9-1-2009

Synthesis, Interdiction, and Protection of Layered Networks

Kevin T. Kennedy

SYNTHESIS, INTERDICTION, AND PROTECTION OF LAYERED NETWORKS

DISSERTATION

Kevin T. Kennedy
Major, USAF

AFIT/DS/ENS/09-01

DEPARTMENT OF THE AIR FORCE

AIR UNIVERSITY

AIR FORCE INSTITUTE OF TECHNOLOGY

Wright-Patterson Air Force Base, Ohio

The views expressed in this research are those of the author and do not reflect the official policy or position of the United States Air Force, Department of Defense, or the United States Government.

AFIT/DS/ENS/09-01

SYNTHESIS, INTERDICTION, AND PROTECTION OF LAYERED NETWORKS

DISSERTATION

Presented to the Faculty

Graduate School of Engineering and Management

Air Force Institute of Technology

Air University

Air Education and Training Command

In Partial Fulfillment of the Requirements for the

Degree of Doctor of Philosophy

Kevin T. Kennedy, B.S., M.S.

Major, USAF

September 2009

APPROVED FOR PUBLIC RELEASE; DISTRIBUTION UNLIMITED

AFIT/DS/ENS/09-01

Synthesis, Interdiction, and Protection of Layered Networks

Kevin T. Kennedy, B.S., M.S.
Major, USAF

Approved:

/signed/ 13 August 2009

Dr. Richard F. Deckro (Chairman) Date

/signed/ 13 August 2009

Dr. James T. Moore (Member) Date

/signed/ 13 August 2009

Dr. Kenneth M. Hopkinson (Member) Date

Accepted:

/signed/ 13 August 2009

M. U. THOMAS Date
Dean, Graduate School of Engineering and Management

AFIT/DS/ENS/09-01

Abstract

This research developed the foundation, theory, and framework for a set of analysis techniques to assist decision makers in analyzing questions regarding the synthesis, interdiction, and protection of infrastructure networks. This includes extension of traditional network interdiction to directly model nodal interdiction; new techniques to identify potential targets in social networks based on extensions of shortest path network interdiction; extension of traditional network interdiction to include layered network formulations; and develops models/techniques to design robust layered networks while considering trade-offs with cost.

These approaches identify the maximum protection/disruption possible across layered networks with limited resources, find the most robust layered network design possible given the budget limitations while ensuring that the demands are met, include traditional social network analysis, and incorporate new techniques to model the interdiction of nodes and edges throughout the formulations. In addition, the importance and effects of multiple optimal solutions for these (and similar) models is investigated. All the models developed are demonstrated on notional examples and were tested on a range of sample problem sets.

Acknowledgements

There are many people that I owe a great deal of thanks for their part in helping me successfully complete this program of study.

I would first like to thank my parents for stressing the value of education. Without their love and sacrifices, I would not be anywhere near where I am today. I would also like to thank my wife and four amazing kids. Their love and understanding provided the support I needed to make it through the difficult times during my research.

Of course, none of this would have been possible without my advisor Dr. Deckro. His hard work, vision, guidance, patience and experience were immeasurable. I have truly enjoyed working for Dr. Deckro and his ability to mentor and challenge his students forced me to become a better student, researcher, and person.

I am also grateful to my committee, Dr. James Moore and Dr. Kenneth Hopkinson. I thank them both for their patience, guidance and feedback on my dissertation.

There have also been numerous others who have helped along the way, that I would like to give my sincere thanks. Dr. Kevin Wood at the Naval Postgraduate School answered my numerous questions about network interdiction and GAMS coding. Dr. Jerry O'Neal also provided invaluable help with C++ programming, especially with the network generators.

Kevin T. Kennedy

Table of Contents

List of Figures

Figure Page

List of Tables

List of Abbreviations

Abbreviation	
AFIT	Air Force Institute of Technology
BLPP	bilevel programming problem
C4I	Command, Control, Communications, Computers, and Intelligence
CI	counter-intelligence
CNAT	Critical Network Analysis Tool
COAL	Committee on Algorithms
COG	center of gravity
COIN-OR	COmputational INfrastructure for Operations Research
CONUS	continental United States
DHS	Department of Homeland Security
DIMACS	Center for Discrete Mathematics and Theoretical Computer Science
DoD	Department of Defense
DTIC	Defense Technical Information Center
FOUO	For Official Use Only
GAMS	General Algebraic Modeling System
GIS	Geographic Information System
GUI	Graphical User Interface
HIIM	Holistic Interpersonal Influence Measure
HSIP	Homeland Security Infrastructure Program
ISR	Intelligence, Surveillance, and Reconnaissance
JFC	Joint Forces Commander
JUNG	Java Universal Network/Graph
LNM	Layered Network Model
KKT	Karush-Kuhn-Tucker
LP	linear program
MANA	Map Aware Non-Uniform Automata
MILDEC	military deception
MIP	Mixed Integer Program
MORS	Military Operations Research Society
NART	Network Analysis Research Tool
NCW	Network Centric Warfare
NGA	National Geospatial-Intelligence Agency
NIPP	National Infrastructure Protection Plan
OA	operational assessment
OPSEC	operations security
OR	operations research
PA	public affairs
PDD	Presidential Decision Directive
PSYOP	psychological operations
SCADA	Supervisory Control and Data Acquisition
SNA	social network analysis
SoSA	system of systems analysis
SVI	supervalid inequality
US	United States
VBA	Visual Basic for Applications
VFT	Value Focused Thinking

Synthesis, Interdiction, and Protection of Layered Networks

I. Introduction

1.1 Background

Network models provide a foundation for analysis in many diverse areas. For example, they are used to study and analyze graphs, supply chains, national infrastructures, social interactions and organizations, to name a few areas. Accordingly, these networks are studied by a variety of interdisciplinary fields including, but not limited to, mathematics, sociology, and operations research. Graph theory, a subset of mathematics, and sociology have developed descriptive measures and techniques to analyze and understand the structure of networks.

Operations researchers have also developed descriptive measures, as well as extending techniques to prescriptive methods to optimize and forecast network performance. As a result, operations research techniques can be used to make networks highly efficient and cost effective. Unfortunately, using only standard synthesis techniques (such as minimum cost synthesis/flow) at the exclusion of others (more robust objective functions) have made many networks increasingly vulnerable to disruptions. [113, p. 235] In addition, networks are often modeled as individual, self-contained units with little regard for interdependencies. However, real world networks are often complex systems of interconnected networks with competing objectives and competitively shared resources. While nature may appear to create random attacks due to weather, earthquakes, and so forth; as the events of September 11th demonstrate, networks are subject to attack from intelligent adversaries who seek to maximize damage. This damage may not be isolated to the specific network attacked.

A great deal of the efforts to identify vulnerabilities in interdependent/layered networks comes from system engineering. This discipline has developed conceptual tools to begin to understand the consequences and effects of actions applied to these layered networks.

1.1.1 System Perspective

> The systems engineering method recognizes each system as an integrated whole even though composed of diverse, specialized structures and subfunctions. It further recognizes that any system has a number of objectives and that the balance between to optimize the overall system functions according to the weighted objectives and to achieve maximum compatibility of its parts. [37, p. 3]

A systems engineering approach allows a broad qualitative understanding of the operational environment in which decisions are made. It promotes "a holistic view of the operational environment that focuses on those key nodes that could influence the outcomes of an operation." [35, p. II-4] A systems view of a military operational environment contains all the elements which are relevant to the current operation under consideration.

The Joint Warfighting Center described the environment in which the military operates as a system of layered networks. [35, p. II-2] Specifically, "system nodes are the tangible elements within a system that can be 'targeted' for action ..." Links are "the behavioral or functional relationships between nodes." [35, p. II-3] For example, nodes may include people, material, or facilities, while links "establish the interconnectivity between nodes that allows them to function as a system." [35, p. II-3] Figure 1.1 is a notional example of the operational environment developed with systems engineering techniques.

With this framework, military decision makers decide on a course of action which aims to "destroy, interrupt, or otherwise affect the relationship" between the nodes. Originally these decisions considered actions in isolation, but decision makers

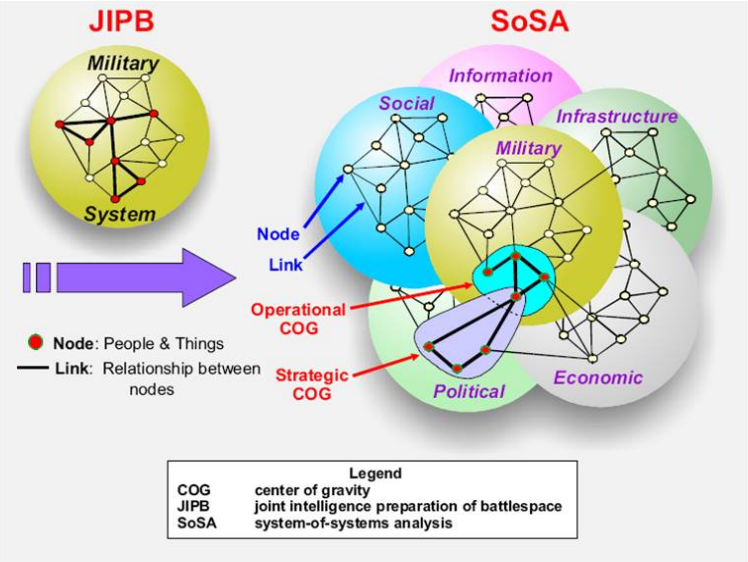


Figure 1.1: Systems Perspective of the Operational Environment [35, p. II-2]

now realize these actions "ultimately influence the system as a whole." [35, p. II-3] This perspective forces decision makers to consider the effects of their decisions across all impacted layers.

System engineering provides a foundation/framework in which to make decisions. In many cases, a quantitative approach may also be beneficial. For example, if quantitative information is available for the networks, then tools from graph theory and operations research can be extended to the system engineering context to provide a rigorous theoretical and mathematical framework from which to make decisions. This framework would need to incorporate the multiobjective (whether competing, symbiotic, integrated, *etc.*) nature of these networks to allow analysis to be done on these layered networks.

1.2 Research Outline

Researchers have demonstrated that the environment in which the military operates can be viewed as layered networks. Some steps to extend existing traditional network models to layered networks have been taken by Wallace [122] and Kennedy

[72]; however, work remained. As a result, the following additional advancements have been made and are described.

To understand the vulnerabilities of layered networks, network interdiction techniques were extended to multiple layers. This identifies the components whose destruction, disruption, or influence would have the greatest intended impact on the layered networks. With these vulnerabilities identified, steps can be taken to minimize the impact of potential attacks across *all* layers of interdependent networks. Therefore, analysis techniques were further extended to identify either additions to the network or components whose fortification would reduce these vulnerabilities. Of course, in addition to analysis of vulnerabilities and protection of existing networks, these techniques can also be used to develop/design new networks.

Simultaneous network interdiction and protection relies on multilevel programming. Solutions from multilevel programs are generally not Pareto optimal. [10, p. 304] This, combined with the fact that network programs often have multiple optimal solutions, leads to circumstances where coalitions may form in order to improve individual solutions. This potential for cooperation is studied in the context of multilevel programming to determine its impact on layered network vulnerability and protection.

1.2.1 Research Objectives

This research has the following contributions:

Formulation/Theory

- Extends network interdiction to direct nodal interdiction; specifically maximum flow nodal interdiction.
- Extends social network analysis methods to determine optimal human interdiction determination.
  - Individual closeness interdiction.
  - All-pairs shortest path interdiction for group closeness interdiction.

- Extends network interdiction to layered networks. This identified the maximum protection/disruption of layered networks with limited resources.
- Extends network synthesis techniques to layered networks considering both cost and risks. This approach allows decision makers to balance robustness, cost, and risk in designing or expanding networks (and in the process identifying edges which make the network vulnerable).
- Determines the impact of coalitions and multiple optimal follower solutions on vulnerabilities and protection decisions.

Methodology/Application

- Provided a methodology to solve above formulations.
- Implemented in General Algebraic Modeling System (GAMS).
- Tested against a developed/notional test problem set.

1.2.2 General Assumptions and Scope

The methods developed in this research are deterministic. Therefore, it is assumed all necessary data is available and not time dependent. For many networks, such as infrastructure networks, this is a reasonable assumption. These networks tend to be fixed assets whose locations are publicly known. In addition, in analyzing one's own network to determine vulnerabilities, it is reasonable to assume all network information is available (i.e. one is not hiding networked components from oneself).

In order to solve multilevel programs, several assumptions are typically made. First, the decision makers on all levels must be rational and can not cooperate with each other. In addition, it is generally assumed (whether true or not), that all lower level programs have unique solutions at all possible values of upper level decisions.

Unfortunately, data for networks of interest may be unavailable and/or restricted. Therefore, notional examples are used for demonstration purposes in unrestricted portions of this research. The methods developed, however, apply to any given network. Therefore, those with access to "real" data can use these methods. For example, infrastructure data for several locations has been obtained by this author, but is generally restricted to "for official use only."

It is also understood that some networks under consideration (especially social networks) change over time. However, this study considered these networks as a snapshot in time (i.e. at the time the decision will be made). Dynamic network research is an evolving field, especially in social network analysis. As measures and techniques are developed to analyze evolving networks, the methods developed in this research can be extended. In other words, the methods developed here serves as a foundation of static networks that can be extended to dynamic networks.

1.3 Document Overview

This document is organized as illustrated in Figure 1.2. Chapter 2 provides the foundational material from the literature review (which is colored green in the Figure 1.2). This includes multiobjective programming, network optimization, and multilevel programming. Two areas which build on this foundation are also provided in the literature: multilevel network optimization which builds on single network optimization, and network interdiction which combines portions of network optimization and multilevel programming. With this foundation, the following chapters develop new formulations with the end goal of contributing to a unified formulation for the analysis of synthesis, interdiction, and protection of layered networks. One of these developments is a new formulation for the synthesis of robust networks. This includes two blocks from Figure 1.2: multiobjective network optimization is built on multiobjective optimization and network optimization; and multiobjective multilayer network optimization which is built on multiobjective network optimization and multilayer network optimization.

Another formulation development is multilayer interdiction which is based on multilayer network optimization and network interdiction. This is followed by a new formulation for human network interdiction which is based on combining network interdiction with traditional social network analysis metrics. Two additional developments are nodal interdiction and coalition/multiple optimal solution formulations.

These are extensions to traditional network interdiction and build on multilevel optimization.

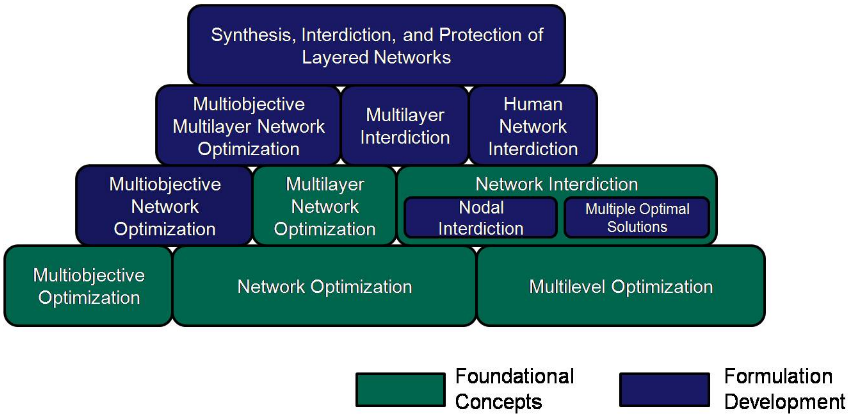

Synthesis, Interdiction, and Protection of Layered Networks
Multiobjective Multilayer Network Optimization
Multilayer Interdiction
Human Network Interdiction
Multiobjective Network Optimization
Multilayer Network Optimization
Network Interdiction
Nodal Interdiction
Multiple Optimal Solutions
Multiobjective Optimization
Network Optimization
Multilevel Optimization
🟩 Foundational Concepts
🟦 Formulation Development

Figure 1.2: Research Blueprint

Figure 1.2 presents a blueprint which is followed in this document. Along the way, these developments are demonstrated using illustrative examples. This research concludes by showing how these techniques could be used to model layered infrastructure networks for synthesis, vulnerabilities, and protection.

II. Literature Review

This chapter reviews relevant literature in network research including formulation and solution techniques. This provides the foundation for extensions which will be developed in the next chapters. This includes multiobjective programming, network optimization, and multilevel programming. Two areas which build on this foundation are also developed in the literature: multilevel network optimization which builds on single network optimization, and network interdiction which combines portions of network optimization and multilevel programming. As demonstrated in Figure 2.1, this literature review provides the foundation for formulation developments discussed in the next chapter.

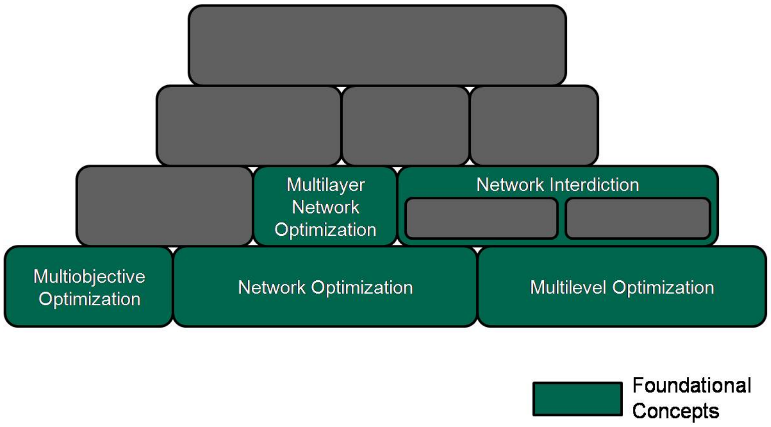


Figure 2.1: Literature Review Roadmap

2.1 Multiobjective Programming

In this section, a portion of multiobjective programming is discussed. Specifically, techniques which focus on formulations to analyze the relationship between robustness (maintaining near optimal solutions despite disruptions) and costs ar explored. As shown in Figure 2.1, this block provides the foundation for multiobjective network optimization.

