z$ is given as $\mathbf{z} = [\mathbf{z}^1 | \dots | \mathbf{z}^K]$, where $\mathbf{z}^k \in \mathbb{R}^q,\ \forall k \in [K]$, is the location of the k-th interdiction. For simplicity, we assume that the constraint set Z is given as

$$Z = \{\mathbf{z} \in \mathbb{R}^{q \times K} : L \leq \mathbf{z}^k \leq U,\ k \in [K]\}, \tag{4–7}$$

where L and U are q-vectors. We assume that each node $i \in N$ is located at a point $\mathbf{p}(i) \in \mathbb{R}^q$, and hence each arc center $(i,j) \in A$ is located at the point $\mathbf{p}(i,j) = 0.5(\mathbf{p}(i) + \mathbf{p}(j))$. We assume that the damage inflicted by an attack on an arc is a function of the distance between the arc and the attack location. For simplicity, we compute distance using the Manhattan norm, i.e., $\|\mathbf{p}\|_1 \equiv \sum_{k=1}^{q} |\mathbf{p}_k|$. For $\hat{\mathbf{z}} \in \mathbb{R}^q$, define $d_{ij}(\hat{\mathbf{z}}) = \|\hat{\mathbf{z}} - \mathbf{p}(i,j)\|_1$ as the distance between $\hat{\mathbf{z}}$ and arc (i,j). In our model, the capacity of arc (i,j) is determined by the distance of the closest attack to $\mathbf{p}(i,j)$. That is, capacity functions are of the form

$$\delta_{ij}(\mathbf{z}) = f_{ij}\left[\min_{k \in [K]} d_{ij}(\mathbf{z}^k)\right], \tag{4–8}$$

where $f_{ij} : \mathbb{R}_+ \to \mathbb{R}_+$ is a function that maps distance to capacity. When $K = 1$, this assumption is intuitive. When $K > 1$, this assumption is conservative (from the interdictor's perspective) because only one of the K interdictions influences the capacity of each arc.

E-MFNIP is the problem that results when Z and δ are defined as in (4–7) and (4–8), respectively. We define $\nu^\star{}_E$ as the optimal objective function value to E-MFNIP.

Remark 4.2. We focus primarily on scenarios in which the original problem may have a set S^1 of (multiple) source nodes and a set S^2 of sink nodes. This problem can be

converted into a problem having only one source and one sink as follows: (i) Create source node 1 and sink node n; (ii) Create arcs $(1, j),\ \forall j \in S^1$, and arcs $(i, n),\ \forall i \in S^2$; and (iii) Define $f_{1j}(d) = M$ for all $j \in S^1$ and $f_{in}(d) = M$ for all $i \in S^2$, where M is a very large constant. □

We now specialize (4–6) to model E-MFNIP as a MIP having a convex continuous relaxation under a certain class of f-functions. In this section, we assume that f_{ij} is given as

$$f_{ij}(d) = \min_{t \in [T]} g_{ij}^t(d), \tag{4–9}$$

where $g_{ij}^t : \mathbb{R}^+ \to \mathbb{R}^+$ is convex and nondecreasing for each $t \in [T]$. Define $h_{ij}^t = g_{ij}^t \circ d_{ij}$ and note that h_{ij}^t is a convex function because g_{ij}^t and d_{ij} are convex and g_{ij}^t is nondecreasing. Moreover, observe that δ_{ij} can be represented as

$$\delta_{ij}(\mathbf{z}) = \min_{t \in [T]} g_{ij}^t \left[\min_{k \in [K]} d_{ij}(\mathbf{z}^k) \right] \tag{4–10a}$$

$$= \min_{t \in [T], k \in [K]} g_{ij}^t \circ d_{ij}(\mathbf{z}^k) \tag{4–10b}$$

$$= \min_{t \in [T], k \in [K]} h_{ij}^t(\mathbf{z}^k), \tag{4–10c}$$

where (4–10b) follows because g_{ij}^t is nondecreasing.

Define λ_{ij}^{tk} to equal 1 if the minimum in (4–10b) is achieved by $t \in [T]$ and $k \in [K]$ and $\lambda_{ij}^{tk} = 0$ otherwise. Using the λ-variables, δ_{ij} can be expressed as

$$\delta_{ij}(\mathbf{z}) = \min \sum_{t \in [T], k \in [K]} \left[h_{ij}^t(\mathbf{z}^k) \right] \lambda_{ij}^{tk}, \tag{4–11a}$$

$$\text{s.t.} \sum_{t \in [T], k \in [K]} \lambda_{ij}^{tk} = 1 \text{ and } \lambda_{ij}^{tk} \in \{0, 1\} \ \ \forall t \in [T],\ k \in [K]. \tag{4–11b}$$

This expression could be substituted for δ_{ij} in Model (4–6) to obtain a valid formulation. Furthermore, observe that decreasing $\delta_{ij}(\mathbf{z})$ relaxes (4–5a), which is the only place $\delta_{ij}(\mathbf{z})$ appears in Model (4–6). Thus, the minimization operator can be dropped from (4–11), because an optimal solution will always exist in which $\delta_{ij}(\mathbf{z})$ takes its smallest value

allowed by (4–11). The resulting MINLP minimizes $\sum_{(i,j)\in A} w_{ij}$ subject to $(u, v) \in \mathcal{U}$, $\mathbf{z} = [\mathbf{z}^1|\cdots|\mathbf{z}^r] \in Z$, $v \in \{0,1\}^{|A|}$, and (4–5b) as well as

$$w_{ij} \geq \sum_{t\in[T],k\in[K]} \left[h_{ij}^t(\mathbf{z}^k)\right] \lambda_{ij}^{tk} - M_{ij}(1 - v_{ij}), \quad \forall (i,j) \in A, \tag{4–12a}$$

$$\sum_{t\in[T],k\in[K]} \lambda_{ij}^{tk} = 1, \quad \forall (i,j) \in A, \text{ and} \tag{4–12b}$$

$$\lambda_{ij}^{tk} \in \{0,1\}, \quad \forall (i,j) \in A,\ t \in [T],\ k \in [K]. \tag{4–12c}$$

To linearize (4–12a), let ϕ_{ij}^{tk} represent the product $\left[h_{ij}^t(\mathbf{z}^k)\right] \lambda_{ij}^{tk}$ and introduce inequalities similar to (4–5) to obtain the following model, valid for any $\bar{M}_{ij}^t$ such that $\bar{M}_{ij}^t \geq h_{ij}^t(\mathbf{z}),\ \forall L \leq \mathbf{z} \leq U,\ (i,j) \in A,\ t \in [T]$.

$$\nu^\star{}_h = \min \sum_{(i,j)\in A} w_{ij} \tag{4–13a}$$

$$\text{s.t. } (u, v) \in \mathcal{U},$$

$$w_{ij} \geq \sum_{t\in[T],k\in[K]} \phi_{ij}^{tk} - M_{ij}(1 - v_{ij}), \quad \forall (i,j) \in A, \tag{4–13b}$$

$$w_{ij} \geq 0, \quad \forall (i,j) \in A, \tag{4–13c}$$

$$\sum_{t\in[T],k\in[K]} \lambda_{ij}^{tk} = 1, \quad \forall (i,j) \in A, \tag{4–13d}$$

$$\phi_{ij}^{tk} \geq h_{ij}^t(\mathbf{z}^k) - \bar{M}_{ij}^t(1 - \lambda_{ij}^{tk}), \quad \forall (i,j) \in A,\ t \in [T],\ k \in [K], \tag{4–13e}$$

$$\phi_{ij}^{tk} \geq 0, \quad \forall (i,j) \in A,\ t \in [T],\ k \in [K], \tag{4–13f}$$

$$v_{ij} \in \{0,1\}, \quad \forall (i,j) \in A, \tag{4–13g}$$

$$\lambda_{ij}^{tk} \in \{0,1\}, \quad \forall (i,j) \in A,\ t \in [T],\ k \in [K], \tag{4–13h}$$

$$\mathbf{z} \in Z.$$

As before, Model (4–13) is a MINLP, but this time its continuous relaxation is convex (because all h-functions are convex). These tractable continuous relaxations can be used to solve (4–13) via, e.g., branch and bound so long as all f-functions can be

expressed as (4–9). In the following section, we discuss techniques for approximating E-MFNIP when f cannot be expressed as (4–9), based on using Model (4–13) with appropriately defined g-functions.

4.3 Solving over a General Capacity Function

In many applications, the f-functions cannot be represented in the form of (4–9). In the remainder of this section, we outline a procedure for using Model (4–13) to lower-bound the optimal E-MFNIP objective function value by defining appropriate piecewise-linear g-functions.

We assume the true δ-functions take the form of (4–8), with f being nondecreasing and concave. These assumptions intuitively match what one would expect from the behavior of a realistic interdiction: Capacity increases as the interdiction moves farther away from an arc (nondecreasing), and a small change in interdiction location is more likely to have a pronounced impact on nearby arcs than on far away arcs (concave). If f is also piecewise-linear, Model (4–13) becomes an exact model using linear g-functions. If f is not piecewise-linear, g-functions can be selected to ensure that (4–13) provides a lower bound on the optimal E-MFNIP objective.

One might reasonably conjecture that, under these assumptions on f, optimal attack locations always coincide with the midpoint of an arc. This conjecture is false, as we illustrate through the following example. Consider the two-dimensional network illustrated in Figure 4-1 in which all odd-numbered nodes are sources and all even-numbered nodes are sinks. The only possible minimum source-sink cut is given by the arcs $(1,2)$, $(3,4)$, and $(5,6)$, where $\mathbf{p}(1,2) = (0,0)$, $\mathbf{p}(3,4) = (0.5,0.5)$, and $\mathbf{p}(5,6) = (1,0)$. If (for $K = 1$) $f_{12}(d) = f_{34}(d) = f_{56}(d) = f(d) \equiv d^{0.75}$ (a strictly concave, nondecreasing function), then attacking either $(0,0)$, $(0.5,0.5)$, or $(1,0)$ reduces the capacity of this cut to $2f(1) = 2$. An attack at $(0.5,0)$, which does not coincide with the midpoint of an arc, results in an even smaller minimum cut capacity of $3f(0.5) = 1.78$.

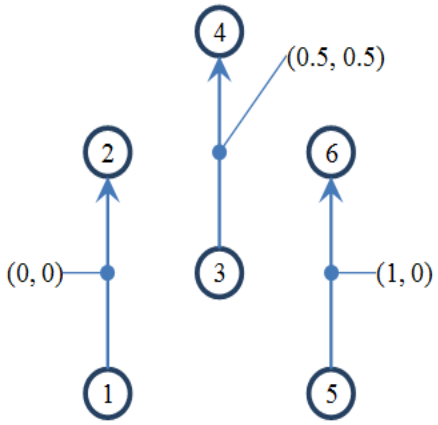

Figure 4-1. Example network with only one possible cut set.

We now describe a technique for approximating δ_{ij}-functions by defining piecewise-linear g_{ij}-functions. Suppose breakpoints $a_0, \ldots, a_T$ are given such that $0 = a_0 < a_1 < \cdots < a_T$, and $d_{ij}(\mathbf{z}) \leq a_T$ for all $L \leq \mathbf{z} \leq U$. For each $t \in [T]$, let g_{ij}^t be the linear function that approximates f_{ij} and is tight at a_{t-1} and a_t, i.e.,

$$g_{ij}^t(d) = \frac{f(a_t) - f(a_{t-1})}{a_t - a_{t-1}}(d - a_t) + f(a_{t-1}). \tag{4–14}$$

See Figure 4-2 for an example f-function and corresponding lower-bounding g-functions.

Theorem 4.1. Let f be concave and nondecreasing, and suppose g_{ij}^t is given as in (4–14). Then (4–13) defined with $h_{ij}^t = g_{ij}^t \circ d_{ij}$ provides a lower bound for E-MFNIP, i.e., $\nu^\star{}_h \leq \nu^\star{}_E$.

Proof. Since f is concave, $g_{ij}^t(d)$ provides a lower bound for $f(d)$ over the interval $d \in [a_{t-1}, a_t]$. Because $d_{ij}(\mathbf{z}) \leq a_T$ for all $L \leq \mathbf{z} \leq U$, there exists a $t \in [T]$ such that $a_{t-1} \leq d_{ij}(\mathbf{z}) \leq a_t$. Therefore, for every $L \leq \mathbf{z} \leq U$, there exists a $t \in [T]$ such that

$$g_{ij}^t \circ d_{ij}(\mathbf{z}) \leq f_{ij} \circ d_{ij}(\mathbf{z}). \tag{4–15}$$

Now, consider any solution $(\hat{\mathbf{z}}, \hat{u}, \hat{v}, \hat{w})$ that is optimal for (4–6), and note that $\nu^\star{}_E = \sum_{(i,j)\in A} \hat{w}_{ij}$. We prove that $\nu^\star{}_h \leq \nu^\star{}_E$ by showing that there exist $\hat{\phi}$ and $\hat{\lambda}$ such

that $(\hat{\mathbf{z}}, \hat{u}, \hat{v}, \hat{w}, \hat{\phi}, \hat{\lambda})$ is a feasible solution to (4–13) having objective value $\nu^\star{}_E$. For each $(i,j) \in A$, let t_{ij} and k_{ij} be defined such that $(t_{ij}, k_{ij}) \in \operatorname{argmin}_{t\in[T],k\in[K]} \left\{ g_{ij}^t \circ d_{ij}(\hat{\mathbf{z}}^k) \right\}$. For $(i,j) \in A$, $t \in [T]$, and $k \in [K]$, define

$$\hat{\lambda}_{ij}^{tk} = \begin{cases} 1 & \text{if } t = t_{ij} \text{ and } k = k_{ij} \\ 0 & \text{otherwise,} \end{cases}$$

and observe that (4–13d) and (4–13h) are satisfied by this choice of $\hat{\lambda}$. Constraints (4–13e) and (4–13f) reduce to $\phi_{ij}^{t_{ij}k_{ij}} \geq h_{ij}^{t_{ij}}(\hat{\mathbf{z}}^k) = g_{ij}^t \circ d_{ij}(\hat{\mathbf{z}}^k)$ and to $\phi_{ij}^{tk} \geq 0$ when $t \neq t_{ij}$ or $k \neq k_{ij}$. Accordingly, define $\hat{\phi}$ as

$$\hat{\phi}_{ij}^{tk} = \begin{cases} \min_{t\in[T],k\in[K]} g_{ij}^t \circ d_{ij}(\hat{\mathbf{z}}^k) & \text{if } t = t_{ij} \text{ and } k = k_{ij} \\ 0 & \text{otherwise,} \end{cases}$$

and note that Constraints (4–13e) and (4–13f) are satisfied.

Next, observe that (4–13b) holds because

$$\begin{aligned} \sum_{t\in[T],k\in[K]} \hat{\phi}_{ij}^{tk} &= \hat{\phi}_{ij}^{t_{ij}k_{ij}} && \text{(4–16a)} \\ &= \min_{t\in[T],k\in[K]} g_{ij}^t \circ d_{ij}(\hat{\mathbf{z}}^k) && \text{(4–16b)} \\ &\leq f_{ij} \circ d_{ij}(\hat{\mathbf{z}}^k) && \text{(4–16c)} \\ &\leq \hat{w}_{ij} + M_{ij}(1 - \hat{v}_{ij}), && \text{(4–16d)} \end{aligned}$$

where (4–16c) follows due to (4–15), and (4–16d) is implied by Constraint (4–5a) of Model (4–6). The remaining constraints of (4–13) are also constraints of (4–6) and hence, $(\hat{\mathbf{z}}, \hat{u}, \hat{v}, \hat{w}, \hat{\phi}, \hat{\lambda})$ is feasible to Model (4–13) with objective value $\sum_{(i,j)\in A} \hat{w}_{ij} = \nu^\star{}_E$. □

Remark 4.3. Observe that f need not be concave in order to apply the approximation MIP in this section. We illustrate the use of convex g-functions to lower-bound a (nonconcave) step function f in Figure 4-3. Such a lower-bounding scheme provides

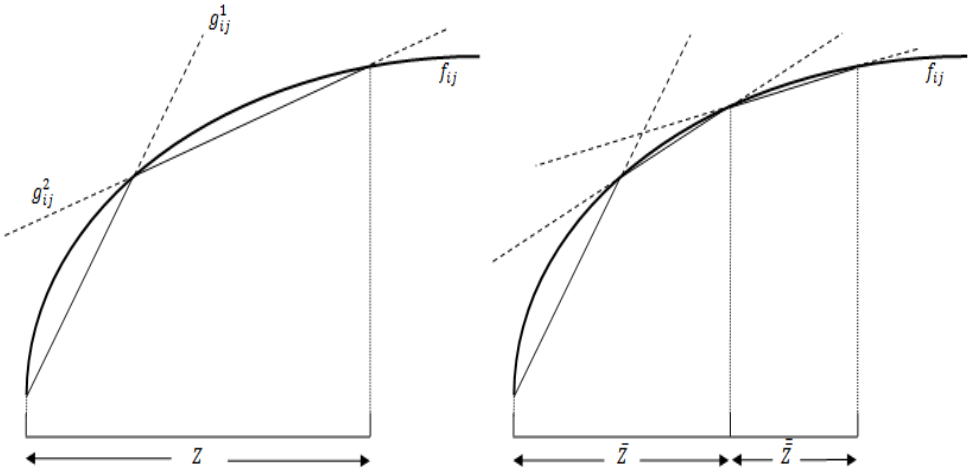

Figure 4-2. f_{ij} lower bounded by g_{ij}^1 and g_{ij}^2 over $\mathbf{z} \in Z$ (left) and refined approximation after partitioning Z (right).

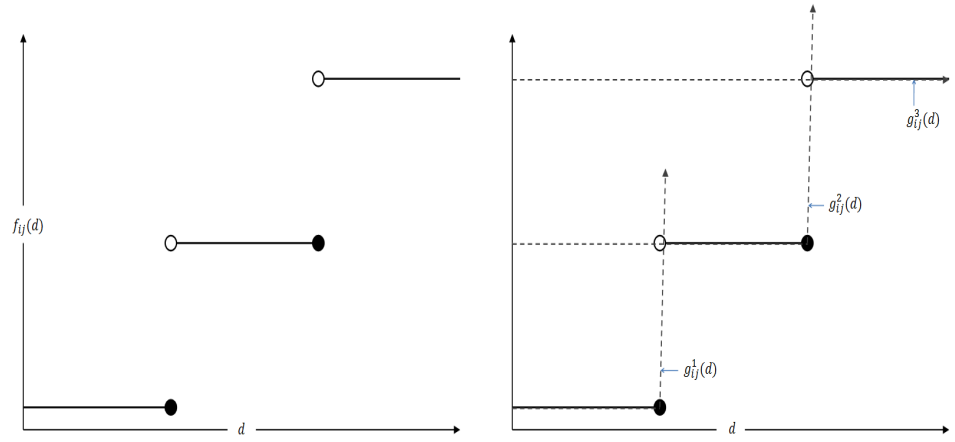

Figure 4-3. Example step function f_{ij} (left) and lower-bounding convex g_{ij}-functions (right).

a way to model capacity functions such as those used in [62] in which arcs are either completely removed (if an attack is within r distance units) or altogether unharmed. □

Remark 4.4. It is possible to refine the g-functions dynamically within the solution of (4–13) to obtain a better approximation for E-MFNIP. This algorithm is based on partitioning a search space Z into regions $\bar{Z}$ and $\bar{\bar{Z}}$, refining the g-functions to more tightly approximate f over these regions, and then branching to determine whether each attack belongs to $\bar{Z}$ or $\bar{\bar{Z}}$. Whether a particular attack is constrained to lie in $\bar{Z}$ or $\bar{\bar{Z}}$

determines how the g-functions can be tightened. As such, we replace the functions g_{ij}^{t} in Model (4–13) with g_{ij}^{tk}. Figure 4-2 demonstrates the partitioning of a set Z into $\bar{Z}$ and $\bar{\bar{Z}}$, resulting in tighter approximations on either side.

The generic scheme we present is based on the following three concepts. One, solving (4–13) over some region $\hat{Z} \subseteq Z$ yields a lower bound on the optimal objective value to (4–6) in which $\mathbf{z} \in \hat{Z}$. An upper bound can also be obtained based on any $\hat{\mathbf{z}} \in Z$ by evaluating $\nu(\hat{\mathbf{z}})$, the maximum flow after attack $\hat{\mathbf{z}}$. Two, note that Z can be divided into two sets $\bar{Z}$ and $\bar{\bar{Z}}$ by choosing some hyperplane H, letting $H_{\leq}$ and $H_{\geq}$ be the two halfspaces induced by H, and setting $\bar{Z} = Z \cap H_{\leq}$ and $\bar{\bar{Z}} = Z \cap H_{\geq}$. Three, a valid branching scheme given $Z = \bar{Z} \cup \bar{\bar{Z}}$ creates $K+1$ subproblems, where in the k-th subproblem, $\mathbf{z}^1, \ldots, \mathbf{z}^{k-1}$ are constrained to come from $\bar{Z}$ and $\mathbf{z}^k, \ldots, \mathbf{z}^K$ must come from $\bar{\bar{Z}}$. Observe that all interdiction solutions that are not feasible to any of these $K+1$ subproblem are symmetric (via reindexing attack locations $\mathbf{z}^1, \ldots, \mathbf{z}^K$) to a solution that is feasible to one of the subproblems.

Using these concepts, we may perform a modified branch-and-bound search of Z. We begin as if solving (4–13) over $\mathbf{z} \in Z$ and an initial set of g-functions via standard branch-and-bound. In this algorithm, a solution $\hat{\mathbf{z}} \in Z$ with objective value $\hat{\nu}$ may be retained as an incumbent solution only if (i) $\hat{\nu}$ is the smallest objective value among known integral solutions and (ii) $\nu(\hat{\mathbf{z}}) - \hat{\nu}$ is less than some tolerance. If $\hat{\mathbf{z}}$ is rejected as an incumbent solution due to reason (ii), the problem is divided into $K+1$ subproblems as described above. In this case, this particular solution to (4–13) is such that for some $t^\star \in [T]$, $k^\star \in [K]$, and $(i^\star, j^\star) \in A$, we have that $v_{i^\star j^\star} = 1$, $\lambda_{i^\star j^\star}^{t^\star k^\star} = 1$, and $g_{i^\star j^\star}^{t^\star k^\star} \circ d_{i^\star j^\star}(\hat{\mathbf{z}}^{k^\star}) < f_{i^\star j^\star} \circ d_{i^\star j^\star}(\hat{\mathbf{z}}^{k^\star})$. To fix this, when we partition Z into $\bar{Z} \cup \bar{\bar{Z}}$ (and specify how many attacks are to be on each side of the partition), we update the g-functions so there exists $t \in [T]$ and $k \in [K]$ such that $g_{ij}^{tk} \circ d_{ij}(\hat{\mathbf{z}}^{k^\star}) = f_{ij} \circ d_{ij}(\hat{\mathbf{z}}^{k^\star})$ for all $(i,j) \in A$. Note that if $\hat{\mathbf{z}}$ is encountered as a solution in one of the subproblems, the updated g-functions

will require that the objective value at $\hat{\mathbf{z}}$ in the modified Model (4–13) matches the true maximum flow after attack $\hat{\mathbf{z}}$.

Figures 4-4 and 4-5 illustrate the branch-and-refine procedure discussed in the previous paragraph. In this example, $q = K = 2$ and $Z = [0, 4] \times [0, 4]$. Arc (i, j) has capacity function $f_{ij}(d) = 10(1 - 0.5^d)$, where d is the distance from $(2, 1)$—the midpoint of arc (i, j)—to the nearest attack. Model (4–13) has been defined using $T = 1$, so that a single linear g-function lower-bounds f_{ij}, i.e., $g_{ij}^{11}(d) = g_{ij}^{12}(d) = g_{ij}(d) = 1.94d$. The solution $\hat{\mathbf{z}}^1 = (0.9, 2.6), \hat{\mathbf{z}}^2 = (2.2, 1.8)$ has been identified as a potential solution to (4–13). Note, however, that $g_{ij}(d) < f_{ij}(d)$ for $d = |(2 - 2.2)| + |1 - 1.8| = 1$. In this instance, we divide Z using the hyperplane $H = \{\mathbf{z} : \mathbf{z}_1 + \mathbf{z}_2 = 4\}$. We then branch and create three subproblems: (i) $\mathbf{z}^1, \mathbf{z}^2 \in \bar{Z} \equiv Z \cap \{\mathbf{z} : \mathbf{z}_1 + \mathbf{z}_2 \leq 4\}$, (ii) $\mathbf{z}^1 \in \bar{Z}$ and $\mathbf{z}^2 \in \bar{\bar{Z}} \equiv Z \cap \{\mathbf{z} : \mathbf{z}_1 + \mathbf{z}_2 \geq 4\}$, and (iii) $\mathbf{z}^1, \mathbf{z}^2 \in \bar{\bar{Z}}$. In each subproblem, g_{ij}^{1k} is replaced by $\bar{g}_{ij}^{1k}(d) = 5d$ if $\mathbf{z}^k \in \bar{Z}$ and by $\bar{\bar{g}}_{ij} = 1.17d + 1.83$ if $\mathbf{z}^k \in \bar{\bar{Z}}$. At this time, the g-functions for each of the other arcs would be simultaneously updated for each $k \in K$. Observe that in order to construct the tightest refinement, H should be chosen parallel to the contours of $d_{ij}(\mathbf{z})$. This is easy to do when using one-norm distance, but may prove difficult if using other definitions of distance.

A major weakness of the above approach is that as many as $K + 1$ subproblems are created with each branch. Therefore, we would expect that the performance of this algorithm suffers when K is large. Indeed, as reported in Section 4.5, our experience suggests that the lower-bounding integer program (4–13) is itself difficult to solve. The computational difficulties associated with this approach motivate an alternative algorithm, which we describe in Section 4.4. □

4.4 Discrete Models

In this section, we develop a methodology for generating solutions to E-MFNIP based on replacing the continuous set $Z \subseteq \mathbb{R}^{q \times K}$ with a discrete set of locations P, a subset of which will be attacked by the leader. The resulting integer program,

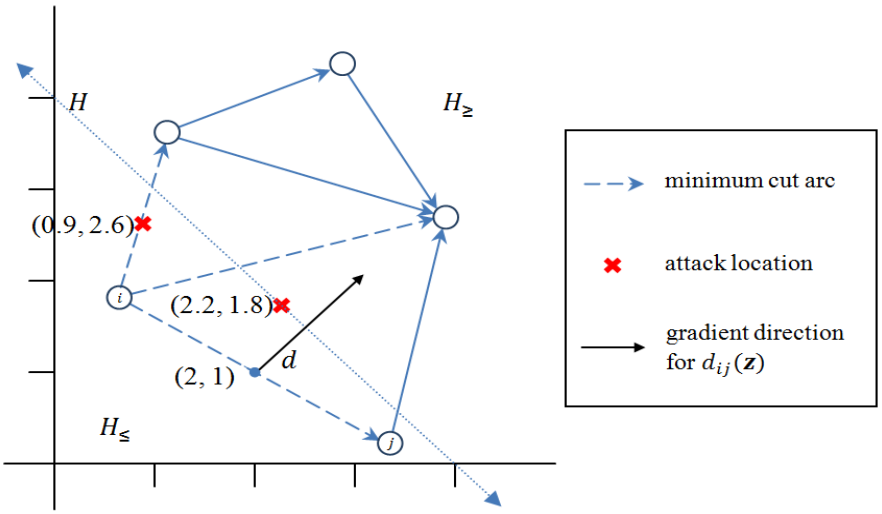

Figure 4-4. Network for partition-and-refine example.

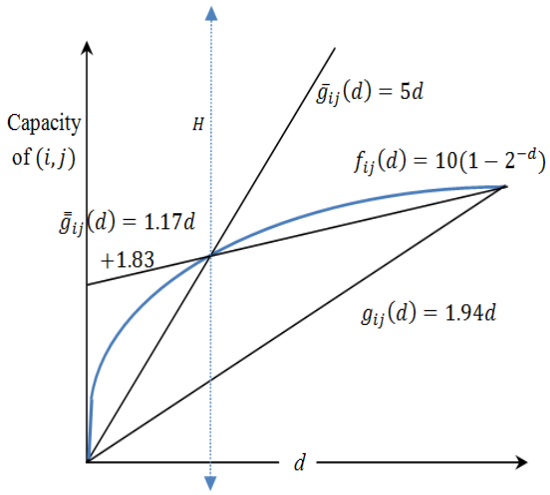

Figure 4-5. Refinement of g-functions for partition-and-refine example.

described in Section 4.4.1, is a restriction of (4–6) and therefore all feasible solutions to this problem yield upper bounds on $\nu^\star_E$. In Section 4.4.2, we propose a modification of this model that provides a lower bound on $\nu^\star_E$ by relaxing the capacity functions. Section 4.4.3 describes a methodology by which this lower bound can be refined iteratively to obtain tighter relaxations.