Ehrgott and Ryan [53] developed a formulation/technique which finds all non-dominated solutions to the tradeoff between cost and robustness. Specifically, they sought to minimize the cost of crew schedules while maximizing robustness. In their formulation, instead of maximizing robustness, they minimized the potential delays resulting from a lack of robustness. Their formulation of the bicriteria problem is as follows: [53, p. 142]

$$\begin{aligned} \text{Min } & z_c = c^T x \\ \text{Min } & z_r = r^T x \\ \text{s.t. } & A_1 x = e \\ & A_2 x = b \\ & x \in \{0, 1\} \end{aligned} \tag{2.1}$$

where $e = (1, 1, \ldots, 1)^T$. The first objective function minimizes cost, the second objective function minimizes effects of lack of robustness, and the constraints define the feasible region. Specifically, the first set of constraints ensure each flight had exactly 1 crew assigned to it, and the second set of constraints are the model's base constraints. [53, p. 142]

The authors first discuss solving this formulation with the popular weighted secularization method. However, they point out that since this problem is a discrete optimization problem, it is well known that a class of efficient solutions known as "unsupported solutions" can never be found. [53, p. 142] For example, consider the following formulation:

$$\begin{aligned} \text{Min } & z_{cr} = \theta c^T x + (1 - \theta) r^T x \\ \text{s.t. } & A_1 x = e \\ & A_2 x = b \\ & x \in \{0, 1\} \end{aligned} \tag{2.2}$$

where θ is a parameter, limited to the range $0 \leq \theta \leq 1$, which is varied to find efficient solutions. All solutions of this formulation will be efficient. [52, p. 97]

However, since this is a discrete formulation (due to the binary variable), there may exist some solutions that this formulation would be unable to find. [52, p. 98]

Therefore, this bicriteria problem was converted into both an ε-constraint problem and an elastic constraint problem. [53, p. 142,144] The concept underlying the ε-constraint is to keep one objective function, and transform all other objective functions into constraints. The ε value places an upper bound on these newly formed constraints. For example, the transformed formulation of Ehrgott and Ryan's formulation in (2.1) becomes [53, p. 142]

$$\begin{aligned} \text{Min } & z_r = r^T x \\ \text{s.t. } & c^T x \leq \varepsilon \\ & A_1 x = e \\ & A_2 x = b \\ & x \in \{0, 1\} \end{aligned} \tag{2.3}$$

where the cost objective function is now a constraint bounded by ε. The ε-constraint method was developed and proven to be able to find all efficient solutions (even for non-convex problems) to models of the form as (2.3). [36, p. 122-123]

The advantages of this formulation are clear:

> Besides being able to generate all efficient solutions by varying the upper bounds on the objective constraints, ... management could simply specify the additional cost they are willing to concede in order to improve robustness. [53, p. 142]

In other words, the effect of various budget levels can be determined simply by varying the ε value accordingly.

Similarly, the research in this dissertation develops a bicriteria model for layered infrastructures. The first objective minimizes costs, while the second objective function maximizes robustness. As done with the Ehrgott and Ryan example, the cost objective function is moved to the constraints using the ε-constraint method.

Therefore, this model is automatically parameterized to consider varying budget levels against robustness.

However, as Ehrgott and Ryan demonstrate, such problems are generally "unsuitable from a computational point of view." [53, p. 144] To overcome this computational burden, they convert their ε-constraint problem into an elastic constraint problem. [53, p. 144] The elastic constraint method relaxes these difficult constraints by allowing them to be violated and penalizing any violation in the objective function. An example of this is demonstrated as follows:

$$\begin{aligned} \text{Min } & z_r = r^T x + p s_u \\ \text{s.t. } & A_1 x = e \\ & A_2 x = b \\ & c^T x + s_l - s_u = \varepsilon \\ & x \in \{0, 1\} \end{aligned} \tag{2.4}$$

As shown by the transformed formulation, the cost constraint is converted to an equality constraint with slack variables added. Any slack, s_u, is then penalized in the objective function. Ehrgott and Ryan proved that this formulation will also find all efficient solutions (by varying p), but is much less computationally burdensome. Again, following their example, the formulation of the bicriteria layered infrastructure problem was transformed into an elastic constraint problem.

2.2 Networks

Networks are studied in a variety of fields. For example, mathematicians developed graph theory, in part, to understand the *structure* of networks. Sociologists have borrowed many of these techniques to describe and explore social networks. [124] In addition, the field of operations research has incorporated many of these techniques and developed the foundation for others. In other words, mathematicians study graphs (which is a mathematical construct consisting of vertices and edges); whereas, operations researchers study networks (which are graphs *plus some data*). [4, p. 4]

Network flow modeling is a subset of mathematical programming that exploits special structures within the problem which allow larger problems to be solved in less time than more general mathematical programming would typically require. Since many real world systems can be modeled as networks, a large number of techniques exist for analyzing single layer networks.

As shown in Figure 2.1, network optimization provides the foundation for three direct formulation developments: multiobjective network optimization, multilayer network optimization, and network interdiction.

To provide this broad foundation, a variety of network models are discussed. Specifically, shortest path, maximum flow, minimum cost network flow, minimum cost cut-sets, network centric operations, social networks, network design, and network interdiction are discussed. This review begins with the classic single layer models from Ahuja *et al.* [2]

2.2.1 Notation

We begin with a directed graph $G = (N, A)$ defined by a set N of n nodes and a set A of m directed arcs. Each arc $(i, j) \in A$ has an associated cost per unit flow c_{ij}. In addition, each arc (i, j) has a maximum amount that can flow through it called its capacity u_{ij} and a lower bound on the minimum amount that must flow l_{ij}. Finally, each node has an associated integer $b(i)$; if $b(i) > 0$ then the node is a supply node, if $b(i) < 0$ then the node is a demand node, and if $b(i) = 0$ then the node is a transshipment node. [2, p. 5]

2.2.2 Shortest Path

Shortest path calculations are used both directly and in the calculations of more complex network flow techniques and social network measures. The idea is simple: suppose we wish to find the shortest path from s to t. Ahuja *et al.* [2, p. 94] provides a linear programming formulation to find the shortest path between any two nodes. Let $c_{i,j}$ be the length of arc i, j, and let $x_{i,j}$ be 1 if arc i, j is chosen for the path, and zero otherwise. In addition, A is an adjacency matrix where each entry a_{ij} is one if there exists an edge from vertex i to vertex j, and zero otherwise.

$$\begin{aligned}
&\min \sum_{(ij)\in A} c_{ij}x_{ij} \\
&\sum_{j:(ij)\in A} x_{ij} - \sum_{j:(ji)\in A} x_{ji} = \begin{cases} 1, & \text{for } i = s; \\ 0, & 0\ \forall i \in N\backslash\{s,t\}; \\ -1, & \text{for } i = t. \end{cases} \\
&x_{ij} \geq 0 \quad \forall (i,j) \in A
\end{aligned} \tag{2.5}$$

Numerous specialized techniques have been developed to exploit the structure of the network in solving the shortest path problem. For example, Dijkstra's algorithm solves the shortest path problem in $O(n^2)$ time. [2, p. 111]

2.2.3 Maximum Flow

The maximum flow problem is stated as follows: "In a capacitated network, we wish to send as much flow as possible between two special nodes, a source node s and a sink node t, without exceeding the capacity of any arc." [2, p. 166]

$$\begin{aligned}
&\text{Maximize } \nu \\
&\text{s.t.} \sum_{\{j:(i,j)\in A\}} x_{ij} - \sum_{\{j:(j,i)\in A\}} x_{ji} = \begin{cases} \nu, & \text{for } i = s; \\ 0, & \text{for all } i \in N - \{s \text{ and } t\}; \\ -\nu, & \text{for } i = t. \end{cases} \\
&l_{ij} \leq x_{ij} \leq u_{ij} \quad \forall (i,j) \in A
\end{aligned} \tag{2.6}$$

where ν is the maximum flow.

There are two basic types of algorithms generally used to solve maximum flow problems: augmenting path algorithms that incrementally augment flow along paths from the source to the sink; and preflow-push algorithms that flood the network and incrementally relieve flow from nodes with excess by sending them forward toward the sink or backward toward the source. [2, p. 167]

2.2.4 Minimum Cost Network Flow

The minimum cost flow problem is simply stated as: "We wish to determine a least cost shipment of a commodity through a network in order to satisfy demands at certain nodes from available supplies at other nodes." [2, p. 4] Mathematically, this is modeled as follows:

$$\begin{aligned} \min & \sum_{(i,j)\in A} c_{ij}x_{ij} \\ \text{s.t.} & \sum_{\{j:(i,j)\in A\}} x_{ij} - \sum_{\{j:(j,i)\in A\}} x_{ji} = b(i) \quad \forall i \in N \\ & l_{ij} \leq x_{ij} \leq u_{ij} \quad \forall (i,j) \in A \end{aligned} \tag{2.7}$$

where, as before, $b(i)$ is 0 for transshipment nodes, greater than zero for supply nodes, and less than zero for demand nodes.

The model in (2.7) can also be written in matrix form as follows:

$$\begin{aligned} \text{Minimize } & \mathbf{cx} \\ \text{s.t. } & \mathcal{N}\mathbf{x} = \mathbf{b} \\ & \mathbf{l} \leq \mathbf{x} \leq \mathbf{u} \end{aligned} \tag{2.8}$$

where $\mathcal{N}$ is an $n \times m$ matrix, called the node-arc incidence matrix of the minimum cost flow variable. Each column $\mathcal{N}_{ij}$ represents the variable x_{ij} with the value $+1$ in the ith row and the value -1 in the jth row. [2, p. 5]

Algorithms for solving the minimum cost network flow problem usually "combine ingredients of both shortest path and maximum flow algorithms," again demonstrating the importance of these formulations and solution techniques. [2, p. 295]

2.2.5 Minimum Cost Cut-set

A network cut is a set of arcs whose deletions disconnects the network into two separate components. A minimum $s-t$ cut is the network cut that has the minimum capacity and disconnects the source from the sink. [2, p. 167] Consider π_i, the dual variable associated with the conservation of flow equation for node i. In addition, let ν_{ij} be the dual variable associated with the capacity constraint of arc (i,j).

$$
\begin{aligned}
\min \sum_{(i,j)\in A} & c_{ij}\nu_{ij} \\
s.t. \pi_i - \pi_j - \nu_{ij} \geq 0 \ \forall & \ \ (i,j) \in A \\
0 \leq \pi_i \leq 1 \ \forall i & \quad \in N \\
0 \leq \nu_{ij} \leq 1 \forall (i,j) & \quad \in A \\
\pi_t = 1; \pi_s = 0 &
\end{aligned}
\tag{2.9}
$$

where c_{ij} is the flow capacity along arc (i,j). It follows that the objective function, $\sum c_{ij}\nu_{ij}$, is the relative cost of cutting the flow of goods.

There is a direct correspondence between maximum flow problems and minimum cut problems. This is stated in the Max-Flow Min-Cut Theorem: "the maximum value of the flow from a source node s to a sink node t in a capacitated network equals the minimum capacity among all $s-t$ cuts." [2, p. 185] Therefore, the theory and algorithms developed for maximum flow problems are also applicable to minimum $s-t$ cut problems.

2.2.6 Network Centric Operations

Networks are increasingly being analyzed in military warfare. Network Centric Warfare is the combination of "strategies, emerging tactics, techniques, and procedures, and organizations that a fully or even partially networked force can employ to create a decisive warfighting advantage." [34, p. 3] In other words, network centric warfare will allow the United States (US) to use its dominance in technology and information to develop a warfighting advantage by "information sharing, shared situational awareness, and knowledge of commander's intent."[34, p. 4]

To provide a foundation of this theory of warfare, the Office of Transformation published four tenets, nine governing principles, and a conceptual framework.

2.2.6.1 Tenets

The four basic tenets of network centric warfare are the following: [34, p. 7]

- A robustly networked force improves information sharing.
- Information sharing enhances the quality of information and shared situational awareness.
- Shared situational awareness enables collaboration and self-synchronization, and enhances sustainability and speed of command.
- These, in turn, increase mission effectiveness.

These tenets were developed to help understand the enhanced power of networked forces. [34, p. 7] However, as discussed later, network centric warfare (including these tenets) implies a symmetric opponent in a conventional conflict. Although the tenets may still be valid against an asymmetric opponent, they focus on the technology to speed the spread of situational awareness, not on how to gather situational awareness from a local population. In addition, there are potential asymmetric advantages to an attacking enemy who is not dependent on these systems. For example, while some enemies use paper maps, the Air Force typically uses FaconView to plan missions. If this system is disabled, it could potentially cause disruptions to planning cycles.

2.2.6.2 Principles

To augment traditional principles of war, the governing principles of a network centric force have also been developed: [34, p. 8]

- Fight first for information superiority
- Access to information: shared awareness
- Speed of command and decision making
- Self-synchronization, dispersed forces: non-contiguous operations
- Demassification
- Deep sensor reach
- Alter initial conditions at higher rates of change
- Compressed operations and levels of war

It is noted that these principles were not designed to replace the "principles of war," but to "provide added direction for executing military operations in the Information Age." [34, p. 8]

2.2.6.3 Conceptual Framework

To provide a foundation for understanding network centric warfare and how its elements fit together, the Office of Transformation developed the conceptual framework depicted in Figure 2.2.[61, p. 4] This framework is a "top-level" representation of network centric concepts and their relations. For example, the foundation of each area is color coded according to its associated domain: physical, information, cognitive, or social.

In addition, with this broad representation, the concept can be decomposed so individual attributes and metrics can be identified for each concept. [61, p. 4] "Each concept in the top-level is described by a set of attributes and metrics at the second level."[61, p. 5] These metrics (combined with other metrics and examples from other sources) allows one to measure the impact of network centric warfare systems, or lack thereof.

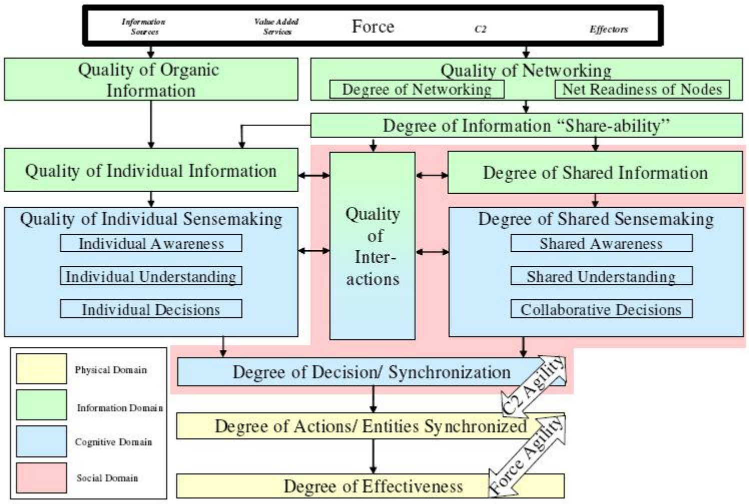


Figure 2.2: Top Level Conceptual Framework

2.2.6.4 Measures

As stated previously, the conceptual framework provides a broad picture of network centric warfare. In addition, the Office of Transformation decomposed this framework and provided second tier metrics. These metrics, along with others, can be found in [34].

Wong-Jiru *et al.* extended these metrics using a multi-layered model. This model breaks Network Centric Warfare (NCW) into processes, people, applications, systems, and the physical network. Each of these are modeled as individual layers, and interactions between layers are modeled as well. [136, p. 2] The (inter-) layer metrics were borrowed from social network analysis (SNA) theory: shortest distance, maximum flow, point connectivity, in/out degree centrality, closeness centrality, flow betweenness centrality, reachability, density, node betweeness centrality, and edge betweenness. [136, p. 40] For each layer, these metrics were plotted on a composite radar chart. A composite layer score was calculated by finding the area under this curve. Finally, a network centrality score was calculated by summing these across all the layers. With these measures, a comparison could be made before and after an (potential) action is taken.

2.2.7 Social Networks

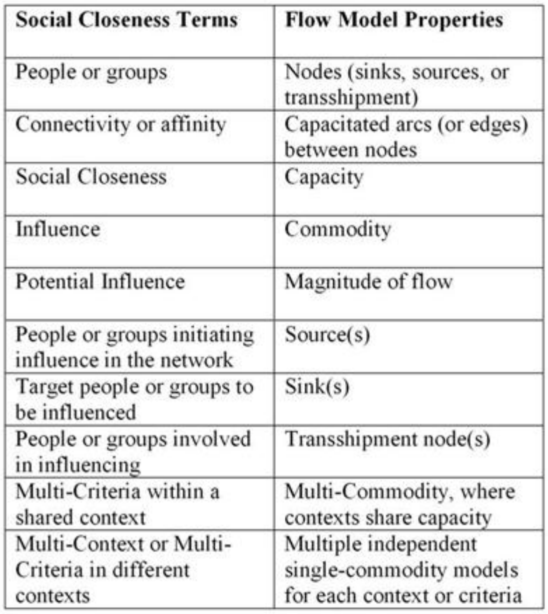
Although often ignored in analyzing other networks, social networks usually play a key role in the operation and development of other types of networks. For example, in infrastructure networks, humans interact with these networks as managers, operators, and users. While humans play an important role in the efficiency and security of infrastructure networks, they are also "the most susceptible to failure and the most adaptable in the management of recovery." [5, p. 73] Therefore, "we will not be able to attack the technical and human portions of the network separately ... destroying terrorists networks requires combinations of physical and social approaches." [116, p. 2-3] The effectiveness and response of a network during and after an attack are determined by humans. For example, "the effectiveness of attacking a power grid may depend on how the operators respond to limit the damage or redirect power." [116, p. 15]

To study social networks, sociologists have borrowed many techniques from graph theory. However, as Clark points out, "SNA measures were designed to help *describe* the network and its topology ..." and while "SNA measures fall short on prescriptive results, many Operations Research techniques were designed with actionable results in mind." To help demonstrate how operations research techniques could be used to make predictions on social networks, Clark developed the "Holistic Interpersonal Influence Measure (HIIM)" (discussed in Section 2.2.7.4). [41, p. 1-8] Renfro developed a measure of flow that he defined as "social closeness" which he proved satisfied all the requirements for flow modeling in linear programming. [99, p. 89] This measure is a capacity bound on potential influence. [99, p. 92] Nesbitt showed that given a measure of flow (in a social network), network interdiction techniques could be used to determine the optimal members to "interdict." [90] Similar work has also been done by Hamill [66], Renfro [99], Clark [41], and Herbranson [68].

In addition, research has been done to demonstrate how social networks can be modeled and analyzed as traditional network flows. For example, in a minimum cost network flow context, cost can take several meanings.

> For instance, 1) a low cost can imply a relationship between actors who have a high level of trust in each other with regard to network operations, 2) a low cost can indicate that analysts have a high level of confidence in the data collected about that relationship, 3) a low cost can imply a relationship that has a low risk of exposure, or 4) costs can represent the monetary expense incurred in commodity exchanges between individuals. In addition, commodities include not only intuitive examples like funds or equipment, but may also include goods with less tangible values such as training or information. [90, p. 8]

A mapping from other social closeness terms to network flow was developed by Renfro and depicted in Figure 2.3. [99, p. 95] or [100]



Social Closeness Terms	**Flow Model Properties**
People or groups	Nodes (sinks, sources, or transshipment)
Connectivity or affinity	Capacitated arcs (or edges) between nodes
Social Closeness	Capacity
Influence	Commodity
Potential Influence	Magnitude of flow
People or groups initiating influence in the network	Source(s)
Target people or groups to be influenced	Sink(s)
People or groups involved in influencing	Transshipment node(s)
Multi-Criteria within a shared context	Multi-Commodity, where contexts share capacity
Multi-Context or Multi-Criteria in different contexts	Multiple independent single-commodity models for each context or criteria

Figure 2.3: Mapping from Social Networks to Network Flows [99, p. 95]

In this research, it is assumed that all actors and links of a social network are known. Therefore, the theory and techniques developed here identify the optimal actor/relationship to target *given the information available.* Of course, this is the underlying assumption of most SNA measures, as they were not, in general, developed for use with missing information in mind. Unfortunately, this is often unrealistic for clandestine networks which need to be secretive in order to minimize detection and survive. In addition, Carley noted that “any isolation is better than none, assuming our goal is to degrade the performance and that we don’t need perfect information to be quite effective.” [32, p. 10]

Incompleteness of the data is not likely to be random. It may reflect the security discipline of the group being targeted or it may reflect biases in data-collection as a result of “lead-following” investigation techniques. [115, p. 262] Therefore, the determination of centrality will depend on “who you know most about, rather than who is central or pivotal in any structural sense.” [115, p. 256] In addition, Sterling found that even a small amount of missing information (less than 10%) can decrease the confidence in some SNA measures (subgroup detection in her case). [117, p. 146]

For analysis of covert networks, Borgatti *et al.* showed that centrality measures are robust to missing information in random graphs. [24] In random networks, he showed that errors in centrality measures increased linearly with the amount of missing information. However, Borgatti noted the degradation in estimation appears faster for cellular networks (as opposed to random networks) and may not be linear. This is a critical issue: in random networks all destabilization tactics (such as isolation of the individual that is highest in centrality) have approximately the same effect; but for networks arranged into cells, this may not be true. Further study is required to determine the impact of missing information on cellular structured networks. In addition, as stated previously, missing data is not likely to be random.

Although complete data is assumed, the techniques developed in this research are deterministic and enable post optimality analysis. This allows an examination of the impact of incomplete or incorrect data.

An important, but little studied (in open sources), topic is the disruption and/or protection of social networks. For hierarchical networks, it has been suggested that this analysis is relatively easy: since groups cannot continue operations without leadership (or often operate less effectively as autonomous units), an effective strategy is to target the leadership. [68, p. 2-23] However, leadership can be replaced and if a mechanism for succession exists, the disruption may be minimal, depending on the closeness and dependencies of the group. Carley *et al.* suggest destabilization occurs when resources, communications, and workload are impacted. [32, p. 4] Geffre develops a criticality measure which combines location, skill/resource connections, and social connections to identify members to target. [62, p. 3-1]

In addition, several traditional SNA measures have been developed to help determine who is important in a social network; some of these are potentially useful in developing plans to disrupt a social network.

> In seeking to incapacitate criminal organizations one obvious approach is to identify those players who are somehow central, vital, key, or pivotal, and target them for removal or surveillance. A central member may play a key role in a network by acting as a leader or serving as a gatekeeper ensuring information flow. [115, p. 264]

Therefore, centrality is “an important ingredient in considering the identification of network vulnerabilities.” [115, p. 264] Specifically, several centrality measures have been considered, including: degree centrality, closeness centrality, and betweenness centrality. These are discussed, along with Borgatti’s “key player” metric. The idea behind the “key player” concept is to break the group into smaller fragmented, less effective groups.

2.2.7.1 Degree Centrality

Degree centrality, $C_D(n_i)$, measures the number of direct connections a node has to other nodes. It has been used as a proxy measure of influence under the assumption that the most connected individual has the most influence. It is calculated as follows: [124, p. 178]

$$C_D(n_i) = \sum_j x_{ij} = \sum_j x_{ji} \tag{2.10}$$

2.2.7.2 Closeness Centrality

One measure of how "key" a member is in a social network is closeness centrality, $C(n_i)$. Closeness centrality is defined in terms of the distance from an individual to all other nodes. For example, a node who "has the shortest possible paths to all the other actors . . . has maximum closeness. [124, p. 184] Therefore, closeness centrality gives a measure of a person's proximity, either virtually or physically, to communicate and/or reach other members of the network.

Mathematically, closeness centrality is calculated as the inverse of the sum of the shortest paths to all other nodes in the network.

$$C(n_i) = \left[\sum_{j=1}^{g} d(n_i, n_j) \right]^{-1} \tag{2.11}$$

where n_i is the node for which centrality is being calculated, $d(n_i, n_j)$ is the distance from node i to node j, and g is the total number of nodes. [57] This measure sums the length of the shortest path from a node to all other nodes and takes the inverse. If no weighting is given to each relation, the distance between a pair of connected nodes is assumed to be one. Unfortunately, closeness centrality becomes "quite arbitrary if the network has arbitrary or fuzzy boundaries." [115, p. 265] It also requires finding all shortest paths from all nodes.

2.2.7.3 Betweenness Centrality

Another suggested measure of how "key" a person is in a network is betweenness centrality, $C_B(v)$. Betweenness centrality is a measure of the proportion of times a node is on the shortest path between other pairs of nodes. Therefore, it is a measure for information control, and/or a person's role as an intermediary such as a broker or gatekeeper. [56]

Freeman [56] developed the following definition:

$$C_B(v) = \sum_{s \neq v \neq t \in V} \frac{\sigma_{st}(v)}{\sigma_{st}} \tag{2.12}$$

where σ_{st} is the number of shortest paths from s to t, and $\sigma_{st}(v)$ is the number of shortest paths from s to t that pass though a node v. "Removing a node of high 'betweenness' will by definition, lengthen the paths connecting several other nodes, rendering communication or transactions between them less efficient." [115, p. 264] This assumes the next shortest path is indeed longer. This assumption, in turn, depends on the network density and arc weighting.

2.2.7.4 Holistic Interpersonal Influence Measure

Clark developed HIIM as a proxy measure of interpersonal influence based on both personal characteristics and social structural characteristics. [41, p. 3-1] Figure 2.4 provides an outline of how this measure is calculated. An analyst begins with demographic and social network data. Individual characteristics and SNA centrality measures are input into a discriminant function where the post posterior probabilities serve as a proxy of individual influence. Interpersonal influence is calculated based on the network topology. The individual measure and interpersonal influence measures are combined to produce the measure of interpersonal influence, HIIM.

2.2.7.5 Key Player Problem 1

Traditional centrality measures help determine structural properties of an open network and a person's role in it. However, these measures do not identify those whose removal would result in a residual network with less cohesion. To show this, Borgatti begins with the seemingly well-suited measure of betweenness centrality.

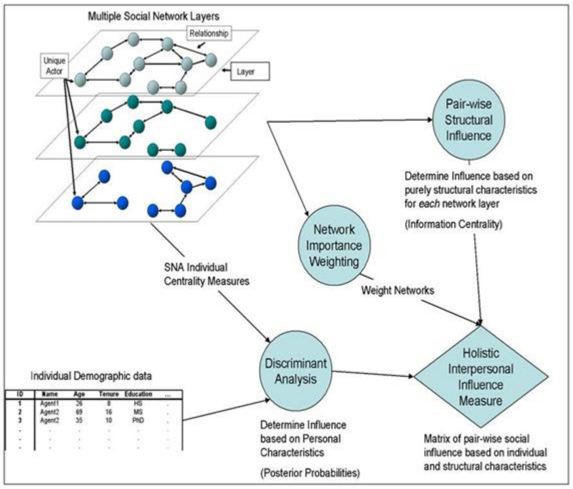

Multiple Social Network Layers
Relationship
Unique Actor
Layer
Pair-wise Structural Influence
Determine Influence based on purely structural characteristics for each network layer
(Information Centrality)
Network Importance Weighting
SNA Individual Centrality Measures
Weight Networks
Individual Demographic data
ID Name Age Tenure Education ...
1 Agent1 26 8 HS
2 Agent2 69 16 MS
3 Agent2 35 10 PhD
Discriminant Analysis
Holistic Interpersonal Influence Measure
Determine Influence based on Personal Characteristics
(Posterior Probabilities)
Matrix of pair-wise social influence based on individual and structural characteristics

Figure 2.4: Holistic Interpersonal Influence Measure Methodology [41, p. 3-3]

However, he shows that removing nodes with the highest betweenness centrality does not (necessarily) maximally fragment the network and does not measure the size of any remaining components that do occur. [23] Therefore, Borgatti develops a measure of degree of "fragmentation;"[1] expressed mathematically as [23, p. 28]

$$D_F = 1 - \frac{2\sum\limits_{i>j} \frac{1}{d_{ij}}}{n(n-1)} \tag{2.13}$$

where d_{ij} is the distance from node i to node j. Note that the reciprocal of the distance must be taken to account for components that are not connected. In other words, if two actors are not connected, then the distance from one to the other is infinite. Since we can not (meaningfully) sum over infinite distance, if the reciprocal is taken, then this measure is zero at the limit. Essentially, this uses the shortest distance as a proxy measure of disruption. In other words, destabilization of the network is based on disrupting the shortest paths in the network which can represent "communication, influence, resources, and so forth." [68, p. 1-3]

[1] Borgatti's definition of fragmentation differs from the definition used by other researchers where it calculates the number of components and diameter of the largest component

This is a key insight to disruption of social networks. If the goal of disruption is to disconnect/isolate portions of the network, calculation of traditional SNA measures such as centrality and reachability becomes problematic as they are designed for connected graphs/networks.

Borgatti suggested finding the maximum of the measure D_F through complete enumeration. However, Herbranson showed through modeling and analyzing the key player problem as an operations research problem that complete enumeration can be very computationally inefficient. [68, p. 4-5]

Herbranson instead enhanced this model by parameterizing the solution space using the size of the set to be removed, included an objective function to represent the ease or difficulty of removing an actor, and the use of arc distance other than one. To do this, he modified the fragmentation measure slightly to

$$F = 1 - \frac{2 \sum\limits_{i>j} \frac{1}{d_{ij}}}{S} \tag{2.14}$$

where d_{ij} is the shortest distance between nodes i and j, and S is $\sum\limits_{i>j} \frac{1}{s_{ij}}$. [68, p. 4-9] In addition, instead of solving this via complete enumeration, Herbranson provided two methods to solve this problem: a dynamic programming approach and an integer programming approach. [68, p. 4-18]

Finally, Herbranson developed an additional model: [68, p. 4-23]

$$\max \quad {}^{D}TF = 1 - \frac{\sum\limits_{j \notin T} \max\{\frac{1}{d_{ij}} | i \in T\}}{n - |T|}, T \subset N \tag{2.15}$$

where T is the target subset determined *a priori*, and d_{ij} is the minimum shortest path distance from any node $i \in T$ to any other node j in the network. ${}^{D}TF$ is the summation of the shortest distances from the set T to all other nodes in the network. Similar to Borgatti's measure, if i and j are not connected, the distance is assumed to be infinite, so the reciprocal is zero. Herbranson developed a heuristic to solve this model.

Regardless of which key player measure and algorithm is used, there is an implicit assumption that targeted actors are accessible. This may not be the case in real-world networks. However, with defined flows, a network fortification/interdiction technique would create a target set which considers these factors.

2.2.8 Infrastructure Networks

The definition of infrastructure networks has evolved considerably, especially since the attacks of September 11, 2001. A comprehensive reference of this evolution is given by Moteff and Parfomak. [88] The definition used here are that given by the USA PATRIOT Act for two reasons. One, it is public law and not as influenced by adminstration policy changes and interpretations (without congressional action) as other definitions. Second, it forms the core for all subsequent definitions of infrastructure and critical infrastructure. The USA PATRIOT Act defines critical infrastructure as

> systems and assets, whether physical or virtual, so vital to the United States that the incapacity or destruction of such systems and assets would have a debilitating impact on security, national economic security, national public health or safety, or any combination of those matters (Sec. 1016(e)). [88, p. 10]

Policy decisions have expanded on this definition and provide explicit illustrations. For example, HSPD 7 defines the following 13 networks as critical infrastructures: agriculture, food, water, public health, emergency services, government, defense industrial base, information and telecommunications, energy, transportation, banking and finance, chemical industry, postal and shipping. [88, p. 11]

These infrastructure networks can be modeled as network flow models. To this end, definitions from Dudenhoeffer *et al.* is used. An infrastructure node is defined as "an entity that acts as a source, produces, consumes, or transforms a resource."

Similarly, an edge is “a physical or virtual entity that acts as a conduit for flow for a physical quantity, information or influence.” [51, p. 479] With these definitions, infrastructure networks can be modeled with maximum flow, minimum cost flow, and so forth.

Objective functions are similarly constructed.

> Prescriptive problems, such as budget allocation, network planning and design, risk management and emergency response problems, aim to optimize the overall network performance based on system-level criteria such as cost minimization, social surplus maximization, risk minimization, or recovery time minimization after network failure. [139, p. 154]

Each of these criteria becomes a candidate to maximize or minimize in an objective function.

2.2.8.1 Interdependencies

The consideration of interdependencies has been used to varying degrees of success; however, almost all research on infrastructure networks considers each network separately, in isolated analysis. This is because these networks are “complex even at an individual level leading to a significant degree of difficulty if the scope is broadened to include multiple systems.” [139, p. 149] However, although each infrastructure is defined and enumerated individually, “each system is composed of numerous interconnected and interdependent cyber, physical, social, and organizational infrastructures, whose relationships are dynamic, nonlinear, probabilistic, and spatially distributed.” [65, p. 33] Therefore, any analysis of infrastructures must take these interdependencies into account.

The growing interdependence and associated vulnerabilities in networked systems has been highlighted in public law. The USA PATRIOT Act states

> Private business, government, and the national security apparatus increasingly depend on an interdependent network of critical physical and

> information infrastructures, including telecommunications, energy, financial services, water, and transportation sectors (Sec. 1016(b)(2)). [88, p. 10]

However, models that account for interdependencies are sparse. "In all cases, the research addressed one infrastructure system and the service it provides and did not consider interdependencies among infrastructure systems." [80, p. 3] The only exception they noted was done by Rinaldi *et al.* [101] who provided useful definitions, but "stopped short of modeling the vulnerability of networks." [80, p. 3] In August 2006, the Idaho National Laboratory published a survey of all available infrastructure models to determine the current state-of-the-art in the field of infrastructure interdependency analysis. They found 30 infrastructure models which perform some level of interdependency modeling. [94] Most of these models are either agent-based or monte carlo simulations; such models give valuable insight, but do not provide deterministic vulnerability analysis. In addition, some of the models are commercial products with limited published documentation, making it difficult to determine what underlying algorithms/methodologies were used. Wallace *et al.* [122] and Kennedy [72] independently developed methods to model layered networks (which could include infrastructure networks). These are discussed in Section 2.4.

To highlight the importance of considering interdependent effects, consider the following. In studying the vulnerability of the Saudi Arabian pipelines, Brown *et al.* noted, "pipeline systems for crude oil and refined petroleum products are sparsely connected because of the enormous expense required to acquire right-of-ways, lay pipe, build pumping stations and maintain the system once it is complete." [27, p. 126] They note that the network (especially pipelines) covers a huge area that cannot be patrolled completely, but "pipelines can usually be repaired fairly quickly." [27, p. 129] Therefore, they state that operational effects of attacking pipelines would "not last for long." Of course, the environmental impact would linger.

However, considering the interdependencies of the pipeline and electrical infrastructures could result in a more robust attack. Even if the goal was only to disrupt the flow through the pipelines, a layered infrastructure analysis may suggest attacking portions of the electric infrastructure. In other words, instead of attacking the pipeline directly, one could attack the transformer that serves the pipeline (along with the other vulnerable targets within the pipeline infrastructure). Brown *et al.* even noted that transformers "pose special difficulties because they are big, heavy, and expensive; few spares exist; and a replacement might have to be ordered from, built by, and shipped from an overseas manufacturer." [27, p. 124] Therefore, a layered analysis would indicate that attacking the transformer would potentially have a significantly longer impact than attacking the pipeline itself.

2.2.8.2 Strategy for Protection

Because of the heavy reliance on infrastructures as part of our way of life, their protection is vital. Strategic guidance/objectives for critical infrastructure protection are provided in a series of National Strategy documents, congressional acts, presidential directives, and DoD directives.

In 1996, President Clinton issued Executive Order 13010 [43] which established the Presidents Commission on Critical Infrastructure Protection (PCCIP) to assess the national dependency on information infrastructures. The commission was charged with developing a comprehensive national policy and implementation strategy for protecting critical infrastructures from physical and cyber threats and assuring their continued operation. In October 1997, this commission published a report that identified eight critical industries susceptible to disruption (through physical and/or cyber attacks). To deal with these vulnerabilities, they recommended a partnership between the public and private sectors to address new vulnerabilities, shared threats, and shared responsibilities.

In 1998, President Clinton issued Presidential Decision Directive (PDD) 63 (titled "Critical Infrastructure Protection") [44] which established the national policy on necessary measures to identify and eliminate significant vulnerabilities to physical and cyber attacks on U.S. critical infrastructures. This PDD has since been superseded by HSPD-7 (to follow).

In 2002, President Bush released the "National Strategy for Homeland Security." The National Strategy focused homeland security functions into six critical mission areas; one of which is protecting critical infrastructure. [129, p. vii] The strategy identifies six critical mission areas. The critical mission area which is relevant to this research is "protecting critical infrastructures and key assets [129, p. 29]." The strategy suggests we "view our vulnerabilities from the perspective of terrorists, and to provide objective data on which to base infrastructure protection standards and performance measures." [129, p. 33] This is the strategy that was followed in this research. The role of a terrorist is assumed to identify those assets/links that could potentially cause the most damage to our nation.

Later that year, the Homeland Security Act created the Department of Homeland Security (DHS) [45]. The act gave DHS responsibility for conducting vulnerability assessments of critical infrastructures and developing a comprehensive plan to secure them. In addition, they were charged with recommending measures necessary to protect critical infrastructures.

In February of 2003, two complimenting strategy documents were released which implement National Strategy for Homeland Security in the critical infrastructure protection area: "National Strategy for the Physical Protection of Critical Infrastructure and Key Assets," [130] and the "National Strategy to Secure Cyberspace" [131].

The "National Strategy for the Physical Protection of Critical Infrastructures and Key Assets" focuses on reducing the Nation's vulnerability by protecting our critical infrastructures from physical attack. To do this, it defines several end state strategic objectives. Two objectives relevant to this research are: "identify and assure the protection of those infrastructures and assets that we deem most critical" [130, p. 2]; and "pursue collaborative measures and initiatives to assure the protection of other potential targets that may become attractive over time." [130, p. 3]." To meet these objectives, the federal government must "identify the critical nodes upon which assets depend; assess associated vulnerabilities; and implement appropriate steps to mitigate those vulnerabilities and protect the infrastructures and assets under its control." [130] For example, this includes "comparing the robustness of different infrastructures at points where key centers or critical nodes are in close proximity to one another and can have cascading effects if attacked." [130, p. 34]

The focus of the "National Strategy to Secure Cyberspace" is on the identification, assessment and protection of interconnected information systems and networks. To achieve this goal it outlines three strategic objectives: "prevent cyber attacks against America's critical infrastructures; reduce national vulnerability to cyber attacks; and minimize damage and recovery time from cyber attacks that do occur." [131, p. viii]

In addition, this strategy outlines initiatives to reduce threats and related vulnerabilities. Two of these initiatives are to develop a methodology to conduct "vulnerability assessments to understand the potential consequences of threats and vulnerabilities; and understand infrastructure interdependencies and improve the physical security of cyber systems and telecommunications." [131, p. 33] One of the goals of this research is to provide one such methodology to identify vulnerabilities across multiple infrastructures.

Later in 2003, President Bush issued "HSPD 7: Critical Infrastructure Identification, Prioritization, and Protection." This directs the identification and prioritization of United States critical infrastructure and key resources and directs protection of them from terrorist attacks. [69]

Finally, in August 2005, the National Infrastructure Protection Plan (NIPP) was released which defines infrastructure protection roles and responsibilities for government and industry. It builds on all previously released Strategies, and it fulfills requirements in HSPD 7 and HSA of 2002.