4.4.1 Discretized E-MFNIP Model

Define $R = \{\mathbf{z} \in \mathbb{R}^q : L \leq \mathbf{z} \leq U\}$ and consider a model that selects K attacks from among a finite set of candidate locations $P \subset R$. Suppose interdictions are placed at locations $S \subseteq P$, where $|S| = K$. From (4–8), the resulting capacity of arc $(i,j) \in A$ is given by $f_{ij}[\min_{\mathbf{p}\in S} d_{ij}(\mathbf{p})] = \min_{\mathbf{p}\in S} f_{ij}[d_{ij}(\mathbf{p})]$, because $f_{ij} : \mathbb{R}_+ \to \mathbb{R}_+$ is nondecreasing. Given P, values $c_{ij}^{\mathbf{p}} \equiv f_{ij}[d_{ij}(\mathbf{p})]$ can be computed a priori for each $\mathbf{p} \in P$ and treated as constants; thus, the objective is to select $S \subset P$ that minimizes the maximum flow over a graph in which the capacity of arc (i,j) is $\min_{\mathbf{p}\in S}\{c_{ij}^{\mathbf{p}}\}$. We refer to this *discretized* problem as DE-MFNIP. To formulate DE-MFNIP as a MIP, let variables $y^{\mathbf{p}}$ equal one if location $\mathbf{p} \in P$ is attacked and zero otherwise. The y-variables are constrained to come from the set $Y = \{y \in \{0,1\}^{|P|} : \sum_{\mathbf{p}\in P} y^{\mathbf{p}} = K\}$. For each $(i,j) \in A$, define $\bar{c}_{ij} \equiv \max_{\mathbf{p}\in P}\{c_{ij}^{\mathbf{p}}\}$ and observe that the capacity of arc (i,j) is $\min_{\mathbf{p}\in P}\{\bar{c}_{ij} - (\bar{c}_{ij} - c_{ij}^{\mathbf{p}})y^{\mathbf{p}}\}$. Given $\hat{y} \in Y$, the follower solves (4–1) with Constraints (4–1c) replaced by $x_{ij} \leq \bar{c}_{ij} - (\bar{c}_{ij} - c_{ij}^{\mathbf{p}})\hat{y}^{\mathbf{p}},\ \forall \mathbf{p} \in P$. Define $\nu_{DE}(\hat{y})$ as the maximum flow resulting from $\hat{y} \in Y$. The value of $\nu_{DE}(\hat{y})$ can be obtained alternatively by solving the following dual model, where $\gamma_{ij}^{\mathbf{p}}$ is the dual variable associated with the new capacity constraints:

$$\nu_{DE}(\hat{y}) = \min \sum_{(i,j)\in A} \sum_{\mathbf{p}\in P} [\bar{c}_{ij} - (\bar{c}_{ij} - c_{ij}^{\mathbf{p}})\hat{y}^{\mathbf{p}}]\gamma_{ij}^{\mathbf{p}}, \tag{4–17a}$$

$$\text{s.t. } (u,v) \in \mathcal{U}, \tag{4–17b}$$

$$\sum_{\mathbf{p}\in P} \gamma_{ij}^{\mathbf{p}} = v_{ij},\ \forall (i,j) \in A. \tag{4–17c}$$

Model (4–17) is a linear program; thus, if (4–17) has an optimal solution, it must have one that is complementary slack with an optimal solution to its dual. Hence, adding the constraints $\gamma_{ij}^{\mathbf{p}} \leq \hat{y}^{\mathbf{p}},\ \forall (i,j) \in A,\ \mathbf{p} \in P$, does not change the optimal objective value. This relationship also permits us to set $\gamma_{ij}^{\mathbf{p}} y^{\mathbf{p}} = \gamma_{ij}^{\mathbf{p}}$, because $\gamma_{ij}^{\mathbf{p}} \leq y^{\mathbf{p}}$ and $y^{\mathbf{p}} \in \{0,1\}$. We

now free $\hat{y}$ as a variable (y) to model DE-MFNIP as the following MIP:

$$\nu^{\star}{}_{DE}(P) = \min \sum_{(i,j)\in A} \sum_{\mathrm{p}\in P} c_{ij}^{\mathrm{p}} \gamma_{ij}^{\mathrm{p}}, \tag{4–18a}$$

$$\text{s.t.} \sum_{\mathrm{p}\in P} \gamma_{ij}^{\mathrm{p}} = v_{ij}, \ \forall (i,j) \in A, \tag{4–18b}$$

$$(u, v) \in \mathcal{U}, \tag{4–18c}$$

$$0 \leq \gamma_{ij}^{\mathrm{p}} \leq y^{\mathrm{p}}, \ \forall (i,j) \in A, \ \mathbf{p} \in P, \tag{4–18d}$$

$$y \in Y. \tag{4–18e}$$

Integer restrictions need not be placed on variables u, v, and γ in Model (4–18), as proven in the following theorem.

Theorem 4.2. There exists an optimal solution to Model (4–18) such that all u-, v-, and γ-variables are binary-valued.

Proof. Let $\hat{y}$ be any element of Y. Consider the problem that results from (4–18) when $y = \hat{y}$ is fixed. Define s^1 and s^2 as vectors of slack variables on constraints $-u_i + u_j - v_{ij} \leq 0$ and $\gamma_{ij}^{\mathrm{p}} \leq 1$, respectively, and observe that Constraints (4–18b)–(4–18d) reduce to a system of the form

$$\begin{bmatrix} C & -I & 0 & I & 0 \\ 0 & I & B & 0 & 0 \\ 0 & 0 & I & 0 & I \end{bmatrix} \begin{pmatrix} u \\ v \\ \gamma \\ s^1 \\ s^2 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}, \tag{4–19}$$

plus nonnegativity constraints on all variables except u. Define $\widetilde{M}$ as the coefficient matrix in (4–19). Observe that C is the transpose of the node-arc incidence matrix for the arc set A, and is therefore totally unimodular (TU). Note that each column of B is made up of all zeros except for the coefficient corresponding to γ_{ij}^{p}, corresponding to a particular $(i,j) \in A$ and $\mathbf{p} \in P$, which equals -1. Appending C with a sequence of

rows or columns, each containing all zeros except for a single 1 or -1, preserves total unimodularity. Therefore, the matrices

$$\begin{bmatrix} C & -I \end{bmatrix}, \begin{bmatrix} C & -I \\ 0 & I \end{bmatrix}, \begin{bmatrix} C & -I & 0 \\ 0 & I & B \end{bmatrix}, \text{ and } \begin{bmatrix} C & -I & 0 \\ 0 & I & B \\ 0 & 0 & I \end{bmatrix}, \tag{4–20}$$

are all TU submatrices of $\widetilde{M}$, and $\widetilde{M}$ is itself TU. Because all of right-hand side values in (4–19) are either 0 or 1, variables u, v, and γ must all be binary-valued in any extreme point solution of (4–18) in which y is fixed to equal $\hat{y}$. Existence of an optimal extreme point solution is guaranteed because this problem is a linear program. □

Remark 4.5. Problem DE-MFNIP remains NP-hard as we briefly illustrate here. Consider a DE-MFNIP instance defined over a network $G(N, A)$ in which the elements of A are ordered as $A = \{(i_1, j_1), \ldots, (i_{|A|}, j_{|A|})\}$ where $i_\ell \in N$ and $j_\ell \in N$ define the origin and destination nodes of the ℓ-th arc. Note that the nodes represented by $i_\ell,\ \ell \in \{1, \ldots, |A|\}$, and $j_\ell,\ \ell \in \{1, \ldots, |A|\}$, need not be distinct from each other. Suppose P is given as $P = \{\mathbf{p}^1, \ldots, \mathbf{p}^{|A|}\}$ and $c_{ij}^{\mathbf{p}^\ell}$ is given as

$$c_{ij}^{\mathbf{p}^\ell} = \begin{cases} 0 & \text{if } i = i_\ell \text{ and } j = j_\ell \\ c_{ij} & \text{otherwise,} \end{cases} \tag{4–21}$$

for $\ell \in \{1, \ldots, |A|\}$. In this case, (4–18) models the problem of finding K arcs that, when removed, result in the smallest maximum flow. This problem is exactly C-MFNIP, which is NP-hard [8], implying that DE-MFNIP is NP-hard in general. DE-MFNIP turns out to be a generalization of another NP-hard problem as well. Suppose that $v_{ij} \in \{0, 1\}$ is fixed for all $(i, j) \in A$ such that arcs $\{(i, j) \in A : v_{ij} = 1\}$ forms a cut set in G (disconnecting flow from node 1 to node n). Then, DE-MFNIP reduces to the p-median problem [68] where P is the set of potential facility locations and $\{(i, j) \in A : v_{ij} = 1\}$ is the set of customers. □

Remark 4.6. Preliminary computational experience indicates that the continuous relaxation of (4–18) can be weak. We now propose a class of valid inequalities that could be used to tighten this relaxation. We say that arc subset $\widetilde{A} \subseteq A$ is a (directed) *path* if there exists a sequence of nodes $\{i_\ell\}_{\ell=1}^{r} \subseteq \{2, \dots, n-1\}$ such that $\widetilde{A} = \{(1, i_1), (i_r, n)\} \cup \left[\bigcup_{\ell=1}^{r-1} \{(i_\ell, i_{\ell+1})\}\right]$.

For a path $\widetilde{A}$, the inequality we propose provides a lower bound on the portion of the objective (4–18a) sum arising from arcs in $\widetilde{A}$. To this end, consider a fixed $\hat{y} \in Y$ and define $\hat{P} = \{\mathbf{p} \in P : \hat{y}^{\mathsf{p}} = 1\}$. Let $\nu^\star_{\widetilde{A}}(\hat{y})$ denote the optimal objective value to Model (4–18) when A is replaced by $\widetilde{A}$ and y is fixed to $\hat{y}$, i.e.,

$$\nu^\star_{\widetilde{A}}(\hat{y}) = \min \sum_{(i,j)\in\widetilde{A}} \sum_{\mathsf{p}\in\hat{P}} c_{ij}^{\mathsf{p}} \gamma_{ij}^{\mathsf{p}}, \tag{4–22a}$$

$$\text{s.t.} \sum_{\mathsf{p}\in\hat{P}} \gamma_{ij}^{\mathsf{p}} = v_{ij}, \ \forall (i,j) \in \widetilde{A}, \tag{4–22b}$$

$$0 \le \gamma_{ij}^{\mathsf{p}} \le 1, \ \forall (i,j) \in \widetilde{A}, \mathbf{p} \in \hat{P}, \tag{4–22c}$$

$$(u, v) \in \mathcal{U}.$$

Because $\widetilde{A} \subseteq A$, the feasible region of (4–22) contains the set of solutions feasible to (4–18) when $y = \hat{y}$. Hence, if $\beta : Y \to \mathbb{R}$ is any function such that $\nu^\star_{\widetilde{A}}(y) \ge \beta(y)$ for all $y \in Y$, the inequality $\sum_{(i,j)\in\widetilde{A}} \sum_{\mathsf{p}\in P} c_{ij}^{\mathsf{p}} \gamma_{ij}^{\mathsf{p}} \ge \beta(y)$ must be valid for (4–18).

Because $\widetilde{A}$ is a path, the constraint set $(u, v) \in \mathcal{U}$ in (4–22) reduces (via aggregating constraints $u_i - u_j + v_{ij} \ge 0$) to $\sum_{(i,j)\in\widetilde{A}} v_{ij} = 1$. Due to Constraint (4–22b), $(u, v) \in \mathcal{U}$ may be replaced in (4–22) by

$$\sum_{(i,j)\in\widetilde{A}} \sum_{\mathsf{p}\in\hat{P}} \gamma_{ij}^{\mathsf{p}} = 1. \tag{4–23}$$

After doing this, note that whenever (4–22c) and (4–23) are satisfied, there exist $v_{ij}, \ (i,j) \in \widetilde{A}$, such that (4–22b) is satisfied. Therefore, (4–22b) may be dropped from Model (4–22). Consequently, solution of Model (4–22) is trivial, and the optimal objective

value is given by

$$\nu^{\star}{}_{\widetilde{A}}(\hat{y}) = \min_{\mathrm{p}\in\hat{P}} \kappa^{\mathrm{p}}, \tag{4–24}$$

where $\kappa^{\mathrm{p}} \equiv \min_{(i,j)\in\widetilde{A}} c_{ij}^{\mathrm{p}}$, $\forall \mathbf{p} \in P$. Thus, using the function $\beta(y) = \min_{\mathrm{p}\in P: y^{\mathrm{p}}=1} \kappa^{\mathrm{p}}$ produces the (nonlinear) valid inequality

$$\sum_{(i,j)\in\widetilde{A}} \sum_{\mathrm{p}\in P} c_{ij}^{\mathrm{p}} \gamma_{ij}^{\mathrm{p}} \geq \min_{\mathrm{p}\in P: y^{\mathrm{p}}=1} \kappa^{\mathrm{p}}. \tag{4–25}$$

We now linearize (4–25) by introducing auxiliary binary variables. To this end, we sort the elements of $\mathbf{p}$ according to κ^{p}: Let $\{\mathbf{p}^{<\ell>}\}_{\ell=1}^{|P|}$ denote unique elements of P such that $\kappa^{\mathrm{p}^{<1>}} \leq \cdots \leq \kappa^{\mathrm{p}^{<|P|>}}$. Define variables π^{ℓ}, $\ell = 1, \ldots, |P|$, to equal 1 if $\kappa^{\mathrm{p}^{<\ell>}}$ is selected as the minimum in (4–25) and 0 otherwise, i.e.,

$$\pi^{\ell} = \begin{cases} 1 & \text{if } y^{\mathrm{p}^{<1>}} = \cdots = y^{\mathrm{p}^{<\ell-1>}} = 0 \text{ and } y^{\mathrm{p}^{<\ell>}} = 1, \\ 0 & \text{otherwise.} \end{cases} \tag{4–26}$$

Due to the cardinality restriction imposed on $y \in Y$, it follows that $\pi^{|P|-K+2} = \pi^{|P|-K+3} = \cdots = \pi^{|P|} = 0$ and $\pi^{|P|-K+1} = 1 - \sum_{\ell=1}^{|P|-K} \pi^{\ell}$. Moreover, the relationship between π and y is captured by the linear constraints

$$\pi^{\ell} \leq y^{\mathrm{p}^{<\ell>}}, \ \forall \ell = 1, \ldots, |P| - K, \tag{4–27a}$$

$$\pi^{\ell} \leq 1 - y^{\mathrm{p}^{<\ell'>}}, \ \forall \ell = 2, \ldots, |P| - K, \ \ell' = 1, \ldots, \ell, \text{ and} \tag{4–27b}$$

$$\pi^{\ell} \geq y^{\mathrm{p}^{<\ell>}} - \sum_{\ell'<\ell} y^{\mathrm{p}^{<\ell'>}}, \ \forall \ell = 1, \ldots, |P| - K. \tag{4–27c}$$