2.2.8.3 Vulnerability

Unfortunately, infrastructures have many characteristics which make them vulnerable. These characteristics include:

> openness and accessibility (designed for efficiency and convenience); extent and ubiquity (vast physical infrastructure); diversity of owners, operators, users, and overseers (controlled by thousands of state and local governments, along with some private business and individual ownership); entwinement with society and the global economy (science and technology, arts, culture, products, and commodities move across international boundaries). [65, p. 34]

To analyze infrastructures and determine vulnerability, system reliability analysis of infrastructure networks generally "pronounce the system robust if there is no single point of failure." [28, p. 530] In addition, fault tree analysis is also used. This method identifies cut-sets that are most likely to disrupt a network, and this technique "pronounces the system robust if the combined probability of occurrence is low." [28, p. 530] "These results must be classified as a guess." [27, p. 105] In addition, these techniques are insufficient for vulnerability analysis due to terrorism. It has been found that a "lone attacker with a high-powered rifle could gravely damage an entire electric power grid by targeting highly reliable components at just a few substations." [27, p. 105] Therefore, Brown *et al.* [27] argue that network-interdiction techniques are more appropriate to determine the criticality of a group of system components.

To further complicate matters, infrastructure networks are "planned, designed and operated by different public, private and/or public-private sectors without explicit coordination." [139, p. 149] As a result, this leads to "wasted resources, operational inefficiencies, and at times cripples some subnetworks completely." [139, p. 150] An additional source of complexity results from the difficulty in enabling coordinated investment decisions due to the disparate nature of the ownership of the different infrastructure network layers. [139, p. 155] Further, these "entities can have different goals, strategies, and financial capabilities." [139, p. 155]

2.2.9 Network Design

In addition to modeling current networks, operations research techniques can also be used to assist in the design and improvement of networks. Frank and Frisch discuss a number of algorithms for special cases of network design. For minimum cost network design, they point out that this can be done "by resorting to linear programming formulations." [55, p. 255]

The minimum cost synthesis problem is simply stated. Given a set of demands, find a network of feasible flows such that the cost, c_{ij}, is minimal. Specifically,

$$\begin{array}{lll} \text{Minimize} & \sum\limits_{(i,j)\in A} c_{ij} z_{ij} & \\ \text{s.t.} & \sum\limits_{\{j:(i,j)\in A\}} x_{ij} - \sum\limits_{\{j:(j,i)\in A\}} x_{ji} = b(i) & \forall i \in N \\ & l_{ij} z_{ij} \leq x_{ij} \leq u_{ij} z_{ij} & \forall (i,j) \in A \\ & z_{ij} \in \{0,1\} & \end{array} \tag{2.16}$$

where, again, $b(i)$ is zero for transshipment nodes, greater than zero for supply nodes, and less than zero for demand nodes. [63, p. 348-349] In addition, $z_{ij} = 1$ if arc (ij) is constructed, and zero otherwise.

This network design problem considers a set of nodes and potential arcs which can be constructed at a fixed cost. The problem then becomes to find the minimum-cost set of arcs to add to the network such that a feasible flow exits. However, as LeBlanc and Boyce pointed out, in reality, network design is usually characterized by multiple levels of decision making. For example, at one level "government or industry officials make one set of decisions, which seek to improve the network's performance." [79, p. 259] At another level, users of the network wish to minimize their costs, and while their decisions "can be predicted, their decisions can not be dictated." [79,

p. 259] Therefore, LeBlanc and Boyce develop a bilevel formulation of the network design problem with user-optimal flows. The specific formulation is transportation specific and is not repeated here. However, it is an important realization that network designers and users often have different objective functions.[1]

2.2.9.1 Vulnerability & Survivability

> Today's interconnected, continent-wide power grids are much better than their local and regional predecessors at providing cheap and reliable power, and they are significantly less prone to local breakdowns. But when they do crash, the consequences are far greater than those of the more frequent and more localized failures of past decades . . . Thus, modern societies have made an unintentional Faustian bargain that brings increases in operational efficiency and capability at the cost of greater susceptibility to widespread catastrophic failures. [85, p. 2]

In addition to simply constructing networks to satisfy demand at the least cost, in most networks it is also important to consider the vulnerability and survivability of the network under failures and/or attacks. The terms "vulnerability" and "survivability" are defined differently in different contexts. This research uses the definitions provided by Clarke and Anandalingam: "Survivability is the ability of a network to perform according to a specification after it has been damaged ... Vulnerability is concerned with the difficulty of destroying the network." [42, p. 921]

Both terms are associated with the potential destruction of a network. Unfortunately, Clark and Anandalingam do not define what it means to destroy a network, so the following discussion from Frank and Frisch is provided:

> A system modeled by a graph may be considered destroyed if, when vertices or branches are removed, the resulting graph G satisfies one or more of the following conditions:
>
> - G contains at least two components.

[1] LeBlanc and Boyce use Bard's original technique to solve the bilevel program which has since been shown does not guarantee optimal solutions.

- There are no direct $s_i - t_i$ paths for specified sets of vertices $\{v_{s_i}\}$ and $\{v_{t_i}\}$.
- The number of vertices in the largest component of G is less than some specified number.
- The shortest $s_i - t_i$ path is longer than some specified number. [55, p. 300-301]

Bullet two is perhaps the most studied due to the importance of connectivity in many networks. For example, "one of the major functions of a communication network is to provide connectivity between users." [64, p. 5] It is also important to note that in most formulations "cost represents the cost of setting up the topology" of the network. While user and other costs are important, "it is usually the case that a topology is designed first and then these other costs are considered in a second stage of optimization." [64, p. 8]

The goal in such a case is to build a minimum-cost network that satisfies the required edge and/or node survivability conditions. Let r_{st} represent the edge survivability requirement that there are at least r_{st} edge disjoint (s,t) paths. In other words, at least r_{st} edges must be removed to disconnect the graph. In addition, let k_{st} and d_{st} represent node survivability as follows: the removal of at most k_{st} nodes leaves at least d_{st} edge disjoint (s,t) paths. With this, the formulation is as follows:

$$\begin{array}{lll}
\min & \sum\limits_{ij \in E} c_{ij} x_{ij} & \\
\text{s.t.} & \sum\limits_{i \in W} \sum\limits_{j \in V \setminus W} x_{ij} \geq r_{st} & \forall (s,t) \in V, s \neq t, \forall W \subseteq V, s \in W, t \notin W \\
& \sum\limits_{i \in W} \sum\limits_{j \in V \setminus (Z \cup W)} x_{ij} \geq d_{st} & \\
& 0 \leq x_{ij} \leq 1 & \forall ij \in E \\
& x_{ij} \text{ integer} & \forall ij \in E
\end{array} \tag{2.17}$$

where (s,t) is the path from s to t; and $\forall z \subseteq V \setminus \{s,t\}$ and $|Z| = k_{st}$.

The objective function is the sum of costs of all edges used in the design. The first constraint states that for each possible subset of nodes, there must be at least r_{st} edges with one endpoint in the subset and the other endpoint outside the subset (in its complement). In addition, the second set of constraints should have the following declaration (but was not included above due to space limitations): $\forall(s,t) \in V, s \neq t, \forall W \subseteq V\backslash Z, s \in W, t \notin W$.

Grötschel *et al.* note that the classical network synthesis problem for multiterminal flows is obtained from this formulation by dropping the second and forth constraints. [64, p. 10]

It should be noted that this formulation is for undirected networks. For directed networks, "we simply replace the notion of an undirected path by a directed one." [64, p. 66]

2.3 Multilevel Programming

Multilevel programs can be viewed as n-person, nonzero-sum games with perfect information. In addition, there is a specified order of play, and non-disjoint control sets. [86, p. 7] Decisions made by higher level players affect the decision space available to lower level players through their objective functions. "Each player's control instruments may allow him to influence but not dictate the policies of another and thereby improve his own performance through the resultant externalities." [86, p. 9]

As shown in Figure 2.1, this area of mathematical programming provides the foundation for network interdiction formulations to be discussed in Section 2.3.7.3. Because different aspects of this foundation are explored, various subsets are discussed including multiojective multilevel programming, multiple followers, multiple optimal solutions, coalitions, solution techniques, reformulation techniques, and special cases.

Mathematically, multilevel programming can be written as a nested optimization problem. Unfortunately, although this formulation represents a variety of practical problems, these problems are very difficult to solve. Even in the simplest case of two levels, where both objective functions and all constraints are linear, this problem is strongly $\mathcal{NP}$-hard. [67, p. 1197] The principal difficulty results from nonconvexity. [86, p. 11] This is true even under the simplifying assumptions (which are usually required for a solution to even exist) of rational players who act noncooperatively.

Bard suggests there are 5 definitions specific to bilevel programming problems (BLPPs) [10, p. 196] (which are a subset of multilevel programming):

- Constraint region of the bilevel programming problem (BLPP): $S \triangleq \{(x, y) : x \in X, y \in Y, A_1x + B_1y \leq b_1, A_2x + B_2y \leq b_2\}$. This constraint region represents all possible choices that can be made by the leader and follower. Savard proved that at least one optimal (global) solution is attained at an extreme point of this set. [104] (as used by [121, p. 6]) This leads to vertex enumeration methods [which are discussed later] to find optimal solutions.
- Feasible set for the follower for each fixed $x \in X$: $S(x) \triangleq \{y \in Y : B_2y \leq b_2 - A_2x\}$. The follower's feasible region is affected by the leader's choice of x.
- Projection of S onto the leader's decision space: $S(X) \triangleq \{x \in X : \exists y \in Y, A_1x + B_1y \leq b_1, A_2x + B_2y \leq b_2\}$. The leader moves first by minimizing x subject to leader's and follower's constraints.
- Follower's rational reaction set for $x \in S(X)$: $P(x) \triangleq \{y \in Y : y \in \arg\min[f(x, \hat{y}) : \hat{y} \in S(x)]\}$. The follower will observe the leader's action, and (assuming he is rational) he will select y from his feasible set that minimizes his objective function.
- Inducible Region: $IR \triangleq \{(x, y) : (x, y) \in S, y \in P(x)\}$. This region is the set over which the leader optimizes his objective function.

The most studied case of the multilevel program is the bilevel case. For $x \in X \subset \mathbb{R}^n$, $y \in Y \subset \mathbb{R}^m$, $F : X \times Y \to \mathbb{R}^1$, and $f : X \times Y \to \mathbb{R}^1$, the BLPP is written as follows:

$$\begin{aligned} \underset{x \in X}{min} \;\; & F(x, y) = c_1x + d_1y \\ & s.t. \;\; A_1x + B_1y \leq b_1 \\ & \underset{y \in Y}{min} \;\; f(x, y) = c_2x + d_2y \\ & s.t. \;\; A_2x + B_2y \leq b_2 \end{aligned} \tag{2.18}$$

where $c_1, c_2 \in \mathbb{R}^n$, $d_1, d_2 \in \mathbb{R}^m$, $b_1 \in \mathbb{R}^p$, $b_2 \in \mathbb{R}^q$, $A_1 \in \mathbb{R}^{p \times n}$, $B_1 \in \mathbb{R}^{p \times m}$, $A_2 \in \mathbb{R}^{q \times n}$, $B_2 \in \mathbb{R}^{q \times m}$.

2.3.1 Multiobjective Multilevel Programming

Wang *et al.* developed a method to generate all nondominated solutions to a multiobjective multilevel program. In this problem, the leader has a multiobjective problem, while the followers all have single objective problems. Let $P^{(n)}$ denote the original problem. $P^{(n-1)}$ is constructed by combining the first and second level problems in problem $P^{(n)}$ excluding the objective function of the lower level problems. [140, p. 179] For example, for a bilevel problem, the follower's objective function is removed (while the constraints remain). In addition, $LP^{(n-1)}$ is formulated by combining the second level through the nth level problems excluding the objective functions at the third through the nth level problems. [140, p. 179] For example, in a bilevel problem, this simply refers to the followers problem (i.e. the follower's objective function and constraints).

As Wang *et al.* point out, these definitions give the set of bases (extreme points) with respect to x corresponding to a given vector y_1 to the first level problem. In other words, the first level constructs the set of feasible extreme points by combining the set of feasible extreme points with respect to y_1 and the set of bases with respect to x provided from the second level problem. [140, p. 180] Nondominated extreme points are found by searching the set of feasible extreme points to this problem. Wang *et al.* show that these nondominated extreme points are also feasbile extreme points in the original problem.

2.3.2 Multiple Followers

Anandalingam noted that most organizations are actually characterized by one higher-level decision maker and k lower-level decision-makers (on an equal level).

He extended the typical bilevel formulation to accommodate this in the following formulation: [6, p. 1025]

$$
\begin{aligned}
&\max_{x_1} \sum_{j=1}^{n} c'_{1j} x_j && \\
&\max_{x_i} \sum_{j=1}^{n} c'_{ij} x_j && i = 2, \ldots, k \\
&\text{s.t. } \sum_{j=1}^{k} A_j x_j \leq b && \\
&\qquad x_i \geq 0 && i = 1, \ldots, k
\end{aligned}
\tag{2.19}
$$

where the c' are vectors. [6, p. 1023] Using the Karush-Kuhn-Tucker (KKT) reformulation techniques on problem (2.19), the model can be expressed using the KKT conditions of all followers as:

$$
\begin{aligned}
&\max_{\{x_1,\ldots,x_k\},w} \sum_{j=1}^{n} c'_{1j} x_j && \\
&\qquad w_i A_i = c_{ii} && i = 2, \ldots, k \\
&\qquad w(\sum_{i=1}^{k} A_i x_i - b) = 0 && \\
&\qquad \sum_{i=1}^{k} A_i x_i \leq b && \\
&\qquad w_i, x_i \geq 0 && i = 1, \ldots, k
\end{aligned}
\tag{2.20}
$$

where $w = (w_1, \ldots, w_k)$ and the constraints are optimality conditions for all the divisional problems. This formulation was extended to allow nonlinear objective functions and constraints as follows:

$$
\begin{aligned}
\max_{\{x_1,\ldots,x_k\}} \; & f_1(x) && \\
\text{s.t. } & \nabla_i f_x(x) - w_i \nabla_i g(x) = 0 && i = 2,\ldots,k \\
& wg(x) = 0 && \\
& g(x) \leq 0 && \\
& w_i \geq 0 && i = 2,\ldots,k
\end{aligned}
\tag{2.21}
$$

where $w = (w_2, \ldots, w_k)$, and ∇_i is the gradient with respect to x_i.

Wang *et al.* independently developed this formulation and made some additional observations. First, they prove (similar to traditional bilevel programs) that at least one optimal solution is a vertex of the constraint region. [123, p. 272] Therefore, vertex enumeration methods would work for this formulation as well. For example, the kth best method could be used where the main difference in its use is in the feasibility test. To check for feasibility, n linear programming problems would be required to be analyzed. [123, p. 275]

In addition, Wang *et al.* show that the branch and bound algorithm developed by Moore and Bard also works for these problems with minor modifications. [123, p. 273] They demonstrate this with several simple examples.

2.3.3 Multiple Optimal Solutions

The BLPP may not have a solution. "If $P(x)$ is not single-valued for all permissible x, the leader may not achieve his minimum payoff over IR." [10, p. 196] "In this case, the follower would be indifferent to any point on that hyperplane; however, the leader might have a specific preference . . . but there may be no way to induce the follower to select that point." [10, p. 302]

Bard suggests there are three possibilities to deal with this problem. "The first would require replacing the 'min' with 'inf' and define ϵ-optimal solutions." [10, p. 303] The second approach "argues for a conservative strategy that redefines the problem ..." with a min max formulation. [10, p. 303] This is known as a "pessimistic strategy." The final option Bard discussed for dealing with multivalued $P(x)$ is "to assume some level of cooperation among the players and rewrite the leaders problem." [10, p. 304] This is known as an "optimistic strategy." Of course, this violates the basic assumption of noncooperation. Bard argues that if players cooperate, then multiobjective programming might be a better alternative. This is discussed in Section 2.3.4.

Bialas and Karwan suggest using an incentive scheme which would provide a "kick back" of level one's earnings to encourage level two to choose its most desirable solution. For example, $f_2^*(x) = f_2(x) + \epsilon f_1(x)$. [17, p. 1008] This may not lead to a unique solution since the leader's solution may also have the same value for multiple values of the follower's solution. However, any of these solutions would satisfy both the leader and follower. In some algorithms, single valued follower solutions are "only needed for an optimal choice for" the leader. [127, p. 184] Here, this assumption is only needed to "get the exact penalty result." [128, p. 399]

Multiple optimal solutions, however, have generally not been a concern in the case of interdiction problems, (where the leader's and follower's decision variables are the same, and the objective functions are negatives of each other). This is because $P(x)$ is always a singleton in this case, as the objective function has the same value for all $y \in Y(x)$. [70, p. 113] However, as is discussed later, network interdiction problems are generally solved by replacing inner optimization problems with their dual formulation. Unfortunately, this introduces an often ignored computational difficulty. These dual formulations themselves can have multiple optimal solutions.

As Smith noted, "the existence of alternative optimal dual solutions in this case implies that several cuts can be generated from each dual solution passed to the subproblem." [109, p. 4] To combat the problem of potentially exponential cuts, Smith developed a cutting plane technique through a reformulation of the problem. [109, p. 5] These methods are, unfortunately, specific to "product placement" formulations of the problem. While not considered here, it should be noted that there is an opportunity to extend these methods to more general formulations.

2.3.4 Coalitions

"[I]t has been shown that Pareto and bilevel optimality are distinct concepts. Even in the case of linear bilevel programming, no simple relationship exists, however close the objectives of the two levels." [84, p. 358] In other words, optimal solutions to multilevel programs are usually not Pareto optimal. Therefore, there are often solutions to the multilevel program in which at least one of the decision makers can achieve a better solution than the optimal multilevel solution (with no change in other decision maker solutions). However, these solutions are not achievable in the multilevel formulation because players can not cooperate to achieve them. The following example taken from Moore demonstrates this: [86, p. 37]

$$\begin{array}{lll} \max\limits_{x\geq 0} F(x,y) = & -x-y & \\ & \max\limits_{y\geq 0} f(x,y) = & 5x+y \\ & \text{s.t.} & -x-\frac{y}{2}\leq -2 \\ & & -\frac{x}{4}+y\leq 2 \\ & & x+\frac{y}{2}\leq 8 \\ & & x-2y\leq 4 \end{array}$$

The optimal solution to this problem is $(\frac{8}{9},\frac{20}{9})$ with an objective function value of $-\frac{28}{9}$. However, "the point $(2,0)$ provides better outcomes for both players but is not in the inducible region." [86, p. 43] The leader's solution improves from $-\frac{28}{9}$ to -2, and the follower's solution improves from $\frac{20}{3}$ to 10. (Note: as Moore points out, if the variables are restricted to integer, the optimal solution is $(1,2)$ with $F=-3$, and the leader and follower do better than the relaxed problem. [86, p. 41])

A natural question arises as to why these points are not the optimal solution. The reason is because of the underlying assumption that players are rational and make decisions sequentially and independently. Therefore, if the leader in the above problem chose $x=2$, then the follower would maximize his objective function and choose $y = 3$ (thus reducing the leader's objective function value i.e. $F = -5$). Therefore, the leader (also a rational decision maker) will not choose $x = 2$ on the hope that the follower will choose $y = 0$. In order to achieve the improved solution, the non-cooperative restriction/assumption must be relaxed.

However, as Bard points out, "if the players are allowed to cooperate, then the preferable strategy would be to seek a Pareto-optimal solution." [10, p. 304] This leads to strategies for coalition formation. In effect, the multilevel solution becomes the "fall-back" solution if the decision makers fail to cooperate. Two types of coalitions are discussed. First, the decision makers may elect to act for the benefit of the group as a whole. This type is discussed next. In the second, and perhaps more realistic case, the decision makers seek to maximize their benefit from cooperating relative to the multilevel solution. This type of coalition and solution methodology is discussed in the next subsection.