Using the π-variables, (4–25) can be represented as

$$\sum_{(i,j)\in\widetilde{A}} \sum_{\mathrm{p}\in P} c_{ij}^{\mathrm{p}} \gamma_{ij}^{\mathrm{p}} \geq \pi^{|P|-K+1} + \sum_{\ell=1}^{|P|-K} (\kappa^{\mathrm{p}^{<\ell>}} - \kappa^{\mathrm{p}^{<|P|-K+1>}}) \pi^{\ell}. \tag{4–28}$$

Model (4–18) can now be tightened by adding (4–28) along with (4–27) and binary restrictions on the π-variables. However, this improvement to the model comes at the expense of adding $O(|P|)$ binary variables and $O(|P|^2)$ constraints. Because of the

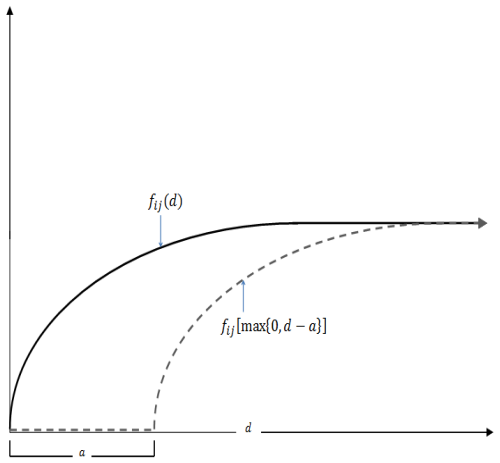

Figure 4-6. Function $f_{ij}(d)$ lower-bounded by relaxed function $f_{ij}\left[\max\{0, d-a\}\right]$.

added complexity, we suggest using a weaker version of (4–28) that includes only $\Phi < |P| - K$ π-variables by replacing $|P| - K$ and $|P| - K + 1$ in (4–27) and (4–28) with Φ and $\Phi + 1$. Based on preliminary computational results, we find that these inequalities are occasionally effective in strengthening the LP relaxation of (4–18). However, they do not appear to be effective in reducing overall computation time for solving these problems, and hence we exclude them from the computational results in Section 4.5. □

4.4.2 Using DE-MFNIP to Provide a Lower Bound for $\nu^\star{}_E$

For $a \geq 0$, the function $c^{\mathrm{p}}_{ij}(a) \equiv f_{ij}\left[\max\{0, d_{ij}(\mathbf{p}) - a\}\right]$ provides a lower bound for $c^{\mathrm{p}}_{ij}(= f_{ij} \circ d_{ij}(\mathbf{p}))$ because f_{ij} is nondecreasing. By replacing $f_{ij}(d)$ in (4–18) with $f_{ij}\left[\max\{0, d-a\}\right]$, we obtain a relaxed model that will be useful in solving E-MFNIP. As seen in Figure 4-6, the function $f_{ij}\left[\max\{0, d-a\}\right]$ lower-bounds $f_{ij}(d)$ by "stretching" the true capacity function away from the attack location by radius a. Let radii $\alpha(\mathbf{p}) \geq 0$ be given for each $\mathbf{p} \in P$ and consider the model DE(P, α) that results from replacing c^{p}_{ij} in Model (4–18) with $c^{\mathrm{p}}_{ij}(\alpha(\mathbf{p}))$:

$$\text{DE}(P,\alpha)\text{: } \nu^\star{}_{DE}(P,\alpha) = \min \sum_{(i,j)\in A} \sum_{\mathrm{p}\in P} c^{\mathrm{p}}_{ij}(\alpha(\mathbf{p}))\gamma^{\mathrm{p}}_{ij}, \tag{4–29}$$

$$\text{s.t. Constraints (4–18b)–(4–18d)}, \quad y \in Y.$$

When $\alpha(\mathbf{p}) = 0,\ \forall \mathbf{p} \in P$, DE$(P, \alpha)$ is identical to Model (4–18); thus, $\nu^\star{}_{DE}(P) = \nu^\star{}_{DE}(P, 0)$.

Theorem 4.3. For $P \subset R$, $\nu^\star{}_{DE}(P) \geq \nu^\star{}_{E}$.

Proof. Clear, due to the fact that $P \subset R$. □

DE-MFNIP is equivalent to the version of E-MFNIP that would result from restricting each $\mathbf{z}^k$ to come from P instead of from R. When $\alpha(\mathbf{p}) > 0$, we have that $c^{\mathrm{p}}_{ij}(0) \geq c^{\mathrm{p}}_{ij}(\alpha(\mathbf{p}))$. This raises the following question: How large must $\alpha(\mathbf{p})$ be before we can guarantee that $\nu^\star{}_{E} \geq \nu^\star{}_{DE}(P, \alpha)$? To address this question, for $\mathbf{p} \in \mathbb{R}^q$ and $a \geq 0$, define $B(\mathbf{p}, a)$ as the ball around $\mathbf{p}$ with radius a, i.e., $B(\mathbf{p}, a) = \{\bar{\mathbf{p}} \in \mathbb{R}^q : \|\mathbf{p} - \bar{\mathbf{p}}\|_1 \leq a\}$. For $\hat{P} \subseteq P$ and $\alpha : \hat{P} \to \mathbb{R}_+$, define $B(\hat{P}, \alpha) = \cup_{\mathrm{p} \in \hat{P}} B(\mathbf{p}, \alpha(\mathbf{p}))$.

Theorem 4.4. Suppose P is a finite subset of R and $\alpha(\mathbf{p})$ is given for each $\mathbf{p} \in P$. If $R \subseteq B(P, \alpha)$, then $\nu^\star{}_{E} \geq \nu^\star{}_{DE}(P, \alpha)$.

Proof. Let $(\hat{\mathbf{z}}, \hat{u}, \hat{v})$ denote an optimal solution to (4–4) in which $\hat{v}$ is binary-valued. For $k \in [K]$, let $\hat{\mathbf{p}}^k \in P$ be a point such that $\left\|\hat{\mathbf{p}}^k - \hat{\mathbf{z}}^k\right\|_1$ is minimized, and note that $\left\|\hat{\mathbf{p}}^k - \hat{\mathbf{z}}^k\right\|_1 \leq \alpha(\hat{\mathbf{p}}^k)$ by hypothesis. Define $\hat{A} = \{(i,j) \in A : \hat{v}_{ij} = 1\}$. Then

$$\nu^\star{}_{E} = \sum_{(i,j)\in\hat{A}} \delta_{ij}(\hat{\mathbf{z}}) = \sum_{(i,j)\in\hat{A}} f_{ij}[\min_{k\in[K]} d_{ij}(\hat{\mathbf{z}}^k)] \tag{4–30a}$$

$$= \sum_{(i,j)\in\hat{A}} \min_{k\in[K]} f_{ij}[d_{ij}(\hat{\mathbf{z}}^k)] \tag{4–30b}$$

$$\geq \sum_{(i,j)\in\hat{A}} \min_{k\in[K]} f_{ij}[\max\{0, d_{ij}(\hat{\mathbf{p}}^k) - \alpha(\hat{\mathbf{p}}^k)\}] \tag{4–30c}$$

$$= \sum_{(i,j)\in\hat{A}} \min_{k\in[K]} c^{\hat{\mathrm{p}}^k}_{ij}(\alpha(\hat{\mathbf{p}}^k)), \tag{4–30d}$$

where (4–30b) follows because f_{ij} is nondecreasing and (4–30c) follows as a result of the triangle inequality $d_{ij}(\hat{\mathbf{z}}^k) + \left\|\hat{\mathbf{p}}^k - \hat{\mathbf{z}}^k\right\|_1 \geq d_{ij}(\hat{\mathbf{p}}^k)$. Now consider the solution to DE(P, α) in which $u = \hat{u}$, $v = \hat{v}$, and all other variables are equal to zero except for

- $y^{\hat{\mathrm{p}}^k} = 1,\ \forall k \in [K]$,

- $y^{\mathbf{p}} = 1$, arbitrarily for $K - r$ additional $\mathbf{p} \in P$, where r is the number of unique points in $\{\hat{\mathbf{p}}^1, \ldots, \hat{\mathbf{p}}^K\}$.
- $\gamma_{ij}^{\hat{\mathbf{p}}^k} = 1$ corresponding to a k that minimizes $d_{ij}(\hat{\mathbf{z}}^k)$ over $k \in [K]$, $\forall (i,j) \in \hat{A}$.

This solution is feasible to DE(P, α) with objective value equal to (4–30d), and therefore $\nu^\star{}_E \geq \nu^\star{}_{DE}(P, \alpha)$. □

Combining the results of Theorems 4.3 and 4.4, we have that $\nu^\star{}_{DE}(P, \alpha) \leq \nu^\star{}_E \leq \nu^\star{}_{DE}(P)$ for any α and P such that $R \subseteq B(P, \alpha)$. Under modest assumptions on the f-functions, P and α may be chosen to guarantee that $\nu^\star{}_E - \nu^\star{}_{DE}(P, \alpha)$ is arbitrarily small, as proven in the following theorem.

Theorem 4.5. Let $P \subset R$, $\alpha : P \to \mathbb{R}_+$, and $a > 0$ be given such that $\alpha(\mathbf{p}) \leq a,\ \forall \mathbf{p} \in P$. For each $(i,j) \in A$, suppose f_{ij} is Lipschitz continuous and define $\tau_{ij} \geq 0$ as the corresponding Lipschitz constant. Then $\nu^\star{}_{DE}(P, \alpha) \geq \nu^\star{}_E - a\tau^\star$, where $\tau^\star$ is a constant that depends only on the τ-values and the network's structure.

Proof. By assumption, we have that

$$\frac{f_{ij}(d_2) - f_{ij}(d_1)}{d_2 - d_1} \leq \tau_{ij},\ \forall 0 \leq d_1 < d_2. \tag{4–31}$$

For any $d \geq 0$, (4–31) implies that

$$f_{ij}(d) - f_{ij}\left[\max\{0, d - a\}\right] \leq \tau_{ij} \min\{d, a\} \leq a\tau_{ij}. \tag{4–32}$$

We now develop a bound for E-MFNIP based on replacing $f_{ij}(d)$ with

$$f_{ij}\left[\max\{0, d - a\}\right] \geq f_{ij}(d) - a\tau_{ij}.$$

Let $\nu^\star{}_E(a)$ denote the optimal objective value to Model (4–4) if $\delta_{ij}(\mathbf{z})$ is given as $\min_{k \in [K]} \left\{ f_{ij}[\max\{0, d_{ij}(\mathbf{z}^k) - a\}] \right\}$, i.e.,

$$\nu^\star{}_E(a) = \min \sum_{(i,j) \in A} \left[\min_{k \in [K]} \left\{ f_{ij}\left[\max\{0, d_{ij}(\mathbf{z}^k) - a\}\right] \right\}\right] v_{ij}, \quad (u, v) \in \mathcal{U},\ \mathbf{z} \in Z. \tag{4–33}$$

As shown in [8], there exists an optimal solution to (4–33) in which all u- and v-variables take binary values. Moreover, given any $u \in \{0,1\}^n$, $v_{ij} = 1$ may be optimal only if $u_i = 0$ and $u_j = 0$. (Otherwise $v_{ij} = 0$ is feasible and has a smaller objective value because $\delta_{ij}(\mathbf{z}) \geq 0,\ \forall \mathbf{z} \in Z$.) Thus, it is valid to replace $\mathcal{U}$ in (4–33) with

$$\bar{\mathcal{U}} \equiv \mathcal{U} \cap \left\{ (u,v) \;\middle|\; \begin{array}{l} v_{ij} \leq u_j,\ \forall (i,j) \in A \\ v_{ij} \leq 1 - u_i,\ \forall (i,j) \in A \\ u \in \{0,1\}^n \\ v \in \{0,1\}^{|A|} \end{array} \right\}. \tag{4–34}$$

Define $\bar{\alpha}(\mathbf{p}) = a,\ \forall \mathbf{p} \in P$ and observe that $\nu^\star{}_{DE}(P,\alpha) \geq \nu^\star{}_{DE}(P,\bar{\alpha}) \geq \nu^\star{}_{E}(a)$, where the second inequality follows because DE-MFNIP is a restriction of E-MFNIP. Therefore, it follows that

$$\begin{aligned}
\nu^\star{}_{DE}(P,\alpha) \geq \quad & \nu^\star{}_{E}(a) && \text{(4–35a)} \\
= \quad & \min_{(u,v)\in\bar{\mathcal{U}},\ z\in Z} \sum_{(i,j)\in A} \left[\min_{k\in[K]} \left\{ f_{ij}\left[\max\{0, d_{ij}(\mathbf{z}^k) - a\} \right] \right\} \right] v_{ij} && \text{(4–35b)} \\
\geq \quad & \min_{(u,v)\in\bar{\mathcal{U}},\ z\in Z} \sum_{(i,j)\in A} \left[\min_{k\in[K]} \left\{ \max\{ f_{ij}(0), f_{ij}(d_{ij}(\mathbf{z}^k) - a\tau_{ij} \} \right\} \right] v_{ij} && \text{(4–35c)} \\
\geq \quad & \min_{(u,v)\in\bar{\mathcal{U}},\ z\in Z} \sum_{(i,j)\in A} \left[\min_{k\in[K]} \left\{ \max\{ f_{ij}(0) + a\tau_{ij}, f_{ij}(d_{ij}(\mathbf{z}^k)) \} \right\} \right] v_{ij} && \text{(4–35d)} \\
& + \min_{(u,v)\in\bar{\mathcal{U}},\ z\in Z} \sum_{(i,j)\in A} (-a\tau_{ij})\, v_{ij} && \text{(4–35e)} \\
\geq \quad & \nu^\star{}_{E} - a\tau^\star, && \text{(4–35f)}
\end{aligned}$$

where (i) (4–35c) follows from (4–32) because f_{ij} is nondecreasing, and (ii) boundedness of $\bar{\mathcal{U}}$ guarantees an optimal solution (with objective value $\tau^\star$) to the minimization problem in (4–35e). □

Theorem 4.5 guarantees that, with a large enough P-set and small enough α-values, the lower bound provided by $\nu^\star{}_{DE}(P,\alpha)$ is arbitrarily close to $\nu^\star{}_{E}$. However, obtaining a lower bound in this fashion is often impractical because the difficulty

assoicated with solving DE-MFNIP grows quickly as P increases. In Section 4.4.3, we propose an alternative technique for developing tight lower bounds based on iteratively building the set P. We now prove conditions under which modification of the set P guarantees an increase in $\nu^{\star}{}_{DE}(P, \alpha)$.

Theorem 4.6. Let P^1 and P^2 be disjoint subsets of R with a finite number of elements, and let $\mathbf{p}^\star \in R \setminus (P^1 \cup P^2)$. Define $\alpha : P^1 \cup \{\mathbf{p}^\star\} \cup P^2 \to \mathbb{R}_+$ and suppose $R \subseteq B(P^1 \cup \{\mathbf{p}^\star\}, \alpha)$. If $B(P^2, \alpha) = B(\mathbf{p}^\star, \alpha(\mathbf{p}^\star))$, then $\nu^{\star}{}_{DE}(P^1 \cup \{\mathbf{p}^\star\}, \alpha) \leq \nu^{\star}{}_{DE}(P^1 \cup P^2, \alpha) \leq \nu^{\star}{}_{E}$.

Proof. Because $B(\mathbf{p}^\star, \alpha(\mathbf{p}^\star)) \subseteq B(P^2, \alpha)$ and $R \subseteq B(P^1 \cup \{\mathbf{p}^\star\}, \alpha)$, it follows that $R \subseteq B(P^1 \cup P^2, \alpha)$. Therefore, $\nu^{\star}{}_{DE}(P^1 \cup P^2, \alpha) \leq \nu^{\star}{}_{E}$ follows from Theorem 4.4.

We now show that $\nu^{\star}{}_{DE}(P^1 \cup \{\mathbf{p}^\star\}, \alpha) \leq \nu^{\star}{}_{DE}(P^1 \cup P^2, \alpha)$. To this end, we show that $c_{ij}^{\mathbf{p}^\star}(\alpha(\mathbf{p}^\star)) \leq c_{ij}^{\bar{\mathbf{p}}}(\alpha(\bar{\mathbf{p}}))$ for every $\bar{\mathbf{p}} \in P^2$. If $d_{ij}(\bar{\mathbf{p}}) \leq \alpha(\bar{\mathbf{p}})$, this is trivially true because $\mathbf{p}(i,j) \in B(\bar{\mathbf{p}}, \alpha(\bar{\mathbf{p}})) \subseteq B(\mathbf{p}^\star, \alpha(\mathbf{p}^\star))$ and therefore $c_{ij}^{\mathbf{p}^\star}(\alpha(\mathbf{p}^\star)) = c_{ij}^{\bar{\mathbf{p}}}(\alpha(\bar{\mathbf{p}})) = 0$. Suppose instead that $d_{ij}(\bar{\mathbf{p}}) > \alpha(\bar{\mathbf{p}})$. Define $\hat{\mathbf{p}}^{ij} = (\alpha(\bar{\mathbf{p}})/d_{ij}(\bar{\mathbf{p}}))\bar{\mathbf{p}} + (1 - \alpha(\bar{\mathbf{p}})/d_{ij}(\bar{\mathbf{p}}))\mathbf{p}(i,j)$ and observe that $\left\|\mathbf{p}(i,j) - \hat{\mathbf{p}}^{ij}\right\|_1 = d_{ij}(\bar{\mathbf{p}}) - \alpha(\bar{\mathbf{p}})$. Moreover, $\hat{\mathbf{p}}^{ij} \in B(\mathbf{p}^\star, \alpha(\mathbf{p}^\star))$ because $B(P^2, \alpha) \subseteq B(\mathbf{p}^\star, \alpha(\mathbf{p}^\star))$, and therefore

$$d_{ij}(\mathbf{p}^\star) \leq \left\|\hat{\mathbf{p}}^{ij} - \mathbf{p}^\star\right\|_1 + \left\|\mathbf{p}(i,j) - \hat{\mathbf{p}}^{ij}\right\|_1 \tag{4–36a}$$

$$\leq \alpha(\mathbf{p}^\star) + d_{ij}(\bar{\mathbf{p}}) - \alpha(\bar{\mathbf{p}}), \tag{4–36b}$$

follows from the triangle inequality. Hence, $\max\{0, d_{ij}(\mathbf{p}^\star) - \alpha(\mathbf{p}^\star)\} \leq \max\{0, d_{ij}(\bar{\mathbf{p}}) - \alpha(\bar{\mathbf{p}})\}$, and because f_{ij} is nondecreasing, it follows that $c_{ij}^{\mathbf{p}^\star}(\alpha(\mathbf{p}^\star)) \leq c_{ij}^{\bar{\mathbf{p}}}(\alpha(\bar{\mathbf{p}}))$.

Now, let $(\hat{y}, \hat{u}, \hat{v}, \hat{\gamma})$ denote an optimal solution to DE$(P^1 \cup P^2, \alpha)$ in which all variables are integral. (Theorem 4.2 guarantees the existence of such a solution.) We now construct a solution $(\tilde{u}, \tilde{v}, \tilde{y}, \tilde{\gamma})$ to DE$(P^1 \cup \{\mathbf{p}^\star\})$ as follows. Define $\tilde{u} = \hat{u}$ and $\tilde{v} = \hat{v}$ and observe that $(\tilde{u}, \tilde{v}) \in \mathcal{U}$. Define $\tilde{y}$ as:

$$\tilde{y}^{\mathbf{p}} = \begin{cases} \hat{y}^{\mathbf{p}} & \text{if } \mathbf{p} \in P^1 \\ \max_{\bar{\mathbf{p}} \in P^2}\{\hat{y}^{\bar{\mathbf{p}}}\} & \text{if } \mathbf{p} = \mathbf{p}^\star. \end{cases} \tag{4–37}$$

If $\sum_{\mathbf{p}\in P^1\cup\{\mathbf{p}^\star\}} \tilde{y}^{\mathbf{p}} < K$, arbitrarily change enough $\tilde{y}$-variables from zero to one so that $\sum_{\mathbf{p}\in P^1\cup\{\mathbf{p}^\star\}} \tilde{y}^{\mathbf{p}} = K$, thus guaranteeing $\tilde{y} \in Y$. Now, for each (i,j), define $\tilde{\gamma}_{ij}$ as:

$$\tilde{\gamma}_{ij}^{\mathbf{p}} = \begin{cases} \hat{\gamma}_{ij}^{\mathbf{p}} & \text{if } \mathbf{p} \in P^1 \\ \sum_{\bar{\mathbf{p}}\in P^2} \hat{\gamma}_{ij}^{\bar{\mathbf{p}}} & \text{if } \mathbf{p} = \mathbf{p}^\star. \end{cases} \tag{4–38}$$

Note that $\tilde{\gamma}_{ij}^{\mathbf{p}^\star} \in \{0,1\}$ because $\sum_{\bar{\mathbf{p}}\in P^2} \hat{\gamma}_{ij}^{\bar{\mathbf{p}}} \leq \hat{v}_{ij} \leq 1$, and (4–18d) is satisfied for $\mathbf{p} \in P^1$ because $\hat{\gamma}$ and $\hat{y}$ are feasible for (4–18). When $\mathbf{p} = \mathbf{p}^\star$, this constraint is also satisfied because $\hat{\gamma}_{ij}^{\mathbf{p}} \leq \hat{y}^{\mathbf{p}}$ for all $\mathbf{p} \in P^2$ implies that either $\max_{\bar{\mathbf{p}}\in P^2}\{\hat{y}^{\bar{\mathbf{p}}}\} = 1$ or $\sum_{\bar{\mathbf{p}}\in P^2} \hat{\gamma}_{ij}^{\bar{\mathbf{p}}} = 0$. Next, we prove that (4–18b) is satisfied each of two cases: $\tilde{\gamma}_{ij}^{\mathbf{p}^\star} = 0$ (case 1) and $\tilde{\gamma}_{ij}^{\mathbf{p}^\star} = 1$ (case 2).

In case 1, we have that $\hat{\gamma}_{ij}^{\mathbf{p}} = 0,\ \forall \mathbf{p} \in P^2$. In this case, we find that $\sum_{\mathbf{p}\in P^1} \hat{\gamma}_{ij}^{\mathbf{p}} = \hat{v}_{ij} = \tilde{v}_{ij}$. Therefore, $\sum_{\mathbf{p}\in P^1\cup\{\mathbf{p}^\star\}} \tilde{\gamma}_{ij}^{\mathbf{p}} = \tilde{v}_{ij}$ and (4–18b) is satisfied. In case 2, we have that $\hat{\gamma}_{ij}^{\mathbf{p}} = 1$ for some $\mathbf{p} \in P^2$, which implies that $\sum_{\mathbf{p}\in P^1} \hat{\gamma}_{ij}^{\mathbf{p}} = 0$ and $\hat{v}_{ij} = \tilde{v}_{ij} = 1$. Thus, $\sum_{\mathbf{p}\in P^1\cup\{\mathbf{p}^\star\}} \tilde{\gamma}_{ij}^{\mathbf{p}} = \tilde{v}_{ij}$ and (4–18b) is satisfied.

We have now shown that $(\tilde{u}, \tilde{v}, \tilde{y}, \tilde{\gamma})$ satisfies all of the constraints of DE$(P^1 \cup \{\mathbf{p}^\star\}, \alpha)$. Therefore, $\nu^\star{}_{DE}(P^1 \cup \{\mathbf{p}^\star\}, \alpha)$ can be upper-bounded by the objective value corresponding to $(\tilde{y}, \tilde{u}, \tilde{v}, \tilde{\gamma})$:

$$\begin{aligned} \nu^\star{}_{DE}(P^1 \cup \{\mathbf{p}^\star\}, \alpha) &\leq \sum_{(i,j)\in A} \left(\sum_{\mathbf{p}\in P^1} c_{ij}^{\mathbf{p}}(\alpha(\mathbf{p}))\tilde{\gamma}_{ij}^{\mathbf{p}} + c_{ij}^{\mathbf{p}^\star}(\alpha(\mathbf{p}^\star))\tilde{\gamma}_{ij}^{\mathbf{p}^\star} \right) && \text{(4–39a)} \\ &\leq \sum_{(i,j)\in A} \left(\sum_{\mathbf{p}\in P^1} c_{ij}^{\mathbf{p}}(\alpha(\mathbf{p}))\hat{\gamma}_{ij}^{\mathbf{p}} + \sum_{\mathbf{p}\in P^2} c_{ij}^{\mathbf{p}}(\alpha(\mathbf{p}))\hat{\gamma}_{ij}^{\mathbf{p}} \right) && \text{(4–39b)} \\ &= \nu^\star{}_{DE}(P^1 \cup P^2, \alpha), && \text{(4–39c)} \end{aligned}$$

where (4–39b) follows because: (i) $\hat{\gamma}_{ij}^{\mathbf{p}} = 1$ for at most one $\mathbf{p} \in P^2$ for each $(i,j) \in A$; and (ii) $c_{ij}^{\mathbf{p}^\star}(\alpha(\mathbf{p}^\star)) \leq c_{ij}^{\mathbf{p}}(\alpha(\mathbf{p}))$ for all $\mathbf{p} \in P^2$ and $(i,j) \in A$. Combining the above results, we have that $\nu^\star{}_{DE}(P^1 \cup \{\mathbf{p}^\star\}, \alpha) \leq \nu^\star{}_{DE}(P^1 \cup P^2, \alpha) \leq \nu^\star{}_{E}$, completing the proof. □

An implication of Theorem 4.6 is that the contents of P can be modified in such a way that guarantees an increase in $\nu^{\star}{}_{DE}(P, \alpha)$ without losing the property that $\nu^{\star}{}_{DE}(P, \alpha) \leq \nu^{\star}{}_{E}$. This result gives rise to a methodology for solving E-MFNIP, which we describe in the following section.

4.4.3 Discretize-and-Refine Solution Methodology

We now describe how (4–18) can be used to solve instances of E-MFNIP. The methods described in this section are correct for general q, but our focus in this section is on problems in which $q = 2$ (e.g., as would be the case in power grid, telecommunications network, and transportation settings).

First, specify an initial set of points $P^0 \subset R$ and values $\alpha^0(\mathbf{p}) \in \mathbb{R}_+$ for each $\mathbf{p} \in P$ such that $R \subseteq B(P^0, \alpha^0)$. An initial lower bound for $\nu^{\star}{}_{E}$ can be obtained by solving DE(P^0, α^0). Solving this problem reveals an initial set of attack locations that are optimal for the relaxed problem. Next, set P^0 is modified (and renamed P^1) to consider a higher density of candidate locations surrounding locations that were optimal for DE(P^0, α^0). Corresponding α^1-values are assigned to the new elements of P^1 in accordance with the assumptions of Theorem 4.6 so that $\nu^{\star}{}_{DE}(P^0, \alpha^0) \leq \nu^{\star}{}_{DE}(P^1, \alpha^1)$. In this fashion, the lower bound is iteratively improved until it is within some acceptable tolerance gap of a known upper bound.

Upper bounds for $\nu^{\star}{}_{E}$ are obtained in two ways: (i) by solving DE$(P, 0)$ for any P (see Theorem 4.3), and (ii) by evaluating the objective value of any feasible solution to (4–4). Recall from Section 4.1 that $\nu([\mathbf{z}^1, \ldots, \mathbf{z}^K])$ denotes the maximum flow resulting from attacks at locations $\mathbf{z}^1, \ldots, \mathbf{z}^K$. We now provide a formal algorithmic description of our "discretize-and-refine" solution algorithm.

Step 1. Select $P^0 \subset R$ and $\alpha^0(\mathbf{p})$, $\forall \mathbf{p} \in P^0$, such that $R \subseteq B(P^0, \alpha^0)$. Set UB $= \infty$ and iteration counter $s = 0$. Let $\varepsilon > 0$ be a given tolerance parameter.

Step 2. Solve DE(P^s, α^s) and obtain an optimal solution $(\hat{y}, \hat{u}, \hat{v}, \hat{\gamma})$. Define $\hat{P} = \{\mathbf{p} \in P^s : \hat{y}^{\mathrm{p}} = 1\}$, and suppose the elements of $\hat{P}$ are $\hat{\mathbf{p}}^1, \ldots, \hat{\mathbf{p}}^K$. Compute $\nu^{\star}{}_{DE}(P^s, \alpha^s)$, $\nu^{\star}{}_{DE}(P^s)$, and $\nu^s \equiv \nu([\hat{\mathbf{p}}^1 | \ldots | \hat{\mathbf{p}}^r])$.

Step 3. If $\min\{\nu^s, \nu^\star{}_{DE}(P^s)\} < \text{UB}$, then set $\text{UB} = \min\{\nu^s, \nu^\star{}_{DE}(P^s)\}$.

Step 4. If $\text{UB} - \nu^\star{}_{DE}(P^s, \alpha^s) < \varepsilon$, then terminate the algorithm with near optimal attack locations $\hat{\mathbf{p}}^1, \ldots, \hat{\mathbf{p}}^K$.