If the restriction against cooperation is dropped, coalitions may form where members of each coalition act to increase their own benefit and/or for the benefit of the coalition as a whole. As Bialas noted, "a formation of a coalition among subsets of the players could provide a means to achieve Pareto-optimality." [18, p. 2440] Chew studied this problem and provided a methodology for prediction coalition formulation. Specifically, Chew defined a "strong contract region" which are points where all objective function levels are increased; and a "weak contract region" where some levels do better, but others remain the same or do worse. The problem of multiple optimal solutions of the follower's problem discussed in the previous section can be viewed as a special case of a weak contract region, and coalitions may form there as well. [38]

The following definitions from Bialas are used to develop the theory of coalition formation: [19, p. 3]. Let $G = \{1, 2, \ldots, n\}$ be the set of n players where 2^G denotes the set of all possible coalitions of G. Let $\mathcal{P} = \{R_1, R_2, \ldots, R_M\}$ be the coalition structure or partition of G into non-empty coalitions. As a result of coalition formation, the objective function of each player in R_j becomes $f'_{R_j}(x) = \sum_{i \in R_j} f_i(x)$. [18, p. 2440] Let $R(i)$ be the unique coalition $R_j \in \mathcal{P}$ such that player $i \in R_j$. Therefore, instead of maximizing $f_i(x)$, player i will now maximize $f'_{R(i)}(x)$. The value of coalition $R_j \in \mathcal{P}$ is given by: [18, p. 2440]

$$v(R_j, \mathcal{P}) \equiv \sum_{i \in R_j} f_i(\hat{x}(\mathcal{P})) \tag{2.22}$$

where $\hat{x}(\mathcal{P})$ is the solution to the n-level optimization problem resulting from the new objective functions.

The core is made up of undominated solution configurations which are the pairs $(r, \mathcal{P})$ where r is an n-dimensional vector whose elements r_i represent the payoff to each player i under coalition structure $\mathcal{P}$. "Once players have negotiated an outcome within the core, no further negotiations or outcomes are possible." [19, p. 3] If no core exists, Willick provides a linear program to determine optimal coalition payoffs to individuals. [134, p. 21] "A solution is an element of the core if it divides the money available from the game in a manner in which every coalition receives at least what it can obtain from playing the game." [135, p. 7] In other words,

> each coalition earns the combined proceeds that each individual coalition member would have received under the original Stackelberg game. Therefore, a player's rational decision may now be altered because he may also be acting for the joint benefit of the members of his coalition. [18, p. 2441]

Willick states that it seems to be reasonable to search for solutions in the core. However, there does not exist an efficient method to find solutions in the core. Willick also points out that sometimes the core is empty and sometimes there are multiple solutions in the core, from which one must choose. "No general existence theorem has been given for the distribution of wealth among the individuals in an n-person game in characteristic function form such that the distribution is always stable." [135, p. 8]

Therefore, the literature does not seem to suggest an efficient method to determine/predict coalition formation. However, it is known that with a mathematical model, modifications to the organizational structure can be made to encourage or dissuade levels from forming coalitions. [38, p. 2]

2.3.4.1 Cooperation via Post-Optimization Analysis

Instead of looking at the "core," Wen and Hsu define a "feasible contraction set, S'." These are the set of points which satisfy the following system of inequalities: [126, p. 356]

$$\begin{aligned} Ax + By &\leq r \\ F(x,y) &\geq F(x^*,y^*) \\ f(x,y) &\geq f(x^*,y^*) \end{aligned} \tag{2.23}$$

These are points which improve at least one level's solution without decreasing the other. Wen and Hsu prove that the interior of this set is empty if and only if (x^*,y^*) is Pareto-optimal.

After a non-Pareto-optimal solution is found from multilevel programming, decision makers may realize that they might benefit by moving to solutions in the feasible contraction set. However, there are usually multiple efficient solutions, all of which would give both decision makers at least as much as the non-Pareto-optimal solution. One method to chose such a point (or at least provide various points from which to choose) is provided by Soismaa using "asymmetric Nash bargaining." [114, p. 429]

For example, let (x^*,y^*) be the optimal solution to a bilevel programming problem. The asymmetric Nash bargaining solution is provided by

$$\begin{aligned} \max \quad \pi(x,y) =& [F(x,y) - F(x^*,y^*)]^{\alpha} \times [f(x,y) - f(x^*,y^*)]^{1-\alpha} \\ \text{s.t.} \ & (x,y) \in S \end{aligned} \tag{2.24}$$

where $\alpha \in [0, 1]$ represents the bargaining power of the upper level decision-maker. [114, p. 429] By using various values of α, the threat-point, ideal-point, and ideal-threat-point solutions defined/developed by Wen can be obtained. [126] In addition, by varying this value "it is possible to trace the whole relevant part of the efficient frontier." [114, p. 430] This allows the decision makers to see all efficient solutions and compromise to choose one among them. [114, p. 431] Wen also provides a numerical example, and points out that if a single point is desired, "it is not obvious" how one would determine the numerical value of the bargaining power parameter. [126]

2.3.5 Solution Techniques for linear BLPPs

Several algorithms have been developed to solve linear BLPPs. Unfortunately, few of these algorithms can be applied to general linear bilevel formulations of modest size or larger. The most widely studied and used algorithms are the following: branch and bound, penalty methods, kth best, and hybrid methods. Because of their wide acceptance and use, these are discussed in turn, along with their respective potential applicability in interdiction type problems.

A classification system was developed by Israeli to match algorithms to formulations that they are best suited to solve. Israeli classified bilevel programs as either "positive" or "negative" according to the relationship between the leader's and follower's objective functions. "Positive" formulations are those where there is a positive correlation between the objective functions of the leader and follower. Israeli continues by pointing out that most existing algorithms for bilevel problems work best for positive formulations. Of course, interdiction problems (where the objective functions are diametrically opposed) are non-positive as there is a strong negative correlation between the objective functions. Unfortunately, "positive algorithms are likely to have poor performance when applied to" interdiction problems. [70, p. 116]

No studies were found that determine which algorithms work well for negatively correlated objective functions.

2.3.6 Reformulation

To facilitate solution methods, multilevel programs are often reformulated to programs with fewer levels. These single level formulations have nonlinear constraints. Therefore, they can be solved via traditional nonlinear techniques. However, due to the difficulties that can arise from solving nonlinear problems with general approaches, specialized algorithms have also been developed which exploit the structure of the single level programs to facilitate solving them with linear programs.

Bard and Moore developed a method to do this using KKT conditions. [11, p. 282] The idea is that the follower's problem is replaced with its KKT conditions which are then appended to the leader's problem. For example, let $u \in R^q$ be the dual variables associated with the follower's constraints, and let $v \in R^m$ be the dual variables associated with $y \geq 0$. Then, formulation (2.18) becomes

$$\begin{aligned} \min\ & c_1x + d_1y \\ \text{s.t.}\ & A_1x + B_1y \leq b_1 \\ & A_2x + B_2y \leq b_2 \\ & uB_2 - v = -d_2 \\ & u(b_2 - A_2x - B_2y) + vy = 0 \\ & x \geq 0, y \geq 0, u \geq 0, v \geq 0 \end{aligned} \tag{2.25}$$

where $c_1 \in \mathbb{R}^n$, $d_1 \in \mathbb{R}^m$, $b_1 \in \mathbb{R}^p$, $b_2 \in \mathbb{R}^q$, $A_1 \in \mathbb{R}^{p\times n}$, $B_1 \in \mathbb{R}^{p\times m}$, $A_2 \in \mathbb{R}^{q\times n}$, and $B_2 \in \mathbb{R}^{q\times m}$.

Hansen *et al.* revised this formulation to explicitly allow constraints to remain with the leader. [67, p. 1195] These are constraints which are only binding on the leader, but can depend on the decisions of the follower. In other words, the constraints are binding on the leader, but include some of the follower's decision variables.

$$\begin{aligned}
\min\ & c_1x + d_1y \\
\text{s.t. } & A_1x + B_1y \leq b_1 \\
& A_2x + B_2y \leq b_2 \\
& uB_1 + vB_2 - w = -d_2 \\
& u(b_1 - A_1x - B_1y) + v(b_2 - A_2x - B_2y) + wy = 0 \\
& x \geq 0, y \geq 0, u \geq 0, v \geq 0, w \geq 0
\end{aligned} \tag{2.26}$$

where $u \in R^p$ are the dual variables associated with the leader's constraints, $v \in R^q$ are the dual variables associated with the follower's constraints, and $w \in R^m$ are the dual variables associated with the $y \geq 0$ constraint.

Anandalingam used a similar method to transform a trilevel problem into a single level problem. However, Sinha pointed out some errors in the development, and published a corrected formulation. To do this, Sinha started with the following trilevel formulation: [108, p. 594]

$$\begin{aligned}
&\max_{\overline{x}_1} f_1(x) = c_{11}x_1 + c_{12}x_2 + c_{13}x_3 \\
&\max_{\overline{x}_2} f_2(x) = c_{21}x_1 + c_{22}x_2 + c_{23}x_3 \\
&\max_{\overline{x}_3} f_3(x) = c_{31}x_1 + c_{32}x_2 + c_{33}x_3 \\
&\quad\text{s.t. } A_{i1}x_1 + A_{i2}x_2 + A_{i3}x_3 \leq b_i \qquad i = 1, 2, \ldots, m \\
&\qquad\quad x_1 \geq 0; x_2 \geq 0; x_3 \geq 0
\end{aligned} \tag{2.27}$$

and transformed (2.27) into the following equivalent problem:

$$
\begin{aligned}
&\max_{x_1} f_1(x) = c_{11}x_1 + c_{12}x_2 + c_{13}x_3 \\
&\quad \text{s.t.} \quad -c_{22(j)} + \sum_{i=1}^{m}(\mu_i + \lambda_i\omega_i)A_{i2(j)} \geq 0, && j = 1, 2, \ldots, n_2 \\
&\qquad \sum_{j=1}^{n_2}[-c_{22(j)} + \sum_{i-1}^{m}(\mu_i + \lambda_i\omega_i)A_{i2(j)}] = 0 \\
&\qquad -c_{23(j)} + \lambda'(-c_{33(j)} + \sum_{i=1}^{m}\omega_i A_{i3(j)}) + \sum_{i=1}^{m}(\mu_i + \lambda_i\omega_i)A_{i3(j)} \geq 0, && j = 1, 2, \ldots, n_3 \\
&\qquad \sum_{j=1}^{n_3}[-c_{23(j)} + \lambda'(-c_{33(j)} + \sum_{i=1}^{m}\omega_i A_{i3(j)}) + \sum_{i=1}^{m}(\mu_i + \lambda_i\omega_i)A_{i3(j)}]x_{3j} = 0 \\
&\qquad -c_{33(j)} + \sum_{i=1}^{m}\omega_i A_{i3(j)} \geq 0, && j = 1, 2, \ldots n_3 \\
&\qquad \sum_{j=1}^{n_3}\mu'_j[-c_{33(j)} + \sum_{i=1}^{m}\omega_i A_{i3(j)}] = 0 && (2.28) \\
&\qquad \sum_{j=1}^{n_3}[-c_{33(j)} + \sum_{i=1}^{m}\omega_i A_{i3(j)}]x_{3j} = 0 \\
&\qquad -\sum_{j=1}^{n_3}(\mu'_j - \lambda' x_{3j})A_{i3(j)} + \lambda_i(A_{i1}x_1 + A_{i2}x_2 + A_{i3}x_3 - b_i) \geq 0, && i = 1, 2, \ldots, m \\
&\qquad \sum_{i=1}^{m}[-\sum_{j=1}^{n_3}(\mu'_j - \lambda' x_{3j})A_{i3(j)} + \lambda_i(A_{i1}x_1 + A_{i2}x_2 + A_{i3}x_3 - b_i)]\omega_i = 0 \\
&\qquad A_{i1}x_1 + A_{i2}x_2 + A_{i3}x_3 \leq b_i && i = 1, 2, \ldots, m \\
&\qquad \mu_i(A_{i1}x_1 + A_{i2}x_2 + A_{i3}x_3 - b_i) = 0 && i = 1, 2, \ldots, m \\
&\qquad \omega_i(A_{i1}x_1 + A_{i2}x_2 + A_{i3}x_3 - b_i) = 0, && i = 1, 2, \ldots, m \\
&\qquad x_1, x_2, x_3, \omega, \mu, \mu' \geq 0; \\
&\qquad \lambda, \lambda' \text{ urs}
\end{aligned}
$$

Bard also used the KKT method to transform a trilevel model into a single level model. First, he replaces the lowest level problem with its KKT conditions. Before repeating the process, the complicating complementary slackness condition is dealt with by assigning a large penalty to it and moving it to the objective function. This results in the following formulation: [9, p. 714]

$$
\begin{aligned}
\max_{x^1} \quad & a^1x^1 + a^2x^2 + a^3x^3 \\
& \max_{x^2,x^3,u} b^2x^2 + b^3x^3 - ku(A^1x^1 + A^2x^2 + A^3x^3 - d) \\
& \text{s.t. } A^1x^1 + A^2x^2 + A^3x^3 \geq d \\
& uA^3 = -c^3 \\
& u \geq 0
\end{aligned}
\tag{2.29}
$$

where k is a sufficiently large finite constant, a, b, and c are constant row vectors of appropriate length; $d \in \mathbb{R}^m$; A^i is an $m \times n_i$ matrix, $i = 1, 2, 3$, and u is an m-dimensional row vector of dual variables. Bard then replaces this new inner problem with its KKT condition to develop the following:

$$
\begin{aligned}
\max_{x,u,\overline{u},v} \quad & a^1x^1 + a^2x^2 + a^3x^3 \\
\text{s.t.} \quad & A^1x^1 + A^2x^2 + A^3x^3 \geq d \\
& uA^3 = -c^3 \\
& \overline{u}[A^2, A^3] = -[b^2, b^3] \\
& ku - v + \overline{u} = 0 \\
& u(A^1x^1 + A^2x^2 + A^3x^3 - d) = 0 \\
& v(A^1x^1 + A^2x^2 + A^3x^3 - d) = 0 \\
& u \geq 0; v \geq 0
\end{aligned}
\tag{2.30}
$$

However, Bard points out that this formulation is necessary for optimal solutions to the original formulation, but not sufficient. Therefore, he develops a simplex-cutting plane algorithm to apply to this formulation to find the optimal solution. This algorithm is discussed later.

White looked at this formulation from Bard, and reformulated it. He shows that the following formulation is equivalent to the original trilevel formulation: [127, p. 186]

$$
\begin{aligned}
\max \quad & f^1 = f^{11}x^1 + f^{12}x^2 + f^{13}x^3 \\
\text{s.t.} \quad & A^1x^1 + A^2x^2 + A^3x^3 \leq b \\
& (v - \lambda u)(b - A^1x^1) - f^{22}x^2 - f^{23}x^3 = 0 \\
& -(v - \lambda u)A^2 \leq -f^{22} \\
& -vA^3 + \lambda f^{33} \leq -f^{23} \\
& -uA^3 \leq -f^33
\end{aligned}
\tag{2.31}
$$

where $x^k \in R_+^{n_k}$, $1 \leq k \leq 3$, $v \in R_+^m$, $u \in R_+^m$, $\lambda \in R_+$, $A^k \in \mathbb{R}^{m \times n_k}$, $k = 1, 2, 3$, $f^{kl} \in \mathbb{R}^{n_l}$

White further reformulates this problem, and proves that solutions to the following formulation are also solutions to (2.31).

$$
\begin{aligned}
\max \quad & f^1x \\
\text{s.t.} \quad & A^1x^1 + A^2x^2 + A^3x^3 \leq b \\
& (v - w)(b - A^1x^1) - f^{22}x^2 - f^{23}x^3 = 0 \\
& -(v - w)A^2 \leq -f^{22} \\
& -vA^3 + \lambda f^{33} \leq -f^{23} \\
& -wA^3 + \lambda f^{33} \leq -f^33
\end{aligned}
\tag{2.32}
$$

where $x^k \in R_+^{n_k}$, $1 \leq k \leq 3$, $v \in R_+^m$, $w \in R_+^m$, and $\lambda \in R_+$.

White continues by developing an algorithm that solves this formulation through a penalty formulation. White's approach is discussed in the next section.

2.3.6.1 Branch and Bound

In many of the reformulations of the multilevel problems, the difficulty in solving the reformulations is in the complementary slackness (or related) conditions. These constraints make the formulations nonlinear which (as the previous section pointed out) can be solved as nonlinear programs. In addition, there are two primary methods to deal with these complementary slackness equations to transform the nonlinear program into one that can be solved as a linear program. One is to attach a penalty to the these functions and move them to the objective function. This method is discussed in the next subsection. The other method is to use a branch-and-bound approach.

Moore and Bard reformulated the bilevel problem to the form shown in (2.25) and developed the branch-and-bound method. The basic idea of their algorithm is to

> suppress the complementarity term and solve the resulting linear program. At each iteration, a check is made to see if [the complementarity term] is satisfied. If so, the corresponding point is in the inducible region, and hence, is a potential solution to (2.25); if not, a branch and bound scheme is used to implicitly examine all combinations of complementary slackness. [11, p. 283]

Shi *et al.* showed how this technique could be easily modified for those problems where the leader's constraints were explicitly kept separate (i.e. formulation (2.26)). [107, p. 534]

Hansen *et al.* extended this approach by noting that at optimum at least one of the follower's constraints is tight. [67, p. 1196] Therefore, they associate a new boolean variable α_i with each constraint in the follower's problem. This variable is 1 if the constraint is tight and 0 otherwise. With this, they prove that for any rational solution, the tightness of the constraints in the follower's subproblem is such that: [67, p. 1198]

$$\begin{aligned} \sum_{i|B^2_{ij}>0} \alpha_i &\geq 1 && \text{if } d^2_j > 0 \\ \sum_{i|B^2_{ij}<0} \alpha_i + \alpha_{m_2+j} &\geq 1 && \text{if } d^2_j < 0 \end{aligned} \tag{2.33}$$

Branching is done by fixing some binary variables, α_i at 0 or at 1. If $\alpha_i = 1$, the ith constraint in the follower's subproblem becomes an equality. If $\alpha_i = 0$, the ith constraint becomes a strict inequality ($>$) and the ith variable in the dual of the follower's subproblem must be 0. Because of the difficulty of dealing with strict inequalities, the authors develop their branch and bound using the dual variable. With this, they develop the algorithm which is shown to outperform (on a set of test problems) the original branch-and-bound algorithm developed by Bard and Moore. [11, p. 1212]

As with many algorithms, a first step in these approaches is to solve a relaxed version. Specifically, the leader's problem is solved while ignoring the follower's objective function (but including the follower's constraints). This has important implications in the case of multiple optimal follower solutions. Since this approach starts with only the objective function of the leader, if the solution is found to be rational, it is the best from the leader's point of view among all rational solutions. [67, p. 1203] Therefore, this method chooses the solution among the follower's alternate optimal solutions that best suits the leader (i.e. the optimistic case). Hansen *et al.* note that it is easy to adapt this method to solve the pessimistic case. To do this, they suggest adding a secondary objective function to the follower's subproblem equal to $-d^1y$ which is only activated in cases of ties for the objective function d^2y. [67, p. 1203]

2.3.6.2 Penalty Method

As discussed in the previous section, another way to overcome the difficulty in complementary slackness equations is to move them to the objective function via a penalty term. For example, in (2.32), the complicating constraint is the equality constraint, so White used a penalty function to move this constraint to the objective function with the following resulting formulation:

$$
\begin{aligned}
\max \;& f^1x - K[(v-w)(b - A^1x^1) - f^{22}x^2 - f^{23}x^3] \\
\text{s.t. } & A^1x^1 + A^2x^2 + A^3x^3 \leq b \\
& -(v-w)A^2 \leq -f^{22} \\
& -vA^3 + \lambda f^{33} \leq -f^{23} \\
& -wA^3 + \lambda f^{33} \leq -f^33
\end{aligned}
\tag{2.34}
$$

where $u \in R^m_+$, $v \in R^m_+$, $w \in R^m_+$, $\lambda \in R_+$, $A^k \in \mathbb{R}^{m \times n_k}$, $k = 1, 2, 3$, $f^{kl} \in \mathbb{R}^{n_l}$

Note, if $\lambda = 0$ is the only solution to the formulation in (2.32), then $w = 0$ in (2.34). White develops an algorithm that finds a solution by increasing values of K until a solution is found which also satisfies the following equation: $(v-w)(b-A^1x^1) - f^{22}x^2 - f^{23}x^3 = 0$. [127, p. 192] However, these solutions are only necessary, not sufficient to solve this original formulation. Therefore, White recommends following the simplex-search developed by Bard and discussed in the hybrid methods below. Essentially, White is replacing step one of Bard's formulation with the formulation (2.34). White points out that the advantage of this method is that it does not require the calculation of an exact penalty parameter as Bard's method does. [127, p. 196]

A simple example of this formulation and algorithm is provided in [83].