Step 5. For each $\hat{\mathbf{p}} \in P \setminus \hat{P}$, set $\alpha^{s+1}(\hat{\mathbf{p}}) = \alpha^s(\hat{\mathbf{p}})$. For each $k \in [K]$:

(a) Construct $2q$ new points: $\mathbf{p}^{k\ell} = \hat{\mathbf{p}}^k + (\alpha^s(\hat{\mathbf{p}})/q)e_\ell,\ \ell \in [q]$, and $\mathbf{p}^{k,q+\ell} = \hat{\mathbf{p}}^k - (\alpha^s(\hat{\mathbf{p}})/q)e_\ell,\ \ell \in [q]$. (Note: e_ℓ is the q-vector with a one in component ℓ and zeros in all other components.)

(b) Define $\alpha^{s+1}(\mathbf{p}^{k\ell}) = \alpha^s(\hat{\mathbf{p}})(1 - 1/q)$ for all $k \in [K],\ \ell \in [2q]$.

Step 6. Set $P^{s+1} = (P^s \setminus \hat{P}) \cup \left(\cup_{k\in[K],\ell\in[2q]}\{\mathbf{p}^{k\ell}\}\right)$. Set $s = s + 1$ and return to Step 2.

In moving from iteration s to iteration $s + 1$, the property that $R \subseteq B(P^{s+1}, \alpha^{s+1})$ is maintained because (i) $R \subseteq B(P^s, \alpha^s)$ and (ii) $B(\hat{\mathbf{p}}^k, \alpha^s) \subseteq B(\{\mathbf{p}^{k\ell}\}_{\ell\in[2q]}, \alpha^{s+1})$, $\forall k \in [K]$. Thus, we guarantee that $\nu^\star{}_{DE}(P^s, \alpha^s) \leq \nu^\star{}_E,\ \forall s$, by Theorem 4.4. Moreover, $\cup_{\ell\in[2q]} B(\mathbf{p}^{k\ell}, \alpha^{s+1}(\mathbf{p}^{k\ell})) = B(\hat{\mathbf{p}}^k, \alpha^s(\hat{\mathbf{p}}^k))$; thus, $B(\{\mathbf{p}^{k\ell}\}_{\ell\in[2q]}, \alpha^{s+1}) \subseteq B(\hat{\mathbf{p}}^k, \alpha^s)$, $\forall k \in [K]$, implying that $\nu^\star{}_{DE}(P^0, \alpha^0) \leq \nu^\star{}_{DE}(P^1, \alpha^1) \leq \cdots \leq \nu^\star{}_{DE}(P^s, \alpha^s) \leq \nu^\star{}_E,\ \forall s$, by Theorem 4.6.

Remark 4.7. In Step 1 of the above algorithm, there are many ways to select P^0 and α^0. Given a selection of P^0, one could choose $\alpha^0(\mathbf{p}) = r,\ \forall \mathbf{p} \in P^0$, where $r > 0$ is the smallest value such that $R \subseteq B(P^0, r)$. Note, however, that the lower bound provided by (4–18) is generally tighter when α is smaller. In order to minimize the value of r needed to ensure that $R \subseteq B(P^0, r)$, one should select the elements of P^0 to be somehow "evenly spaced." Viewed from another perspective, for fixed $r > 0$, one should construct P^0 using the fewest possible points so that $R \subseteq B(P^0, r)$. For $q = 2$ and $R = \{\mathbf{z} : L \leq \mathbf{z} \leq U\}$, selecting $P^0 = \{L - r < \mathbf{z} < U + r :\ z_1 = kr,\ z_1 + z_2 = 2\ell r,\ k, \ell \in \mathbb{Z}\}$ guarantees $R \subseteq B(P^0, r)$ with with no overlap between $B(\mathbf{p}, r)$ and $B(\bar{\mathbf{p}}, r)$, for any $\mathbf{p}, \bar{\mathbf{p}} \in P^0$. □

4.5 Computational Results

In this section, we summarize computational results obtained from implementing the methods of Sections 4.3 and 4.4 on two-dimensional undirected test networks. We begin

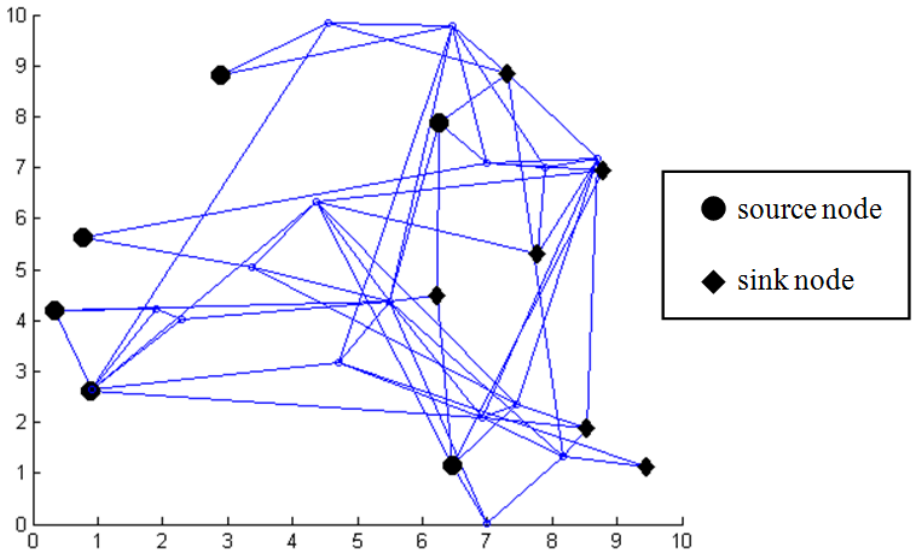

Figure 4-7. Test network.

by comparing the quality of E-MFNIP solutions on a test network to solutions which could be obtained by solving the MFNIP model of [8]. Figure 4-7 displays a network with six source nodes, six sink nodes, 16 intermediate nodes, and 67 undirected edges represented using $2 \times 67 = 134$ directed arcs. The network is set inside a ten-by-ten two-dimensional square. Each edge (i, j) is assigned a nominal capacity $c_{ij} \in \{1, \ldots, 10\}$ and the capacity of edge (i, j) is computed as $c_{ij}(1 - b^d)$ where d is the distance to the nearest attack and $0 < b < 1$ is a constant. For small values of b, each arc's capacity is close to c_{ij} except when d is close to zero; thus, we would expect the E-MFNIP optimal solution to resemble an optimal solution for MFNIP at these values of b. When b is larger, each attack has effect over a greater area, and thus, we would expect E-MFNIP solutions to differ significantly from MFNIP.

We solve E-MFNIP (using discretize-and-refine) and MFNIP (using the model of [8]) using $K = 3$ attacks over this network for each $b \in \{0.01, 0.1, 0.3, 0.5, 0.7, 0.9\}$ and compare the best E-MFNIP solution for each case to the E-MFNIP solution that would result from attacking the midpoints of each arc attacked in the MFNIP solution.

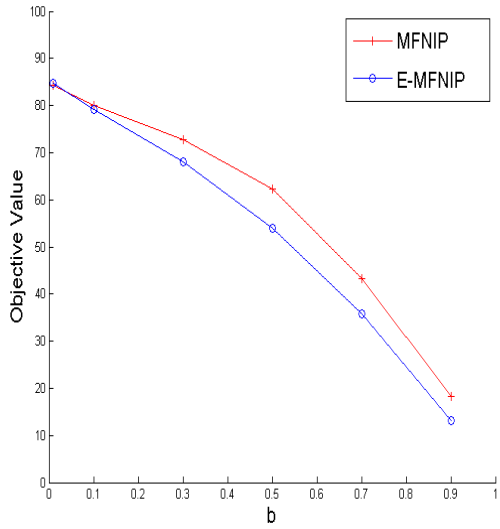

Figure 4-8. Comparison of objective values as b is varied.

Each E-MFNIP instance was solved to within 2% of optimality. Figure 4-8 plots the the resulting E-MFNIP objective values against b. As anticipated, the gap between E-MFNIP and MFNIP increases as b increases. When $b = 0.9$, the E-MFNIP solution results in a maximum flow value that is 29% less than the flow resulting from the MFNIP solution. Figure 4-9 displays the optimal E-MFNIP solutions (representing attack locations with a star) for $b = 0.01, 0.1, 0.3, \ldots, 0.9$. The MFNIP solution is not pictured because it is almost identical to the E-MFNIP solution for $b = 0.01$. As b increases, the optimal attack locations move from a closely-packed formation to a more spread out formation. This behavior has an intuitive explanation: When b is large, attacks have a significant effect on a larger range of distances meaning that a pair of attacks may have a diminished effect due to redundancy if located too close together.

We now compare the discretized methods of Section 4.4 with the lower-bounding model of Section 4.3. All of the computational runs in this result set were executed on a Dell Power Edge 2600 machine with two Pentium 4 3.2GHz/1M cache processors, using CPLEX 12.1 to solve all of the integer programming models. The discretize-and-refine procedure, coded in C++, uses the Concert Technology 2.9 library to solve a sequence

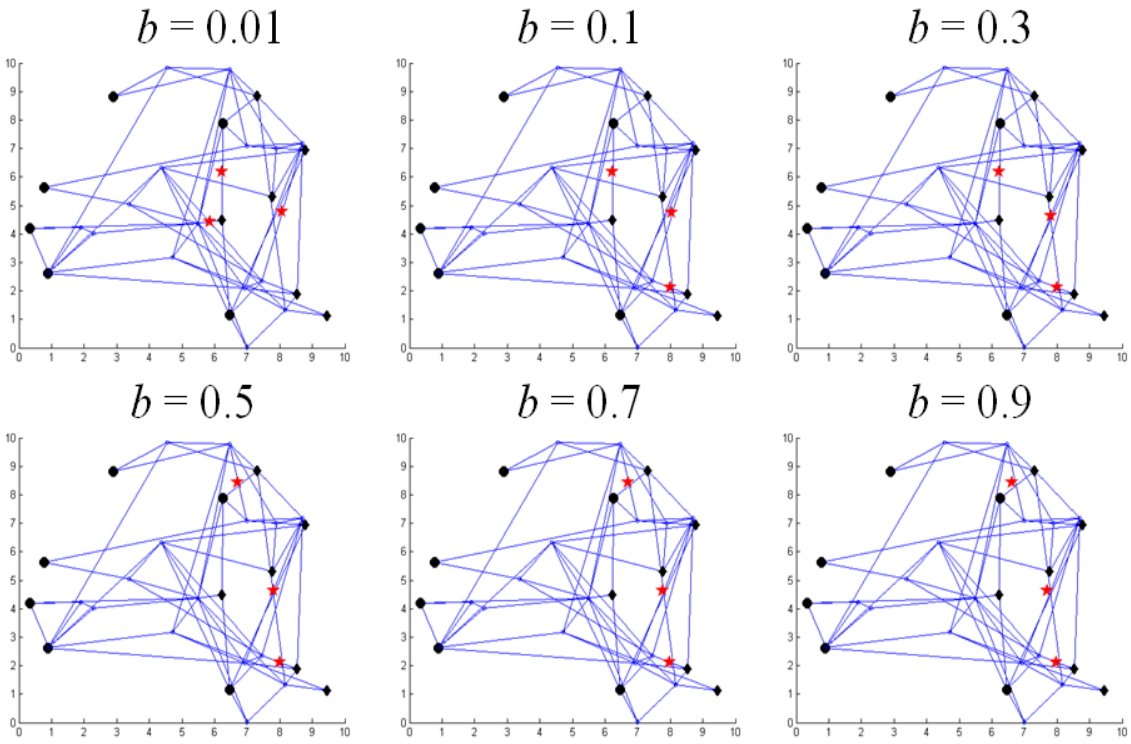

Figure 4-9. Plot of attack locations as capacity parameter b is varied.

of integer programs with the same version of CPLEX. Networks were constructed by randomly locating nodes inside a ten by ten square and then adding an edge between each pair of nodes with a certain probability. The capacity of each edge (i, j) is computed as in the example network of Figure 4-7 using $b = 0.5$.

We now describe in further detail the six-step process used to generate these networks. One, parameters $n \in \mathbb{Z}_+$, $S \in \mathbb{Z}_+$, and $\eta \geq 1$ are specified a priori to control the number of nodes, number of sources and sinks, and the arc density, respectively. Two, the two-dimensional square $R = \{\mathbf{z} \in \mathbb{R}^2 : \ 0 \leq z_1 \leq 10, \ 0 \leq z_2 \leq 10\}$ is divided into S equally sized bands, $B_\sigma = \{\mathbf{z} \in \mathbb{R}^2 : \ 0 \leq z_1 \leq 10, \ (\sigma - 1)(S/10) \leq z_2 \leq \sigma(S/10)\}, \ \sigma \in [S]$. Three, node locations $\mathbf{p}(i), \ i \in [n]$ are chosen uniformly in R such that at least two nodes are located in B_σ for each $\sigma \in [S]$. Four, for each $\sigma \in [S]$, the left-most node in B_σ (i.e., the one with the smallest z_1-coordinate) is designated as a source node, and the right-most node is designated as a sink. Five, for each $\sigma \in [S]$, edges are constructed to

Table 4-1. Random network generation profiles.

Profile	n	S	η	Arcs
A	28	6	1.75	78
B	28	10	1.75	58
C	28	6	1.40	130
D	28	10	1.40	140
E	43	6	2.00	134
F	43	10	2.00	116
G	43	6	1.60	228
H	43	10	1.60	200

connect all of the nodes located in B_σ by a single path that passes through each of the nodes from left to right. Six, for each $i, j \in [n]$ such that $i < j$, edge (i, j) is created with probability $\eta^{-\|\mathbf{p}(i)-\mathbf{p}(j)\|_1}$.

Using the process described above, an instance was generated from each of eight different network profiles, A through H. A summary of the profiles is provided in Table 4-1. Note that the numbers of nodes and arcs reported in this table do not include additional nodes and arcs used to convert the problem to one having a single source and sink, and that the network of Figure 4-7 was generated from profile C.

We solve each of the eight test networks for $K = 1, 3, 5, 7$ using the following approaches based upon the models in Sections 4.3 and 4.4:

1. **Lower-bounding Model** (4–13)**:** We solve this model using $T = 1$ and $T = 2$ linear underestimating g-functions for each arc. Following (4–14), the g-functions are computed using breakpoints $a_0 = 0$, $a_1 = d^\star$ ($a_0 = 0$, $a_1 = 2$, and $a_2 = d^\star$) for $T = 1$ ($T = 2$), where $d^\star = \min\{d : d \geq d_{ij}(\mathbf{z})\ \forall \mathbf{z} \in Z\}$. The algorithms corresponding to $T = 1$ and $T = 2$ are labeled as [L:T1] and [L:T2], respectively. For these algorithms, we report the best known lower bound $\underline{\mathbf{O}}$.

2. **Lower-bounding Model** (4–18)**:** We solve DE(P, α), where P and α are chosen according to Remark 4.7 for $r = 1$ and $r = 0.5$ (denoted [L:r1.0] and [L:r0.5]). We report the best known lower bound $\underline{\mathbf{O}}$.

3. **Upper-bounding Model** (4–18)**:** We solve DE$(P, 0)$, where P is chosen according to Remark 4.7 for $r = 1$ and $r = 0.5$ (denoted [U:r1.0] and [U:r0.5]). We also solve DE$(P, 0)$ using $P = \{\mathbf{p}(i, j) : (i, j) \in A\}$ (denoted [U:mp]). For each profile, we report the the best known upper bound $\overline{\mathbf{O}}$.

4. **Discretize-and-refine:** We use the approach described in Section 4.4.3, where P^0 and α^0 are chosen according to Remark 4.7 using $r = 1$. We report $\overline{\mathbf{O}}$ and $\underline{\mathbf{O}}$, the best known upper and lower bounds, respectively.

Discretize-and-refine instances were solved to within one percent of optimality or terminated after two hours. For each instance, we report the solution time **T(s)** (or the optimality gap for instances not complete within two hours). We report "$<$" for instances that solved within one second and "–" for instances that terminated with an optimality gap in excess of 1000%.

The results of these experiments are contained in Tables 4-2 and 4-3. For each instance, the greatest lower bound on the optimal objective value is indicated in italics and the least upper bound is indicated in bold. The discretize-and-refine method [D&R] appears to be an effective approach, identifying near-optimal solutions (and proving their near-optimality) in 20 of 32 instances. We provide a more detailed view of [D&R] by closely examining the algorithm's statistics for one of the test instances. Figure 4-10 displays the upper- and lower-bound objective values at each iteration of [D&R] for the F network with $K = 5$. This instance solved in 37 iterations, requiring a total of 6674 seconds. Figure 4-11 breaks down the solution time for each iteration by three subroutines: (i) Lower-bound model solution, (ii) Upper-bound model solution, and (iii) model update between iterations. The solution time for the lower-bound model grows quickly as the algorithm progresses, dominating the time for both the upper-bound solution and model update. That is, solutions of good quality are found relatively quickly in [D&R], but proving optimality requires a significant amount of computation. Because discretize-and-refine converges to an optimal solution, [D&R] provides a standard by which to compare the other solution approaches.

The methods derived from (discrete) Model (4–18) consistently provide much better bounds on the E-MFNIP solution than do either of the methods derived from Model (4–13). In most instances with $K > 1$, [L:T1] and [L:T2] fail to produce a strictly positive lower bound within two hours. This may be due to the weakness

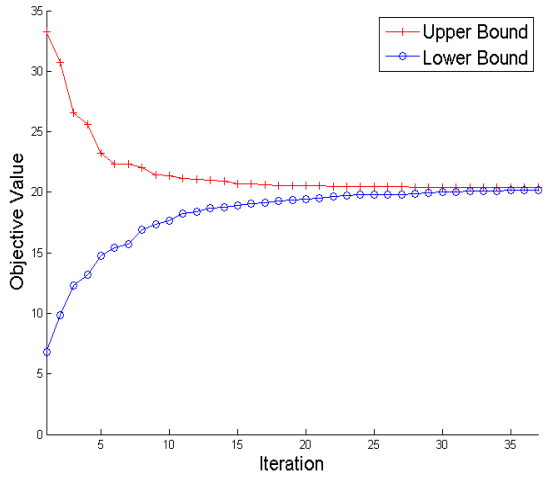

Figure 4-10. Upper and lower bounds obtained at each iteration of discretize-and-refine for F network with $K = 5$ attacks.

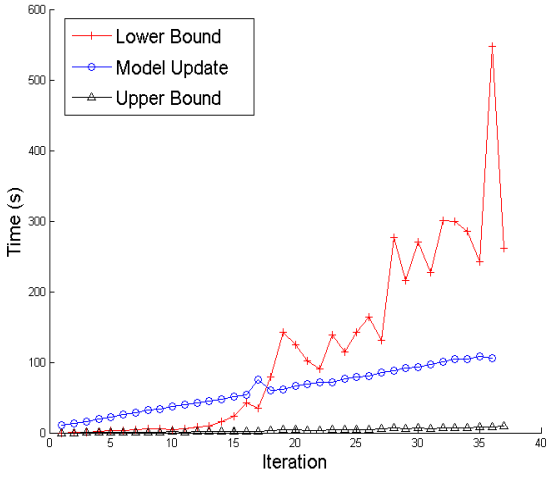

Figure 4-11. Solution times for three subroutines at each iteration of discretize-and-refine.

of Model (4–13) resulting from the inclusion of multiple levels of big M constraints. Algorithms [L:r1.0], [L:r0.5], [U:r1.0], [U:r0.5], and [U:mp]—all single-solve implementations of Model (4–18)—consistently solve within seconds and provide much tighter bounds, even in the cases ([L:r1.0] and [U:r1.0]) that use the smallest P-set. Algorithm [U:mp] produces the best solution among all algorithms in 27 of 32 instances, suggesting that solving Model (4–18), with P defined as the set of arc midpoints, may be an

effective heuristic method for identifying E-MFNIP solutions. However, we caution future researchers against relying solely upon [U:mp] to solve E-MFNIP because (i) [U:mp] provides no lower bound and therefore no guarantee of optimality, and (ii) our experience suggests that the quality of [U:mp] solutions is heavily dependent upon the type of f-functions used. For instance, consider any network in which the distance between any two arc midpoints is at least $r > 0$ and the capacity functions are given as

$$f_{ij}(d) = \begin{cases} 0 & d \leq r \\ c_{ij} & d > r, \end{cases} \tag{4–40}$$

for all $(i,j) \in A$. Thus, if (i,j) and (i',j') are unique arcs such that $r < \|\mathbf{p}(i,j) - \mathbf{p}(i',j')\|_1 \leq 2r$, an attack at $0.5(\mathbf{p}(i,j) + \mathbf{p}(i',j'))$ effectively removes both arcs while an attack at $\mathbf{p}(i,j)$ or $\mathbf{p}(i',j')$ would only remove one.

The results for [U:r1.0] and [U:r0.5] suggest that high-quality solutions can be obtained by solving a single instance of (4–18) if a large enough P-set is used. The number of variables and constraints in Model (4–18) is $O(P|A|) = O(|A|/r^2)$, so increasing P (or decreasing r) can quickly lengthen model setup and solution times and increase memory requirements. Alternatively, given that the lower-bounding Model (4–18) appears to be much more difficult to solve than the corresponding upper-bounding model, another possible heuristic approach is to apply the discretize-and-refine algorithm of Section 4.4.3 to the upper-bounding model only.

Table 4-2. Computational results for networks A through D.

		[L:T1]		[L:T2]		[L:r1.0]		[L:r0.5]		[U:r1.0]		[U:r0.5]		[U:mp]		[D&R]		
Net.	K	$T(s)$	$\underline{O}$	$T(s)$	$\underline{O}$	$T(s)$	$\underline{O}$	$T(s)$	$\underline{O}$	$T(s)$	$\overline{O}$	$T(s)$	$\overline{O}$	$T(s)$	$\overline{O}$	$T(s)$	$\underline{O}$	$\overline{O}$
A	1	$<$	6.49	74	20.5	$<$	18.8	1	20.6	$<$	28.0	1	25.7	$<$	**23.8**	61	*23.6*	23.8
	3	3226	2.51	428%	1.70	$<$	4.30	1	7.28	$<$	17.8	1	16.3	$<$	**10.8**	200	*10.7*	10.8
	5	–	0.00	–	0.00	$<$	0.61	3	2.26	$<$	14.9	4	11.5	$<$	**4.46**	789	*4.46*	4.49
	7	–	*0.00*	–	*0.00*	$<$	*0.00*	1	*0.00*	1	14.0	2	9.18	$<$	**0.00**	1142	*0.00*	**0.00**
B	1	$<$	20.6	44	53.5	$<$	48.6	4	55.9	$<$	62.2	2	62.2	1	**61.3**	215	*60.7*	61.4
	3	44.3%	6.51	273%	8.99	$<$	20.8	2	31.3	$<$	47.7	2	45.1	1	**39.9**	2783	*39.6*	40.0
	5	–	0.00	–	0.00	$<$	12.7	2	20.0	$<$	40.9	3	35.7	2	**28.3**	5339	*28.1*	28.4
	7	–	0.00	–	0.00	$<$	6.59	3	12.2	$<$	35.4	2	27.8	1	**18.6**	1111	*18.6*	18.7
C	1	7	25.3	36.7%	47.5	1	62.4	4	66.9	$<$	81.2	4	77.1	6	**74.2**	157	*73.6*	74.3
	3	339%	2.39	–	0.13	1	22.9	2	33.8	1	60.6	1	51.5	2	**45.7**	4054	*45.4*	45.8
	5	–	0.00	–	0.00	$<$	11.2	5	20.7	1	50.9	3	41.3	2	**31.4**	1.13%	*31.1*	31.4
	7	–	0.00	–	0.00	$<$	5.25	5	11.8	$<$	45.2	1	33.8	7	**21.8**	3.21%	*21.2*	21.9
D	1	2	46.0	39.4%	97.1	2	131.0	13	143.6	1	161.1	12	158.5	5	157.3	4846	*155.7*	**157.3**
	3	281%	6.89	–	0.31	2	67.4	11	91.0	$<$	121.5	3	120.8	13	**116.7**	5.34%	*110.9*	116.8
	5	–	0.00	–	0.00	2	38.4	13	62.9	$<$	104.1	3	99.5	4	**92.2**	10.2%	*84.1*	92.7
	7	–	0.00	–	0.00	1	22.5	11	43.9	$<$	91.4	2	85.0	1	**73.6**	11.8%	*66.9*	74.7

Table 4-3. Computational results for networks E through H.

		[L:T1]		[L:T2]		[L:r1.0]		[L:r0.5]		[U:r1.0]		[U:r0.5]		[U:mp]		[D&R]		
Net.	K	$T(s)$	$\underline{O}$	$T(s)$	$\underline{O}$	$T(s)$	$\underline{O}$	$T(s)$	$\underline{O}$	$T(s)$	$\overline{O}$	$T(s)$	$\overline{O}$	$T(s)$	$\overline{O}$	$T(s)$	$\underline{O}$	$\overline{O}$
E	1	7	8.92	6766	24.3	1	21.1	8	23.3	$<$	29.0	3	27.4	12	**26.6**	308	*26.5*	26.7
	3	449%	0.48	–	0.00	$<$	5.45	1	8.69	1	19.2	1	15.7	2	**11.7**	1504	*11.6*	11.7
	5	–	0.00	–	0.00	$<$	0.78	2	2.72	1	14.6	1	10.3	1	**3.39**	738	*3.39*	3.42
	7	–	0.00	–	0.00	$<$	0.00	1	0.14	1	12.6	1	8.63	1	**0.808**	3472	*0.806*	0.810
F	1	13	15.1	23.2%	38.7	1	39.5	7	47.2	1	54.5	6	53.9	3	**52.6**	2036	*52.2*	52.6
	3	726%	0.88	–	1.43	1	19.4	6	24.1	1	40.2	4	38.5	7	**33.2**	1.92%	*32.7*	33.3
	5	–	0.00	–	0.00	$<$	6.83	2	12.0	$<$	33.2	$<$	28.4	1	**20.3**	6674	*20.2*	20.4
	7	–	0.00	–	0.00	$<$	2.92	15	7.25	1	29.3	3	24.5	1	**13.8**	4875	*13.7*	13.8
G	1	47	16.4	296%	12.1	2	38.3	10	45.5	$<$	57.1	12	56.6	49	55.8	3826	*54.6*	**55.1**
	3	–	0.00	–	0.00	2	12.0	7	19.9	1	40.7	6	38.6	21	33.8	4.89%	*32.1*	**33.7**
	5	–	0.00	–	0.00	2	3.08	4	7.14	$<$	30.3	3	24.7	9	17.7	11.4%	*15.8*	**17.6**
	7	–	0.00	–	0.00	1	0.00	3	2.39	$<$	26.8	2	18.7	3	**9.26**	13.5%	*8.61*	9.77
H	1	46	25.2	64.6%	43.7	10	68.8	16	75.0	3	85.9	16	84.0	27	84.1	7060	*83.0*	**83.8**
	3	–	0.00	–	0.00	2	27.0	15	38.8	1	62.7	12	57.3	26	**55.2**	7.67%	*51.6*	55.6
	5	–	0.00	–	0.00	4	13.6	3	23.4	1	51.8	3	43.9	6	**38.4**	5.76%	*36.7*	38.8
	7	–	0.00	–	0.00	2	6.66	16	14.9	1	45.7	11	38.4	28	**29.9**	16.34%	*27.2*	31.6

CHAPTER 5
CONCLUSION

In this dissertation, we perform in-depth analysis of several interdiction problems. In Chapter 2, we extend the BiPSNIP model of [32] in a number of ways, beginning with a reformulation of the problem that has fewer constraints and has a tighter continuous relaxation. We develop a class of facet-defining inequalities and a corresponding separation procedure. These inequalities generalize the BiSNIP inequalities given in [32], and they appear to significantly improve computational efficiency. We solve an instance derived from the U.S. road network and discuss the results. The difficulty of the problem instances investigated here lend evidence to the necessity of developing effective heuristic techniques for this class of problems. Another interesting follow-on investigation may seek to find approximation algorithms that yield high-quality solutions using reasonable computational effort.

In Chapter 3, we analyze further analyze the BiPSNIP model by characterizing the convex hull of integer-feasible solutions to the single-scenario cardinality-constrained problem. This polyhedron has an exponential number of facets that arise due to the inclusion of a constraint limiting the number of interdictions. The cuts defining the single-scenario polyhedron are then reconsidered from the perspective of solving a multiple-scenario problem. Consistent with the NP-hardness result from [51], we argue that the multiple-scenario polyhedron may contain facets that are difficult to separate; however, the single-scenario cuts are still able to improve the LP relaxation significantly when B is large relative to n.

Our experience is that the multiple-scenario polyhedron varies drastically depending on the relative ordering of $\bar{p}$-values in the different scenarios. An interesting experiment could compare instances in which the $\bar{p}$-values from different scenarios were correlated, and report results similar to those given in Table 3-2. Moreover, although the one-scenario

cuts clearly tighten the LP relaxation in many instances, efficient implementation procedures have not been studied extensively. We leave these studies for future work.