2.3.6.3 Penalty on Duality Gap

Anandalingam and White made the following observation, "For a given value of x, the leader's decision vector, the follower is at his rational reaction set when the duality gap of the second-level problem becomes zero." [8, p. 1170] This leads to another method to transform a bilevel problem into a single level model. The problem is transformed by adding a term to the leader's objective function that minimizes the duality gap of the follower's problem. For example, using Israeli's notation, a BLPP can be transformed into the following formulation by penalizing the follower's duality gap: [70, p. 120]

$$\begin{aligned} \min_{x,y,w} \;& c_1^T x + c_2^T y - k(c_3^T y - w^T(b - Bx)) \\ \text{s.t. } & Dx \leq d \\ & Ay + Bx \leq b \\ & w^T A \geq c_3 \\ & x \geq 0, y \geq 0, w \geq 0 \end{aligned} \tag{2.35}$$

where $\mathbf{c}$, $\mathbf{y}$, $\mathbf{u} \in \mathbb{R}^n$, $\mathbf{b} \in \mathbb{R}^m$, $A \in \mathbb{R}^{m \times n}$ and $\mathbf{c}^T$ is the transpose of the column vector $\mathbf{c}$. For k sufficiently large, the duality gap must be zero. Israeli mentions two algorithms that have been developed to solve this formulation [see 8, p. 120]. Anandalingam also provides a short algorithm to solve this formulation. [7, p. 240] Notably, instead of finding the exact penalty, k, he suggests starting with a low value and increasing k in discrete steps.

In addition, White and Anandalingam noted that if b and d are almost negatively correlated (as they are in interdiction problems–see Israeli below) then the following steps for the initial value of w may be helpful. First, select $(\overline{x}, \overline{y}) \in \arg\max_x \;\; \min_y [ax - dy : (x, y) \in Z]$. Second, select $w^1(K) \in \arg\max[\hat{F}(\overline{x}, \overline{y}, w, K) : w \in W]$. [128, p. 406-407]

White and Anandalingam compared this method to the other penalty method (of penalizing complementary slackness of KKT conditions) as used by Bard in [9]. They show that the duality gap method provides a lower upper bound on the optimal solution used by the algorithm, so it is likely to converge more rapidly. [128, p. 413]

2.3.6.4 k^{th} Best

The k^{th} best algorithm repeatedly finds the "next best" solution to the leader's problem, until a solution is found in the inducible region. [16, p. 213] This is clearly a positive algorithm, as the follower's objective function is ignored while the leader's objective function is solved repeatedly until a solution would also be optimal for the follower's problem.

The k^{th} best algorithm has been proven to find the optimal solution by Wen and Bialas. They did this by considering the rational reaction set of each level. For example, let S^1 be the feasible set for the first level, S^2 be the rational reaction set for the second level, and S^3 be the rational reaction set for the third level. Wen and Bialas showed that "if x is an extreme point of S^3, the x is an extreme point of S^2 as well as S^1." [125, p. 369] Therefore, since the optimal point must occur at an extreme point, they show one can examine the extreme points of S^3 to find these solutions.

A related algorithm is vertex enumeration. This is based on the observation that if the set of rational solutions is nonempty, "at least one optimal solution of [the linear bilevel program] is obtained at an extreme point of the polytope defined by" the combined set of leader's and follower's constraints. [67, p. 1195]

2.3.6.5 Hybrid Methods

In addition to solving bilevel programs, many of these techniques have been combined to solve more difficult trilevel problems. For example, Anandalingam and Apprey suggested combining the k^{th} best method with the penalty function approach to solve a trilevel formulation. Specifically, the leader's problem is solved (without regard to follower's problems). Next, the follower's (bilevel) problem is solved using a penalty method. If the two solutions match, then the optimal solution has been found. If not, the next best solution to the leader's problem is found and the algorithm is repeated. [7, p. 241] Wen and Bialas use a very similar method; however, instead of using a penalty method to solve the bilevel follower's problem, they use a complementary pivot algorithm. [125, p. 370-371]

Bard proposed a simplex-cutting plane algorithm to solve a trilevel program. The first step, instead of solving the leader's problem, solves the formulation in (2.30). The second step is to fix the leader's variables, and solve the remaining bilevel problem. If the solutions match, then the optimal solution is found. Otherwise, "a simplex-type search is conducted" to arrive at a local optimum. A cut is then added which makes the incumbent solution infeasible (i.e. $f^1 = f^1 + \varepsilon$). Finally, the last step is designed to find a point of intersection between the cut and the level one inducible region. [9, p. 715]

2.3.6.6 Other Methods for BLPPs

Other methods have been developed, but are not discussed here for various reasons. For example, Bard developed a "grid search" algorithm, but it "only works for BLPPs whose solutions are known to be Pareto-optimal." [11, p. 289] The "parametric complementary pivot" approach is not guaranteed to converge and the leader's objective function coefficients associated with the follower's variables must be nonnegative. [11, p. 289] Ben-ayed and Blair present simple examples where both of these methods fail to find optimal solutions. [13]

2.3.7 Special Cases

A special case/simplification of multilevel programs occurs with interdiction problems. Interdiction problems occur when decision makers are assumed to have diametrically opposed objective functions. For example, if the follower's problem is a linear program, then the formulation would be a linear system interdiction.

2.3.7.1 Linear System Interdiction

$$\begin{aligned} z^* &= \min_{x \in X} \max_{y \in Y(x)} \mathbf{c}^T\mathbf{y} \\ X &= \{\mathbf{x} \in \{0,1\}^n | \mathbf{Rx} \leq \mathbf{r}\} \\ Y(x) &= \{\mathbf{y} | \mathbf{Ay} \leq \mathbf{b}, 0 \leq y \leq \mathbf{U}(\mathbf{1} - \mathbf{x})\} \end{aligned} \tag{2.36}$$

where $\mathbf{c}, \mathbf{y}, \mathbf{u} \in \mathbf{R}^n$, $\mathbf{c}, \mathbf{b} \in R^m$, $\mathbf{A} \in R^{m \times n}$.

Brown *et al.* showed how a formulation of this type can be solved in [26]. The basic idea is to transform the problem into a "cost attack" in which y's use of resources that x attacks is penalized in the objective function by a coefficient of P.

$$\begin{aligned} \max_{x \in X} \min_{y} \quad & \mathbf{cy} + \mathbf{x}^T PFy & & \\ \text{s.t.} \quad Ay &\geq b & & (v) \\ Fy &\leq u & & (w) \\ Cx &\leq d & & \\ y &\geq 0 & & \end{aligned} \tag{2.37}$$

The dual of the inner problem can be taken to form a single maximization.

$$\begin{aligned} \max_{x \in X, v, w} \quad & b^T v + uw \\ \text{s.t.} \quad A^v + F^T w &\geq c + F^T Px \\ Cx &\leq d \\ v \geq 0, w &\leq 0 \end{aligned} \tag{2.38}$$

As a single level problem with no complementary slackness conditions, this formulation is clearly easier to solve. A similar technique can be used on more difficult problems as well. The next section discusses formulations in which the followers problem is a mixed-integer formulation.

2.3.7.2 Mixed-Integer Linear System-Interdiction

If the follower's system can be modeled with a Mixed Integer Program (MIP), then the mixed-integer linear system-interdiction problem can be written as: [70, p. 46]

$$\begin{aligned} z^* &= \min_{x \in X} \max_{y \in Y(x)} \mathbf{c}^T \mathbf{y} \\ X &= \{\mathbf{x} \in \{0,1\}^n | \mathbf{Rx} \leq \mathbf{r}\} \\ Y(x) &= \{\mathbf{y} | \mathbf{Ay} \leq \mathbf{b}, 0 \leq y \leq \mathbf{U}(\mathbf{1} - \mathbf{x}), \mathbf{y} \in Y_{INT}\} \end{aligned} \tag{2.39}$$

where $\mathbf{c}, \mathbf{y}, \mathbf{u} \in \mathbf{R}^n$, $\mathbf{c}, \mathbf{b} \in R^m$, $\mathbf{A} \in R^{m \times n}$, and Y_{INT} represents integer (or binary) restrictions on none, some, or all of the variables $\mathbf{y}$. It is clear from the formulation that when activity j is interdicted ($x_j = 1$) then the upper bound on y_j is changed from u_j to 0.

As Israeli noted, to use Benders' partitioning, a reformulation is necessary.

> In Benders' decomposition the feasible region of the subproblem is fixed, independent of the first level variables ($\mathbf{x}$ in our case) while the objective function changes at every iteration. To obtain this situation in our case, we force the interdiction through a penalty term in the objective function, which will ensure that the use of an interdicted activity is not cost-effective. Then we can leave interdicted activities free in the subproblem (their upper bounds are not affected by $\mathbf{x}$), knowing for sure that these activities will not be used in an optimal solution. [70, p. 47]

The following formulation accomplishes this:

$$\begin{aligned} z^{**} &= \min_{x \in X} \max_{y \in Y} \mathbf{c}^T \mathbf{y} - \mathbf{x}^T \mathbf{V} \mathbf{y} \\ X &= \{\mathbf{x} \in \{0,1\}^n | \mathbf{Rx} \leq \mathbf{r}\} \\ Y &= \{\mathbf{y} | A\mathbf{y} \leq \mathbf{b}, 0 \leq \mathbf{y} \leq U, \mathbf{y} \in Y_{INT}\} \end{aligned} \tag{2.40}$$

With this, Israeli shows Benders' partitioning can be used and the master problem becomes [70, p. 49]

$$\begin{aligned} &\min_{x \in X, z} z \\ &\text{s.t. } z \geq \mathbf{c}^T \hat{\mathbf{y}} - \mathbf{x}^T \mathbf{V} \hat{\mathbf{y}} \qquad\qquad \hat{\mathbf{y}} \in \hat{\mathbf{Y}} \end{aligned} \tag{2.41}$$

where the subproblem is the inner maximization problem in formulation (2.41). Israeli also proves convergence and discusses methods to tighten the penalty term, V, [70, p. 50], or the master problem can be replaced with a set-covering problem [70, p. 51]

2.3.7.3 Network Disruption and Interdiction

A subset of linear programming interdiction is network interdiction. In this case, the follower's problem can be solved as a network problem (maximum flow, minimum cost flow, and so forth) Examinations of this special case are extensive; therefore, the entire next section is devoted to it. Just as different algorithms have been developed to exploit the structure of different network problems, interdiction algorithms have been developed which exploit the special structure as well. Therefore, the network interdiction section is broken into sections according to the underlying network type.

2.4 Layered/Interdependent networks

As discussed in Section 2.2.8, interdependencies are generally ignored when analyzing large networks such as infrastructure networks. These interdependencies can be especially critical in vulnerability analysis because they can potentially allow cascading effects across multiple networks. Therefore, it is vital that these interdependencies be considered in vulnerability identification and protection/fortification strategies. To do this, there is a need "to develop broad-based resource allocation procedures that capture these interactions vis-*à*-vis investment decision making." [139, p. 151]

As shown in Figure 2.1, multilayer network optimization is based on (single layer) network optimization. To discuss developments in this area, definitions of interdependencies are presented first. This is followed by two independent efforts from the literature that account for these interdependencies.

2.4.0.4 Interdependencies

An infrastructure dependency is defined as "a linkage or connection between two infrastructures, through which the state of one infrastructure influences or is correlated to the state of the other." [101, p. 14] Similarly, an interdependency is defined as "a bidirectional relationship between two infrastructures through which the state of each infrastructure influences or is correlated to the state of the other." [101, p. 14]

Rinaldi *et al.* suggest there are four types of interdependencies: physical, cyber, geographic, and logical. An infrastructure is physically dependent on another infrastructure if it requires material produced by another. Cyber dependency occurs when the state of an infrastructure is dependent on information sent through the information infrastructure. For example, energy and water infrastructures depend heavily on the use of Supervisory Control and Data Acquisition (SCADA) systems to conrol their functions. This type of interdependency is discussed extensively in *National Strategy to Secure Cyberspace* which was reviewed in Section 2.2.8.2.

Infrastructures are geographically interdependent if they are in close spacial proximity. For instance, if energy and telecommunications lines are attached to a bridge, both would be affected if the bridge is destroyed to affect transportation. Finally, logical dependencies are those relationships between infrastructures not included in the other categories [101, p. 14-16]

Wallace *et al.* independently developed a list of five types of interdependencies between differing networks: input, shared, exclusive-or, mutually dependent, and co-located. [122, p. 8] Input dependence results when one network requires input from another network. This is the same as physical interdependency defined by Rinaldi. Shared dependence occurs when some physical components are active in multiple networks. Exclusive-or dependence means that only one network (of a group of networks) can provision one service/resource at a time. A set of networks is said to be *mutually dependent* if the operation of one of the networks requires the other networks in the set. This is related to, but more broad than, the cyber dependency discussed by Rinaldi. Finally, physical components are said to be co-located if they are within a prescribed geographic region or area. Again, this is the same as geographic dependency defined by Rinaldi.

Therefore, based on Rinaldi and Wallace *et al.*, research has defined the following types of dependencies: physical/input, mutually dependent/cyber, geographic/co-located, shared, exclusive-or, and logical. With these types of dependencies defined, it may be possible to estimate the impact of targeted effects across all network layers. The level and reach of effects will depend on the degree of coupling, type of coupling, and adaptability to change between the layers. For example, tightly coupled systems have little slack in their connecting links, whereas loosely coupled systems can often accommodate failures by adapting. [101, p. 19]

While the focus of this research is on infrastructure interdependence, it is recognized that interdependencies do not just occur in infrastructure networks. Another interdependency which is examined in this research is that between social networks and infrastructure networks. For example, social networks require a stable, complete communication network to maximize efficiency and unity. [68, p. 1-2]

2.4.0.5 Restoration Model

Under normal operating conditions, Wallace *et al.* assumed infrastructure networks operate independently as a minimum cost network flow problems. However, disruptions to one or more of these networks create unmet demand which requires consideration of interdependencies. [122, p. 32] To model this, Wallace *et al.* developed a "restoration" model which models these interdependencies to help prioritize different demands for the same service.

Instead of a minimum cost objective function, the objective function in the restoration model is changed to minimize unmet demand. The constraints for those nodes who are not dependent on another network are largely unchanged in the conversion from a minimum cost formulation to a restoration formulation. The main exception is that a slack variable is added to demand constraints to capture any unmet demand.

For interdependent nodes/arcs, a new set of constraints (which mirror the independent nodes) is added. These new constraints have an additional binary variable, y, that enables modeling of the interdependency. Interdependencies are modeled as follows: if an interdependent node does not receive its demand from one (another infrastructure) network, it is not available for supply or transshipment in other networks. In other words, "constraints are included in this restoration model to shift the connector variable from 1 (operating) to 0 (failed) when the required demand isn't met at a dependent node." [122, p. 20] For example, if a telephone switching station does not receive its demand from the power network, it will not be able to function as a transshipment node for telephone calls.

This restoration model (modified to be consistent with previous notation) is as follows:

$$
\begin{array}{lll}
\text{Minimize} & \sum\limits_k s_k + \sum\limits_k b(1-y_k) & \\
\text{s.t.} & \sum\limits_{\{j:(i,j)\in A\}} x_{ij} - \sum\limits_{\{j:(j,i)\in A\}} x_{ji} \leq b & \forall \text{ (independent) Supply Node} \\
 & s_k + (\sum\limits_{\{j:(i,j)\in A\}} x_{ij} - \sum\limits_{\{j:(j,i)\in A\}} x_{ji}) = -b & \forall \text{ (independent) Demand Node} \\
 & \sum\limits_{\{j:(i,j)\in A\}} x_{ij} - \sum\limits_{\{j:(j,i)\in A\}} x_{ji} = 0 & \forall \text{ (independent) Transshipment Node} \\
 & l_{ij} \leq x_{ij} \leq u_{ij} \quad \forall (i,j) \in A & \\
 & \sum\limits_{\{j:(i,j)\in A\}} x_{ij} - \sum\limits_{\{j:(j,i)\in A\}} x_{ji} \leq b y_k & \forall \text{ (interdependent) Supply Node} \\
 & s_k + (\sum\limits_{\{j:(i,j)\in A\}} x_{ij} - \sum\limits_{\{j:(j,i)\in A\}} x_{ji}) = -b y_k & \forall \text{ (interdependent) Demand Node} \\
 & \sum\limits_{\{j:(i,j)\in A\}} x_{ij} - \sum\limits_{\{j:(j,i)\in A\}} x_{ji} = 0 & \forall \text{ (interdependent) Transshipment Node} \\
 & x_{ij} \leq u_{ij} y_k & \\
 & s_l \leq (1-y)b & \\
 & x_{ij} \leq u_i j & \\
 & x_{ij} \geq 0 & \\
 & y \in \{0,1\} & \\
 & s_{ij} \geq 0 &
\end{array}
\tag{2.42}
$$

The objective function minimizes the total shortfall (slack) plus unmet interdependent demand. Note, there is no consideration for partial slack at interdependent nodes because they control the operation of nodes in other subsystems. [122, p. 34] The constraints are as described in the previous two paragraphs.

2.4.0.6 Kennedy Model

Instead of only considering interdependencies after a disruption, Kennedy *et al.* took a different approach. [73] They started with single layer networks and modified the formulation slightly to allow for multiple layered modeling. This is done with two sets of variables. The first set contains the original (individual) network variables which model the infrastructure characteristics. The second set of variables captures interdependent elements. One advantage of this formulation is that it may be decomposed by variable type and solved to optimality using a Benders' partitioning based solution approach.