In Chapter 4, we develop a new model for maximum flow interdiction which explicitly models the geometric relationship between an attack and network components. The problem we model, denoted E-MFNIP, generalizes the traditional discrete maximum flow interdiction problem, which is known to be NP-hard. Due to the structure of the capacity functions we use, existence of a compact single-stage integer linear programming formulation is unlikely so we develop alternate solution approaches which fall in two classes: (i) methods from Section 4.3 that use continuous variables to select attack coordinates, and (ii) methods derived from solving a discretized version of E-MFNIP, denoted DE-MFNIP. Computational results indicate that the methods derived from DE-MFNIP are much more effective in obtaining high-quality solutions to E-MFNIP than the Section 4.3 methods and are much faster, too.

Among the discretization-based approaches we develop, the central contributions of this chapter are DE-MFNIP models that give upper and lower bounds for E-MFNIP. We show how these models can be combined into a discretize-and-refine algorithm to obtain optimal solutions for E-MFNIP. This algorithm successfully identifies optimal solutions, but requires iterative solution of DE-MFNIP instances that become progressively more difficult to solve. Most of the effort associated with discretize-and-refine comes from improving the lower bound that ultimately provides the guarantee of optimality, suggesting that heuristic approaches based on solving DE-MFNIP may also be effective.

Based on the analysis in Chapter 4, many of the most effective methods for solving E-MFNIP are based on solving DE-MFNIP. Thus, future research developing improved solution techniques for DE-MFNIP could be of benefit in solving E-MFNIP. We identify some potential discretization-based E-MFNIP heuristics, but other types of heuristics and approximations could prove useful as well. Additional investigations may

seek to find best implementations of the algorithms we propose as well as extend our techniques to different types of distance and capacity functions.

REFERENCES

[1] R. D. Wollmer, "Removing arcs from a network," *Operations Research*, vol. 12, pp. 934–940, 1964.

[2] A. McMasters and T. Mustin, "Optimal interdiction of a supply network," *Naval Research Logistics Quarterly*, vol. 17, pp. 261–268, 1970.

[3] United Nations. (2012, June) Treaty on the non-proliferation of nuclear weapons. [Online]. Available: http://www.un.org/disarmament/WMD/Nuclear/NPT.shtml

[4] S. H. Lubore, H. D. Ratliff, and G. T. Sicilia, "Determining the most vital link in a flow network," *Naval Research Logistics Quarterly*, vol. 18, pp. 497–502, 1971.

[5] P. M. Ghare, D. C. Montgomery, and W. C. Turner, "Optimal interdiction policy for a flow network," *Naval Research Logistics Quarterly*, vol. 18, pp. 37–45, 1971.

[6] H. D. Ratliff, G. T. Sicilia, and S. H. Lubore, "Finding the n most vital links in flow networks," *Management Science*, vol. 21, pp. 531–539, 1975.

[7] H. W. Corley and H. Chang, "Finding the n most vital nodes in a flow network," *Management Science*, vol. 21, pp. 362–364, 1974.

[8] R. K. Wood, "Deterministic network interdiction," *Mathematical and Computer Modeling*, vol. 17, pp. 1–18, 1993.

[9] C. A. Phillips, "The network inhibition problem," in *Proceedings of the 25th Annual ACM Symposium on the Theory of Computing*, 1993, pp. 776–785.

[10] J. O. Royset and R. K. Wood, "Solving the bi-objective maximum-flow network-interdiction problem," *INFORMS Journal on Computing*, vol. 19, pp. 175–184, 2007.

[11] D. S. Altner, O. Ergun, and N. A. Uhan, "The maximum flow network interdiction problem: Valid inequalities, integrality gaps, and approximability," *Operations Research Letters*, vol. 38, pp. 33–38, 2010.

[12] D. Fulkerson and G. Harding, "Maximizing the minimum source-sink path subject to a budget constraint," *Mathematical Programming*, vol. 13, pp. 116–118, 1977.

[13] B. Golden, "A problem in network interdiction," *Naval Research Logistics Quarterly*, vol. 25, pp. 711–713, 1978.

[14] M. O. Ball, B. L. Golden, and R. V. Vohra, "Finding the most vital arcs in a network," *Operations Research Letters*, vol. 8, pp. 73–76, 1989.

[15] H. W. Corley and D. Y. Sha, "Most vital links and nodes in weighted networks," *Operations Research Letters*, vol. 1, pp. 157–160, 1982.

[16] K. Malik, A. Mittal, and S. Gupta, "The k most vital arcs in the shortest path problem," *Operations Research Letters*, vol. 8, pp. 223–227, 1989.

[17] A. Bar-Noy, S. Khuller, and B. Schieber, "The complexity of finding most vital arcs and nodes," University of Maryland, Institute of Advanced Computer Studies, College Park, MD, Tech. Rep., 1995.

[18] L. Khachiyan, E. Boros, K. Borys, K. Elbassioni, V. Gurvich, G. Rudolf, and J. Zhao, "On short paths interdiction problems: Total and node-wise limited interdiction," *Theory of Computing Systems*, vol. 43, pp. 204–233, 2008.

[19] E. Israeli and R. K. Wood, "Shortest-path network interdiction," *Networks*, vol. 40, pp. 97–111, 2002.

[20] R. D. Wollmer, "Algorithms for targeting strikes in a lines-of-communication network," *Operations Research*, vol. 18, pp. 497–515, 1970.

[21] C. Lim and J. C. Smith, "Algorithms for discrete and continuous multicommodity flow network interdiction problems," *IIE Transactions*, vol. 39, pp. 15–26, 2007.

[22] J. C. Smith, C. Lim, and F. Sudargho, "Survivable network design under optimal and heuristic interdiction scenarios," *Journal of Global Optimization*, vol. 38, pp. 181–199, 2007.

[23] R. L. Church, M. P. Scaparra, and R. S. Middleton, "Identifying critical infrastructure: The median and covering facility interdiction problems," *Annals of the Association of American Geographers*, vol. 94, pp. 491–502, 2004.

[24] R. L. Church and M. P. Scaparra, "The r-interdiction median problem with fortification," *Geographical Analysis*, vol. 39, pp. 129–146, 2007.

[25] F. Liberatore, M. P. Scaparra, and M. S. Daskin, "Analysis of facility protection strategies against an uncertain number of attacks: The stochastic r-interdiction median problem with fortification," *Computers and Operations Research*, vol. 38, pp. 357–366, 2011.

[26] M. P. Scaparra and R. L. Church, "A bilevel mixed-integer program for critical infrastructure protection planning," *Computers and Operations Research*, vol. 35, pp. 1905–1923, 2008.

[27] ——, "An exact solution approach for the interdiction median problem with fortification," *European Journal of Operations Research*, vol. 189, pp. 76–92, 2008.

[28] K. J. Cormican, D. P. Morton, and R. K. Wood, "Stochastic network interdiction," *Operations Research*, vol. 46, pp. 184–197, 1998.

[29] U. Janjarassuk and J. T. Linderoth, "Reformulation and sampling to solve a stochastic network interdiction problem," *Networks*, vol. 52, pp. 120–132, 2008.

[30] O. Berman, D. Krass, and C. W. Xu, “Locating flow-intercepting facilities: new approaches and results,” *Annals of Operations Research*, vol. 60, pp. 121–143, 1995.

[31] A. Gutfraind, A. Hagberg, and F. Pan, “Optimal interdiction of unreactive Markovian evaders,” in *Proceedings of the 6th International Conference on Integration of AI and OR Techniques in Constraint Programming for Combinatorial Optimization Problems*, Springer, Berlin, 2009, pp. 102–116.

[32] D. P. Morton, F. Pan, and K. J. Saeger, “Models for nuclear smuggling interdiction,” *IIE Transactions*, vol. 39, pp. 3–14, 2007.

[33] F. Pan and D. P. Morton, “Minimizing a stochastic maximum-reliability path,” *Networks*, vol. 52, pp. 111–119, 2008.

[34] M. D. Bailey, S. M. Shechter, and A. J. Schaefer, “SPAR: stochastic programming with adversarial recourse,” *Operations Research Letters*, vol. 34, pp. 307–315, 2006.

[35] R. Hemmecke, R. Schultz, and D. L. Woodruff, “Interdicting stochastic networks with binary interdiction effort,” in *Network Interdiction and Stochastic Integer Programming*, D. L. Woodruff, Ed. Kluwer, Boston, MA, 2002.

[36] D. L. W. Harald Held, Raymond Hemmecke, “A decomposition algorithm applied to planning the interdiction of stochastic networks,” *Naval Research Logistics*, vol. 52, pp. 321–328, 2005.

[37] H. Bayrak and M. D. Bailey, “Shortest path network interdiction with asymmetric information,” *Networks*, vol. 52, pp. 133–140, 2008.

[38] G. G. Brown, M. Carlyle, J. Salmerón, and R. K. Wood, “Defending critical infrastructure,” *Interfaces*, vol. 36, 2006.

[39] J. Salmerón, “Deception tactics for network interdiction: A multi-objective approach,” Operations Research Department, Naval Postgraduate School,, Monterey, CA, Tech. Rep., 2011.

[40] C. M. Rocco and J. E. Ramirez-Marquez, “A bi-objective approach for shortest-path network interdiction,” *Computers and Industrial Engineering*, vol. 59, pp. 232–240, 2010.

[41] N. Assimakopoulos, “A network interdiction model for hospital infection control,” *Computers in Biology and Medicine*, vol. 17, pp. 413–422, 1987.

[42] Y. Wang and S.-K. Au, “Spatial distribution of water supply reliability and critical links of water supply to crucial water consumers under an earthquake,” *Reliability Engineering and System Safety*, vol. 94, pp. 534–541, 2009.

[43] G. Brown, M. Carlyle, D. Diehl, J. Kline, and K. Wood, "A two-sided optimization for theater ballistic missile defense," *Operations Research*, vol. 53, pp. 745–763, 2005.

[44] J. Salmerón, K. Wood, and R. Baldick, "Analysis of electric grid security under terrorist threat," *IEEE Transactions on Power Systems*, vol. 19, pp. 905–912, 2004.

[45] ——, "Worst-case interdiction analysis of large-scale electric power grids," *IEEE Transactions on Power Systems*, vol. 24, pp. 96–104, 2009.

[46] J. P. Watson, R. Murray, and W. E. Hart, "Formulation and optimization of robust sensor placement problems for drinking water contamination warning systems," *Journal of Infrastructure Systems*, vol. 15, pp. 330–339, 2009.

[47] G. G. Brown, W. M. Carlyle, R. Harney, E. Skroch, and R. K. Wood, "Interdicting a nuclear-weapons project," *Operations Research*, vol. 57, pp. 866–877, 2009.

[48] J. F. Benders, "Partitioning procedures for solving mixed-variables programming problems," *Numerische Mathematik*, vol. 4, pp. 238–252, 1962.

[49] M. Scaparra and R. Church, "Protecting supply systems to mitigate potential disaster: A model to fortify capacitated facilities," University of Kent, Working Paper 209, 2010.

[50] P. Cappanera, "Maria paola scaparra," *Transportation Science*, vol. 45, pp. 64–80, 2011.

[51] F. Pan, "Stochastic network interdiction: Models and methods," Ph.D. dissertation, University of Texas at Austin, 2005.

[52] A. J. Miller and L. A. Wolsey, "Tight formulations for some simple mixed integer programs and convex objective integer programs," *Mathematical Programming*, vol. 98, pp. 73–88, 2003.

[53] G. L. Nemhauser and L. A. Wolsey, *Integer and Combinatorial Optimization*, 2nd ed. New York, NY: John Wiley & Sons, 1999.

[54] L. A. Wolsey, *Integer Programming*. New York, NY: Wiley-Interscience, 1998.

[55] H. D. Sherali, W. P. Adams, and P. J. Driscoll, "Exploiting special structures in constructing a hierarchy of relaxations for 0-1 mixed integer programs," *Operations Research*, vol. 46, pp. 396–405, 1998.

[56] H. D. Sherali and W. P. Adams, "A hierarchy of relaxations between the continuous and convex hull representations for zero-one programming problems," *SIAM Journal on Discrete Mathematics*, vol. 3, no. 3, pp. 411–430, 1990.

[57] ——, "A hierarchy of relaxations and convex hull characterizations for mixed-integer zero-one programming problems," *Discrete Applied Mathematics*, vol. 52, no. 1, pp. 83–106, 1994.

[58] E. Balas, S. Ceria, and G. Cornuéjols, “A lift-and-project cutting plane algorithm for mixed 0-1 programs,” *Mathematical Programming*, vol. 58, pp. 295–324, 1993.

[59] L. Lovász and A. Schrijver, “Cones of matrices and setfunctions and 0-1 optimization,” *SIAM Journal on Optimization*, vol. 1, pp. 166–190, 1991.

[60] M. R. Garey and D. S. Johnson, *Computers and intractability: a guide to the theory of NP-completeness*. W. H. Freeman, 1979.

[61] S. Neumayer, G. Zussman, R. Cohen, and E. Modiano, “Assessing the impact of geographically correlated network failures,” in *MILCOM, 2008 Proceedings IEEE*, 2008, pp. 1–6.

[62] ——, “Assessing the vulnerability of the fiber infrastructure to disasters,” in *INFOCOM, 2009 Proceedings IEEE*, 2009, pp. 1566–1574.

[63] S. Neumayer and E. Modiano, “Network reliability with geographically correlated failures,” in *INFOCOM, 2010 Proceedings IEEE*, 2010.

[64] P. Agarwal, A. Efrat, S. Ganjugunte, D. Hay, S. Sankararaman, and G. Zussman, “Network vulnerability to single, multiple, and probabilistic physical attacks,” in *MILCOM, 2010 Proceedings IEEE*, 2010, pp. 1824–1829.

[65] ——, “The resilience of WDM networks to probabilistic geographical failures,” in *INFOCOM, 2011 Proceedings IEEE*, 2011.

[66] A. Bernstein, D. Bienstock, D. Hay, M. Uzunoglu, and G. Zussman, “Power grid vulnerability to geographically correlated failures analysis and control implications,” Department of Electrical Engineering, Columbia University, Tech. Rep., 2011.

[67] W. Wu, B. Moran, J. Manton, and M. Zukerman, “Topology design of undersea cables considering survivability under major disasters,” in *WAINA, 2009 Proceedings IEEE*, 2009.

[68] P. Mirchandani and R. Francis, *Discrete Location Theory*. John Wiley & Sons, 1990.

BIOGRAPHICAL SKETCH

Kelly Sullivan was born in 1984 in Little Rock, Arkansas. He received B.S. and M.S. degrees in Industrial Engineering from the University of Arkansas in 2006 and 2008, respectively. He joined the Industrial and Systems Engineering department at the University of Florida in August 2008 and will complete his Ph.D. in August 2012. His research interests include integer programming, robust optimization, large-scale optimization, interdiction, and homeland security applications.