For example, each network k from the set of layered networks is modeled as a directed graph $G_k = [N_k, A_k]$ where N_k is the set of nodes and A_k is the set of arcs creating the network topology for the k^{th} network.

$$G_k = [N_k, A_k] \text{ where } A_k \subseteq \{(i,j) : i,j \in N_k\} \tag{2.43}$$

Further, each arc $(i,j) \in A_k$ has an associated cost per unit flow c_{ijk}, a maximum capacity u_{ijk}, and a minimum flow requirement l_{ijk}.

$$l_{ijk} \leq x_{ijk} \leq u_{ijk} \quad \forall \ (i,j) \in A_k, k = 1, \ldots, K. \tag{2.44}$$

Finally, each node has an associated integer $b_k(i)$; if $b_k(i) > 0$ then the node is a supply node, if $b_k(i) < 0$ then the node is a demand node, and if $b_k(i) = 0$ then the node is a transshipment node.

To model the interdependencies of the network, some new notation is introduced. As discussed previously, Wallace *et al.* developed a restoration model which used a binary variable y to capture the current state of interdependent nodes. To increase the flexibility of this model, the notation used in the Kennedy model allows some layers involving an interdependent edge to be available, while some are not–something that could not be captured in Wallace's model. Therefore, instead of assuming that an impacted edge affects all layers, Kennedy *et al.*'s notation introduces a variable, $w \in W$ which identifies a common effect option that impacts a subset of the interdependent edges (and/or nodes). [72]

For example, consider an arc that appears in three levels of a multilayered network: $x_{2,3,1}, x_{9,13,2}$, and $x_{24,23,3}$. One effect option, w_1, may affect all three levels ($x_{2,3,1}, x_{9,13,2}$, and $x_{24,23,3}$), while w_2 may affect only edges $x_{2,3,1}, x_{9,13,2}$. A variable which captures the effect that effect options have on an edge is also required. The variable $\delta_{(i,j,k),w}$ is also introduced.

$$\delta_{(i,j,k),w} \begin{cases} \neq 0, & \text{if } w \text{ has an effect on interdependent edge } (i,j); \\ = 0, & \text{otherwise.} \end{cases} \tag{2.45}$$

For a given scenario, an additional variable type was introduced which indicates which of the w options is selected/occured or so forth. For a given $w \in W$ define the decision to employ effect option w as

$$y_w = \begin{cases} 1, & \text{if option } w \text{ is selected/occurs;} \\ 0, & \text{otherwise..} \end{cases} \tag{2.46}$$

The impact of a common effect option has been represented by $\boldsymbol{\delta}_w$. The cost or benefit of this option and all other common effect options are defined by the following vector of costs (or benefits):

$$\mathbb{C}^T = \left[\mathbb{C}_1, \mathbb{C}_2, \ldots, \mathbb{C}_{|W|}\right]. \tag{2.47}$$

Costs or benefits associated with the individual networks are defined by

$$\mathbf{C}^T = \left[\mathbf{c}_1^T, \mathbf{c}_2^T, \ldots, \mathbf{c}_k^T\right]. \tag{2.48}$$

The actual elements of $\mathbb{C}^T$ and $\mathbf{C}^T$ may be positive, negative or zero, as dictated by the situation being modeled. While the objective function has been expressed as a minimization, it may be stated as a maximization (with any necessary variations) as the particulars of the problem under consideration require. For example, if $\mathbb{C}^T$ were the benefits from flow in a given arc and $\mathbf{C}^T$ were the benefits of some upgrade option y_i, the model would select the best upgrade packages for the entire layered system.

Minimum Cost Network Flow

Given the definitions from the previous sections, the minimum cost network flow formulation across multiple layers is:

$$
\begin{aligned}
\min \quad & \mathbf{C}^T\mathbf{x} + \mathbb{C}^T\mathbf{y} \\
s.t. \quad &
\end{aligned}
$$

$$
\begin{bmatrix}
A_1 & & & \\
\cdots & & & \\
I_1 & & & \\
 & A_2 & & \\
 & \cdots & & \\
 & I_2 & & \\
 & & \ddots & \\
 & & & A_K \\
 & & & \cdots \\
 & & & I_K \\
 & & &
\end{bmatrix}
\mathbf{x} - \mathbf{D}\mathbf{y} +
\begin{bmatrix}
\mathbf{0} \\ \mathbf{s}_1 \\ \mathbf{0} \\ \mathbf{s}_2 \\ \vdots \\ \mathbf{0} \\ \mathbf{s}_K
\end{bmatrix}
=
\begin{bmatrix}
\mathbf{b}_1 \\ \boldsymbol{\mu}_1 \\ \mathbf{b}_2 \\ \boldsymbol{\mu}_2 \\ \vdots \\ \mathbf{b}_K \\ \boldsymbol{\mu}_K
\end{bmatrix}
\tag{2.49}
$$

$$
\begin{aligned}
s_k &\geq 0 \quad \forall k \\
x_k &\geq 0 \quad \forall k \\
\mathbf{y} &\in \{0, 1\}
\end{aligned}
$$

where $\mathbf{x}$ represents a column vector formed by the $\mathbf{x}_k$ vectors, $\mathbf{D}$ represents the matrix formed by the columns, $\boldsymbol{\delta}_w$, associated with the effects options, and $\mathbf{y}$ represents a column vector formed by the previously defined decision variables y_w. This includes flow-balance and bounding constraints for each single-layer problem.

In [73, p. 15], Kennedy *et al.* show how this formulation can be solved via Benders' partitioning.

Minimum Cut – Maximum Flow Formulation

In this example, from Kennedy *et al.*, given layered networks with interdependent arcs, the objective is to minimize the combined cost of cutting all networks using individual and shared elements in the overall cut set. A cost is associated with each of these arcs, and the goal is to find the minimum cost set of arcs which determines this partition. [2]

For example, suppose the desired objectives/effects are to prevent military transportation, electricity flow, and land-based telecommunications to a specified island. Further, assume that only one bridge connects to the island, and all power and telecommunication lines are tied beneath and across the bridge. One obvious solution is to bomb the bridge, severing the bridge itself as well as the power and telecommunication lines. However, if the costs are too high (i.e. civilian casualties resulting from an inability to exit the island), then another form of attack may be more appropriate. Perhaps a less costly attack would be to target the bridge with an electronic attack, disrupting power and telecommunications, and kinetically bombing the military transportation hub on the island. While these objectives are often considered in isolation, the Kennedy model incorporates such options.

Consider π_{i_k}, the dual variable associated with the conservation of flow equation for node i of network k. In addition, let ν_{ijk} be the dual variable associated with the capacity constraint of arc(i,j) of network k. A minimum cut formulation for each of the k networks would be

$$\begin{aligned}
&\min \sum_{(i,j)\in A_k} c_{ijk}\nu_{ijk} \\
&s.t. \\
&\quad \pi_{i_k} - \pi_{j_k} - \nu_{ijk} \geq 0 \ \forall \ \ (i,j) \in A_k \qquad (2.50) \\
&\quad 0 \leq \pi_{i_k} \leq 1 \ \forall i \quad \in N \\
&\quad 0 \leq \nu_{ijk} \leq 1 \forall (i,j) \quad \in A \\
&\quad \pi_{t_k} = 1; \pi_{s_k} = 0
\end{aligned}$$

where c_{ijk} is the flow capacity along arc(i,j) of network k. It follows then that the objective function, $\sum c_{ijk}\nu_{ijk}$, is the relative cost of cutting the flow of goods in network k. To incorporate the interdependencies previously described, let

$$y_w = \begin{cases} 1, & \text{if } w \in W_{\mathcal{I}} \text{ is chosen} \\ 0, & \text{otherwise.} \end{cases} \tag{2.51}$$

The "cost," $\mathbb{C}_w$, then represents the relative cost of cutting the interdependent arcs associated with using option w. The commonality model then becomes

$$\begin{aligned} \min \quad & \sum_{k\in K} \sum_{(i,j)\in A_k} c_{ijk}\nu_{ijk} + \sum_{w\in W} \mathbb{C}_w y_w \\ s.t. \quad & \\ & \pi_{i_k} - \pi_{j_k} + \nu_{ijk} + \delta_{(i,j,k),w} y_w \geq 0 \ \forall \ (i,j) \in A_k, k \in K \\ & \pi_{t_k} - \pi_{s_k} \geq 1 \ \forall \ s,t \in N_k, k \in K \\ & \sum_{w\in W_\gamma} y_w \leq 1 \ \forall \ \gamma \in \Gamma \end{aligned} \tag{2.52}$$

where

$$\delta_{(i,j,k),w} = \begin{cases} 1, & \text{if arc}(i,j) \text{ of network } k \text{ is affected by option } w \in W \\ 0, & \text{otherwise} \end{cases} \tag{2.53}$$

Note, however, this formulation is not solved directly. Since valuable information is gained from the dual variables of the Benders' subproblem, when the y variables have been fixed, the dual of Benders' subproblem is solved instead. This dual is a maximum flow formulation given as

$$
\begin{array}{lll}
\max & \displaystyle\sum_{k \in K} \sum_{(i,j) \in A_k} x_{t_k, s_k} + \sum_{w \in W} \mathbb{C}_w \overline{y}_w & \\
s.t. & & \\
& \displaystyle\sum_{j:(i,j) \in A_k} x_{ijk} - \sum_{j:(j,i) \in A_k} x_{jik} = 0 \quad \forall\ k \in K & \\
& x_{ijk} \leq c_{ijk} - c_{ijk} \delta_{(i,j,k),w} \overline{y}_w \qquad \forall\ (i,j) \in A_k, k \in K, w \in W & \\
& x_{ijk} \geq 0 \qquad \forall\ (i,j) \in A_k, k \in K &
\end{array}
\tag{2.54}
$$

As with the minimum cost network flow formulation, Kennedy *et al.* demonstrate this formulation can be effectively solved with Benders' partitioning. As is typical with Benders' partitioning, information between the master problem and subproblem is passed back and forth until some stopping criteria is reached.

2.5 Network Interdiction and Fortification

Network interdiction is a special case of bilevel programming. These types of problems are also referred to as "attacker-defender" problems. In this case, the defender is a network operator who seeks to protect and operate the network while the attacker seeks to maximally disrupt this network. As shown in Figure 2.1, network interdiction is based on network optimization and multilevel programming.

The specific formulation of network interdiction problems depends on the network under consideration. As such, several specific interdiction formulations are discussed: shortest path, maximum flow, minimum cost, multicommodity flow, system flow, and facility location. First, some assumptions and introductory material are discussed.

The key assumptions in attacker-defender problems are: [28, p. 532]

- The attacker's and defender's actions are sequential
- The attacker has a perfect model of how the defender will (or should) optimally operate the system, even after an attack
- The attacker will manipulate that system to his best advantage.

These assumptions mirror those for general multilevel programs discussed in Section 2.3. Brown *et al.* state that the last two assumptions are “strong, but prudent” because the defender can do no worse should the attacker have a less-than-perfect model or fails to implement a perfect attack. [28, p. 532] In effect, this is a conservative strategy to protect against worst-case attacks.

This research examines some of these assumptions to determine their impact on the problems/solutions. Previous research has shown the importance of these assumptions. For example, Brown *et al.* have found that “secrecy and deception are valuable.” [27, p. 41] In addition, “[o]ne insight from these military and diplomatic exercises is that the use of deception and secrecy can contribute significantly to the successful defense of our critical infrastructure, or to successful attacks on our adversary’s infrastructure.” [28, p. 542]

Some additional assumptions are also made to simplify the presentation and provide a foundation. First, it is assumed that interdiction is binary. In other words, if a node/arc is interdicted, it is completely destroyed. Similarly for fortification, if a node/arc is fortified, it can not be destroyed. Extensions to allow partial interdiction/fortification exist in the literature, but are not discussed here.

In addition, it is assumed that the interdictor/attacker has insufficient resources to disconnect s from t. Otherwise, a simpler minimum cut algorithm can be used. Finally, it is assumed that only the edges are interdicted and that edges are directed. Again, this is not a limiting assumption as “extensions of our techniques to handle undirected networks and/or node interdiction are also straightforward.” [70, p. 19]

“In this problem, the defender and attacker play a zero sum game, i.e., the defender tries to minimize the same objective function that the attacker tries to maximize.” [138, p. 712] Since the inner problem can be solved as a network problem, this nested “max-min” structure has an exploitable structure. By taking the dual of the inner minimization (network) problem, the problem is converted into a maximization problem. This allows one to formulate a single model in which the leader's decision variables and the follower's (dual) decision variables are simultaneously optimized. [138, p. 714] This technique was first developed and used by Fulkerson in 1975. [58]

Finally, some algorithms have also been developed which extend the network interdiction problem to a protection problem. In this case, the network user (follower) knows that an attack is pending and uses its (limited) resources to protect a portion of his assets. Such a network protection problem is a trilevel problem (i.e. min-max-min). In cases where algorithms have been developed to exploit the special structure of the network, these trilevel algorithms are included in the discussion as well.

2.5.1 Shortest Path Interdiction & Fortification

This subsection begins with a mathematical formulation of the shortest path problem. Ahuja *et al.* [2, p. 94] provides a linear programming algorithm to solve any network for the shortest path between any two nodes. The general formulation (discussed in Section 2.2.2) is as follows:

$$\min \sum_{(ij)\in A} c_{ij}x_{ij} \tag{2.55}$$

$$\sum_{j:(i,j)\in A} x_{ij} - \sum_{j:(j,i)\in A} x_{ji} = \begin{cases} 1, & \text{for } i = s; \\ 0, & 0\ \forall i \in N\ \{s,t\}; \\ -1, & \text{for } i = t. \end{cases}$$

$$x_{ij} \geq 0 \ \ \forall (i,j) \in A$$

Suppose someone wished to "attack" a given network in an effort to maximize a shortest path between two nodes. If an attacker has sufficient resources, he/she could find the minimum cost (or any other) cut set required to disconnect the two nodes in a network. This would result in disjoint networks in which no path would exist from the pair of arbitrarily chosen nodes. However, in some cases, this may be impractical because either the attacker does not possess sufficient resources, a target is inaccessible, or completely "cutting" an arc may be impractical.

In these cases, a shortest path interdiction formulation can be used. The following formulation maximizes the shortest $s - t$ path length in a directed network by interdicting arcs. This approach is based on the work of Israeli and Wood [71].

$$z^* = \max_{\delta \in \Delta} \min \sum_{(ij) \in A} (c_{ij} + \delta_{ij} d_{ij}) x_{ij} \tag{2.56}$$

$$\sum_{j:(i,j)\in A} x_{ij} - \sum_{j:(j,i)\in A} x_{ji} = \begin{cases} 1, & \text{for } i = s; \\ 0, & 0 \ \forall i \in N\{s,t\}; \\ -1, & \text{for } i = t. \end{cases}$$

$$x_{ij} \geq 0 \quad \forall k \in A$$

where $\Delta = \{\delta \in \{0,1\}^{|A|} | r^T \delta \leq D\}$; c_{ij} is the nominal integer length of arc (i,j); d_{ij} is the added integer delay if arc ij is interdicted, $x_{ij} = 1$ if arc (i,j) is traversed in the shortest path (= 0 otherwise); $\delta_{ij} = 1$ if arc (i,j) is interdicted; r_{ij} is the resource required to interdict arc (i,j), and D is the total amount of interdiction resource available.

Fulkerson was the first to show that this formulation can be written as a single level MIP. [58, 59] Israeli developed this idea and solution methodology. By first fixing x, taking the dual of the inner minimization problem, making "a few simple modifications" and releasing x, the following formulation results: [70, p. 19]

$$\begin{aligned} z^* &= \max_{\delta,\pi} \pi_t - \pi_s \\ \text{s.t. } & \pi_j - \pi_i - d_{ij}\delta_{ij} \leq c_{ij} \\ & \pi_s = 0 \\ & \delta \in \Delta \end{aligned} \tag{2.57}$$

This formulation could be solved directly as a mixed-integer program (using branch-and-bound for example). However, Israeli found that "when possible delays d_{ij} are large, the linear program (LP) relaxation of the model is weak and this results in excessive enumeration and unsatisfactory computation times." [70, p. 20] Instead, he developed a decomposition algorithm.

<u>*Solution Technique*</u>

Israeli and Wood [71] provided a means to solve the shortest-path network interdiction problem with Benders' decomposition. In their formulation, the subproblem has fixed values of δ resulting in a shortest path formulation.

$$\min \sum_{(i,j)\in A} (c_{i,j} + \hat{\delta_{ij}} d_{ij}) x_{i,j} \tag{2.58}$$

$$\sum_{j:(i,j)\in A} x_{ij} - \sum_{j:(j,i)\in A} x_{ji} = \begin{cases} 1, & \text{for } i = s; \\ 0, & 0\ \forall i \in N\ \{s,t\}; \\ -1, & \text{for } i = t. \end{cases}$$

$$x_{ij} \geq 0 \quad \forall (i,j) \in A$$

where c_{ij} is the nominal integer length of arc (i,j), d_{ij} is the added integer delay if arc (i,j) is interdicted, $\hat{\delta} = 1$ if arc (i,j) is interdicted, and $x_{ij} = 1$ if arc (i,j) is traversed in the shortest path ($= 0$ otherwise). In the first iteration, $\hat{\delta}$ is set to zero. In subsequent iterations, this value is passed from the master problem.

The master problem uses the flow values (x_{ij}) from the subproblem as fixed values to determine interdiction strategies δ.

$$\max_{\delta\in\Delta} z \tag{2.59}$$

$$z \leq c_{i,j}^T \hat{x_{ij}} + \delta_{ij} D \hat{x_{ij}} \quad \forall x_{ij} \in X$$

where c_{ij} is the nominal integer length of arc (i, j), $\delta = 1$ if arc (i, j) is interdicted, and $x_{ij} = 1$ if arc (i, j) is traversed in the shortest path, $\hat{x_{ij}}$ is fixed values transferred from the subproblem, and D is the total amount of interdiction resource available.

This process is repeated until the objective values from the master problem and subproblem are equal (within a user defined tolerance).

As noted previously, large delays (d values) can lead to weak convergence. Israeli noted some supervalid inequality (SVI)s that may speed convergence. [71, p. 100-102] In addition, a modified covering decomposition algorithm was developed which ignores the delay (particularly useful when interdiction completely destroys arcs). This approach replaces the master problem in the above formulation with the following: [71, p. 103]

$$\begin{aligned} &\text{Find } x \in X \\ &\quad \text{s.t. } \tilde{y}^T x \geq 1 \ \ \forall \ \ (\hat{x}, \hat{y}) \in \hat{X}\hat{Y} \end{aligned} \tag{2.60}$$

where $\tilde{y} \equiv (\text{diag}(1-\hat{x}))\hat{y}$. The constraint $\tilde{y}^T x \geq 1$ is the covering constraint such that "if the interdictor wishes to force the follower to traverse a path other than $\hat{y}$ then a new interdiction plan $\hat{x}'$ must interdict some arc that is not interdicted by $\hat{x}$ but is used by the follower in response to $\hat{x}$." [71, p. 103] Therefore, with this replacement, the algorithm generates new interdiction plans until the master problem becomes infeasible. At this point, the best found solution is provably optimal. [71, p. 103]

Fortification

Suppose the follower knew an attack was pending and had a limited budget to fortify some elements of his network. Let the set of feasible defense plans be given by $G = \{\mathbf{g} \in \{0,1\}^{|A|} | \mathbf{Hg} \leq \mathbf{h}\}$ where $g_k = 1$ means arc k can not be interdicted. With this, the following formulation finds the optimal defense strategy for the network user: [70, p. 61]

$$\begin{aligned} &\min_{g \in G} \max_{x \in X(g)} \min_{y \in Y(x)} \mathbf{c}^T \mathbf{y} \\ &\qquad \textit{where} \\ &\qquad\quad G = \{\mathbf{g} \in \{0,1\}^{|A|} | \mathbf{Hg} \leq \mathbf{h}\} \\ &\qquad X(g) = \{\mathbf{x} \in \{0,1\}^{|A|} | \mathbf{Rx} \leq \mathbf{r}, 0 \leq \mathbf{x} \leq 1 - \mathbf{g}\} \\ &\qquad Y(x) = \{y | \mathbf{y} \text{ is an incidence vector } s - t \text{ path that is feasible with respect to } x\} \end{aligned} \tag{2.61}$$

Israeli suggests a solution technique which involves a nested decomposition algorithm where the master problem is given by [70, p. 62]

$$\begin{aligned} \min_{g \in G} \quad & z \\ \text{s.t.} \quad & z \geq \mathbf{c}^T \mathbf{y}(\mathbf{x}(\hat{\mathbf{g}})) - \mathbf{g}^T \mathbf{V}_D \mathbf{x}(\hat{\mathbf{g}}) \qquad \forall \;\; \hat{\mathbf{g}} \in \hat{G} \end{aligned} \tag{2.62}$$

This master problem suggests a new defense plan $\mathbf{g}$ and updates $\underline{z}_D$. The subproblem (which is a maximimal shortest path) solves the system-interdiction problem associated with g, adds the solution to $\hat{G}$, updates $\overline{z}_D$, and is given by [70, p. 63]

$$\begin{aligned}
&\max_{x \in X(\hat{g})} \min_{y \in Y(x)} \mathbf{c}^T \mathbf{y} \\
&\quad \text{where} \\
&\quad X(\hat{g}) = \{x \in \{0,1\}^{|A|} | \mathbf{R}\mathbf{x} \leq \mathbf{r}, 0 \leq \mathbf{x} \leq 1 - \hat{\mathbf{g}}\} \\
&\quad Y(x) = \{\mathbf{y} | \mathbf{y} \text{ is an incidence vector for an } s - t \text{ path that is feasible with respect to } \mathbf{x}\}
\end{aligned} \tag{2.63}$$

where $\hat{G}$ is a subset of all possible defense plans. [70, p. 63] Israeli continues to discuss methods for determining a small, but valid, penalty vector v_D. [70, p. 64]

2.5.2 Maximum Flow Interdiction

As discussed in Section 2.2.3, the maximum flow problem can be formulated as follows:

$$\begin{aligned}
\max \quad & x_{ts} \\
\text{s.t.} \quad & \sum_j x_{sj} - \sum_j x_{js} - x_{ts} = 0 \\
& \sum_j x_{ij} - \sum_j x_{ji} = 0, \quad \forall \quad (i,j) \in A \\
& \sum_j x_{tj} - \sum_j x_{jt} + x_{ts} = 0 \\
& 0 \leq x_{ij} \leq u_{ij}, \quad \forall \quad (i,j) \in A \\
& x_{ts} \geq 0
\end{aligned} \tag{2.64}$$

where x_{ts} is an artificial arc from t to s.

Interdiction of this maximum flow can be written as the following bilevel program: [137, p. 5]

$$
\begin{aligned}
\min_{\gamma\in\Gamma}\max_{x} \quad & x_{ts} & & \\
\text{s.t.} \quad & \sum_j x_{sj} - \sum_j x_{js} - x_{ts} = 0 & & \\
& \sum_j x_{ij} - \sum_j x_{ji} = 0, & \forall \;\; i \in \text{N-\{s,t\}} & \qquad (2.65)\\
& \sum_j x_{tj} - \sum_j x_{jt} + x_{ts} = 0 & & \\
& x_{ij} - u_{ij}(1-\gamma_{ij}) \leq 0, & \forall \;\; (i,j) \in A & \\
& x_{ts} \geq 0, & \forall (i,j) \in A \cup \{(t,s)\} &
\end{aligned}
$$

where $\Gamma \equiv \{\gamma_{ij} | \gamma_{ij} \in \{0,1\} \forall (i,j) \in A, \sum_{(i,j)\in A} r_{ij}\gamma_{ij} \leq R\}$.

Wood shows this can be transformed into the following (single level) integer program:

$$
\begin{aligned}
\min \quad & \sum_{(i,j)\in A} u_{ij}\beta_{ij} & & \\
\text{s.t.} \quad & \alpha_i - \alpha_j + \beta_{ij} + \gamma_{ij} \geq 0, & \forall (i,j) \in A & \\
& \alpha_t - \alpha_s \geq 1 & & \qquad (2.66)\\
& \sum_{(i,j)\in A} r_{ij}\gamma_{ij} \leq R & & \\
& \alpha_i \in \{0,1\}, & \forall i \in N & \\
& \beta_{ij}, \gamma_{ij} \in \{0,1\}, & \forall (i,j) \in A &
\end{aligned}
$$

where $\alpha_i = 1$ for i on the t side of the cut and $\alpha_i = 0$ for i on the s side of the cut; γ_{ij} is 1 if (i,j) is a forward arc across the cut which is to be broken; β_{ij} is 1 if (i,j) is a forward arc across the cut, but is not to be broken; and all other β_{ij} and γ_{ij} are zero.

This formulation could be solved directly, but as Uygun notes, the resource constraint $\sum_{(i,j)\in A} r_{ij}\gamma_{ij} \leq R$ makes this problem difficult to solve. [120, p. 9] To combat this difficulty, Uygun uses Lagrangian relaxation to move this constraint to the objective function (building on work done the year before by Bingol [20]).

However, he discovered that "problematic" R values exist which leads to large gaps from optimality, for which he had to resort to time consuming branch-and-bound to solve. [120, p. 42] In addition, this procedure has difficulty finding the optimal solution when many of the arcs have the same capacity. [50, p. 51]

To combat the problems with problematic Rs, Cormican uses a Benders' partitioning technique to find an exact solution to the single objective case.

Solution Technique

Cormican took this formulation and developed a solution technique using Bender's partitioning. Cormican started with the following equivalent formulation of the bilevel maximum flow interdiction problem: [46, p. 16]

$$
\begin{aligned}
\min_{\gamma\in\Gamma}\max_{x} \quad & x_{ts} - \sum_{(i,j)\in A} \gamma_{ij}x_{ij} & & \\
\text{s.t.} \quad & \sum_{j} x_{sj} - \sum_{j} x_{js} - x_{ts} = 0 & & \\
& \sum_{j} x_{ij} - \sum_{j} x_{ji} = 0, & \forall \;\; i \in \text{N-\{s,t\}} & \qquad (2.67)\\
& \sum_{j} x_{tj} - \sum_{j} x_{jt} + x_{ts} = 0 & & \\
& 0 \le x_{ij} \le u_{ij}, & \forall (i,j) \in A &
\end{aligned}
$$

where $\Gamma \equiv \{\gamma_{ij} | \gamma_{ij} \in \{0,1\} \forall (i,j) \in A, \sum\limits_{(i,j)\in A} r_{ij}\gamma_{ij} \le R\}$.

Using Bender's partitioning, Cormican shows this can be broken into the following master and subproblem: [46, p. 17-18]

Master Problem: MASTER(X)

$$\begin{aligned} &\min_{\gamma \in \Gamma} z \\ &\text{s.t. } z \geq x_{ts}^k - \sum_{(i,j)\in A} x_{ij}^k \gamma_{ij} \\ &\quad \gamma_{ij} \in \{0,1\}, \qquad \forall (i,j) \in A \end{aligned} \tag{2.68}$$

Subproblem: SUB(γ)

$$\begin{aligned} &\max_x x_{ts} - \sum_{(i,j)\in A} x_{ij}\hat{\gamma}_i j \\ &\text{s.t. } \sum_j x_{sj} - \sum_j x_{js} - x_{ts} = 0 \\ &\quad \sum_j x_{ij} - \sum_j x_{ji} = 0, \qquad \forall \; i \in \text{N-\{s,t\}} \\ &\quad \sum_j x_{tj} - \sum_j x_{jt} + x_{ts} = 0 \\ &\quad 0 \leq x_{ij} \leq u_{ij}, \qquad \forall (i,j) \in A \end{aligned} \tag{2.69}$$

Finally, Cormican provided the following algorithm which finds the optimal solution: [46, p. 18-19]

Benders' Decomposition Algorithm for Network Interdiction
Input: Network $G = (N, A)$, arc capacities u_{ij}, arc interdiction costs r_{ij}, interdiction budget R, special nodes s and t, convergence tolerance *toler*.
Output: Interdiction vector γ^*, which is the solution within ($100 \times$ *toler*)% of optimality.

1. Solve maximum flow problem MF for flow values x^1; Let $X' = \{x^1\}$; let $k = 2$; Let $UB = z(MF)$
2. Solve MASTER(X') for γ; Let $LB = z(MASTER(X'))$
3. Solve $SUB(\hat{\gamma})$ for x^k; Let $X' = X \cup \{x^k\}$; If $z(SUB) \leq UB$ then let $UB = z(SUB(\hat{\gamma}))$ and $\gamma^* = \hat{\gamma}$
4. If $UB - LB \leq LP \times toler$ then stop: Interdiction set γ^* is a solution to the network interdiction problem with objective function value within ($100 \times$ *toler*)% of the optimal objective value.

5. Let $k = k + 1$; Go to step 2.

2.5.2.1 Extension

An important extension to this formulation is given by Royset and Wood who developed a bi-objective maximum flow network interdiction problem. Here, instead of cost being a fixed constraint, a second objective was to minimize the cost. In other words, they seek to find Pareto-optimal solutions with respect to minimizing post-interdiction maximum flow and minimizing "total interdiction cost." This is an important consideration because often one must plan for various resource availabilities. In addition, one may wish to consider tradeoffs between cost, risk, and effectiveness. [102, p. 175] Royset and Wood note that the "the efficient frontier can be identified by solving [Formulation (2.66)] over a sufficiently wide range of R-values." [102, p. 180] However, they use a weighted-sums scalarization of the objectives using Lagrangian relaxation. Lagrangian relaxation was used to move the resource constraint to the objective function and find the efficient frontier in [102].

Other extensions to the maximum flow interdiction model have also been developed. Wood extended the maximum flow interdiction problem to account for cardinality constraints [137, p. 8], partial arc interdiction [137, p. 8], multiple sources and sinks [137, p. 9], undirected networks [137, p. 9], multiple resources [137, p. 10], and multiple commodities [137, p. 11]. Wood also discusses using valid inequalities [137, p. 12], and a cutset based reformulation [137, p. 14] to speed solution times.

In another extension, Whiteman modified the maximum flow interdiction problem to solve multiple sets of objectives, each with their own flow capacity goals. Here a simultaneous cut is required for each objective requiring that most variables and constants pick up an additional index, k, to track each objective. He notes that "γ_{ij} variables do not require the additional index, since if an arc is targeted for any objective, it is always targeted." [133, p. 3] This formulation is as follows: [133, p. 4]

$$
\begin{aligned}
\min \quad & \sum_{arcs} \gamma_{ij} & & \\
\text{s.t.} \quad & \alpha_{ik} - \alpha_{jk} + \beta_{ijk} + \gamma_{ij} \geq 0 & \forall \ \text{arcs}, k & \\
& \sum_{arcs} u_{ij}\beta_{ijk} \leq G_k & \forall \ \ k & \qquad (2.70) \\
& \alpha_{ik} = 0 & \forall \ \text{sinks}, k & \\
& \alpha_{ik} = 1 & \forall \ \text{sources}, k & \\
& \alpha_{ik}, \beta_{ijk}, \gamma_{ij} \in \{0, 1\} & &
\end{aligned}
$$

The objective function minimizes the number of arcs broken, the first constraint requires any arcs spanning the cut under consideration ($|\alpha_i - \alpha_j| = 1$) be designated as targeted ($\gamma_{ij} = 1$) or untargeted ($\beta_{ij} = 1$). It should be noted that a cost coefficient c could be multiplied against γ in the objective function to model the desirability of targeting a facility γ. [132, p. 21] The second constraint requires that spanning arcs which are not broken have a combined capacity of no more than the flow capacity goal. The last set of constraints specify which nodes are designated as sources and sinks which must always be on their respective sides of the cut. One way to specify an arc as untargetable would be to add a set of constraints of the form: $\gamma_{ij} = 0$ for all untargetable arcs. To determine an efficient frontier, these objectives may be varied one objective at a time. For example, sometimes a "modest reduction in interdiction level requirements can significantly reduce weapon requirements." [133, p. 8]

Whiteman proposed additional extensions to his model. The first was risk assessment. The model above assumes a targeted arc will be completely destroyed with probability one. This first extension was to allow monte carlo simulations where this probability can be between 0 and 1. After a set number of iterations, a probability of interdiction at specified levels is obtained. [133, p. 9]

A second extension occurs when there is a fixed number of weapons to apply to a given objective. This formulation exchanges the resource constraint and the objective function. Therefore, the flow capacity will be minimized with the given number of weapons. For the multiobjective version, "a relative weight must be assigned" to each objective. [133, p. 10] A third extension allows for variable weapon requirements by allowing non-unity coefficients on the γ_{ij} variables. Similarly, variable target effects can be modeled. Partial interruptions can be modeled by modifying the resource constraint as follows: $\sum_{arcs}(u_{ij}\beta_{ij} + \nu_{ij}\gamma_{ij}) \leq G$ where ν_{ij} is the capacity remaining when arc ij is targeted. [133, p. 11]

2.5.3 Minimum Cost Network Flow

As discussed in Section 2.2.4, a minimum cost network flow program seeks to determine the least expensive way to route commodities through a network. [2, p. 357-397] Let $\mathbf{c}$ be a vector of component operating costs (and/or penalties), and $\mathbf{y}$ be the system operating decisions or activities, and $\mathbf{y} \in Y$ be constraints on that operation. The result is the following formulation: [28, p. 533]

$$\begin{aligned} \min_{y \geq 0} \quad & \mathbf{cy} \\ \text{s.t.} \quad & A\mathbf{y} = \mathbf{b} \\ & F\mathbf{y} \leq \mathbf{u} \end{aligned} \tag{2.71}$$

where the first constraint corresponds to a general operating constraint, and the second set of constraints correspond to capacity limitations for asset $k \in K$. It is assumed that an attacker would seek to maximize this minimum cost. Therefore, if it is assumed an attack on asset k causes the loss of all its capacity u_k, then the following minimum cost network interdiction formulation results: [28, p. 533]

$$\begin{aligned} \max_{x \in X} \min_{y \geq 0} \quad & \mathbf{cy} \\ \text{s.t.} \quad & A\mathbf{y} = \mathbf{b} \\ & F\mathbf{y} \leq U(1 - \mathbf{x}) \end{aligned} \tag{2.72}$$

where $U = diag(\mathbf{u})$ and x is a vector of attack decisions.

As an alternate formulation, the use of "attacked" capacities can be penalized to make them "uneconomical." [28, p. 533] Let $P = diag(\mathbf{p})$, then

$$\begin{aligned} \max_{x \in X} \min_{y \geq 0} \quad & (\mathbf{c} + \mathbf{x}^T PF)\mathbf{y} \\ \text{s.t.} \quad & A\mathbf{y} = \mathbf{b} \\ & F\mathbf{y} \leq U(1 - x) \end{aligned} \tag{AD1}$$

With this new formulation, the dual of the inner minimization can be taken which results in the following formulation: [28, p. 533]

$$\begin{aligned} \max_{\beta \leq 0, \theta, x} \quad & \mathbf{b}^T \theta + \mathbf{u}\beta \\ \text{s.t.} \quad & A^T \theta + F^T \beta - F^T P \mathbf{x} \leq \mathbf{c}^T \\ & \mathbf{x} \in X \end{aligned} \tag{2.73}$$

where θ is the dual variable associated with the first constraint and β is the dual variable associated with the second set of constraints.

Brown *et al.* note that this formulation can be solved directly, or via Benders' partitioning. In fact, they point out that the first step of Benders' would be to fix $\mathbf{x}$ and take the dual which would result in the formulation above. [28, p. 533]

A more explicit formulation is given by Nesbitt in [90]. The primary difference is that Nesbitt's formulation included constructs for potential missing information. Nesbitt's technique for missing information is discussed in Section 2.5.7.

Let N be a set of nodes, A be the set of arcs, $(i, j) \in A$ be the arc directed from node i to node j, (i', j') be commodity from source i' bound for destination node j', $c_{i,j}$ be the cost per unit flow of commodity over arc (i, j), $b_{i',j'}$ be the net amount of flow originating at i' with destination j', $u_{i,j}$ be the upper bound on commodity flow on arc $(i, j) \in A$, and $x_{i,j,i',j'}$ be the flow on arc $(i, j) \in A$ of commodity from node i' to node j'.

$$\begin{aligned}
\min_{x} \quad & \sum_{(i,j)\in V} c_{i,j} \sum_{i' \neq j'} x_{i,j,i',j'} && && (2.74)\\
\text{s.t.} \quad & \sum_{j:(n,j)\in A} x_{n,j,i',j'} - \sum_{i:(i,n)\in A} x_{i,n,i',j'} = \begin{cases} b_{i',j'}, & n = i'; \\ b_{j',i'}, & n = j'; \\ 0, & \text{otherwise.} \end{cases} && \forall \;\; n, i', j' \in N, i' \neq j' \\
& 0 \leq \sum_{i' \neq j'} x_{i,j,i',j'} \leq u_{i,j} && \forall \;\; (i,j) \in A \\
& x_{i,j,i',j'} \geq 0 && \forall \;\; (i,j) \in A
\end{aligned}$$

In this case, a network interdiction formulation seeks to maximize this minimum cost network flow. This formulation requires a couple of additional parameters/variables. Let $d_{i,j}$ be the cost imposed per unit flow on arc $(i, j) \in A$ when an arc is attacked, *maxattacks* be the maximum number of components that the attacker can target, and $y_{i,j}$ be 1 if the arc $(i, j) \in A$ is attacked and 0 otherwise.

$$\begin{aligned}
\max_{Y \in \Upsilon} \min_{x,w} \quad & \sum_{(i,j)\in V} (c_{i,j} + d_{i,j} y_{i,j}) \sum_{i' \neq j'} x_{i,j,i',j'} \\
\text{s.t.} \quad & \sum_{j:(n,j)\in A} x_{n,j,i',j'} - \sum_{i:(i,n)\in A} x_{i,n,i',j'} = \begin{cases} b_{i',j'}, & n = i'; \\ b_{j',i'}, & n = j'; \\ 0, & \text{otherwise.} \end{cases} && \forall \;\; n, i', j' \in N, i' \neq j' \\
& 0 \leq \sum_{i' \neq j'} x_{i,j,i',j'} \leq u_{i,j} && \forall \;\; (i,j) \in A \\
& && (2.75) \\
& x_{i,j,i',j'} \geq 0 && \forall \;\; (i,j) \in A \\
& && (2.76)
\end{aligned}$$

where

$$\Upsilon = \begin{cases} \sum\limits_{(i,j) \in V} y_{i,j} \leq maxattacks, & \\ y_{i,j} \in \{0,1\}, & \forall (i,j) \in A. \end{cases} \tag{2.77}$$

This inner objective calculates the cost of flowing commodities through the network given that some arcs have been attacked ($y_{i,j} = 1$) and therefore have a higher cost. The first constraint enforces flow balance while the second constraint ensures flow is less than capacity.

By taking the dual of the inner minimization problem, the following equivalent formulation can be found:

$$\begin{aligned} \max_{Y \in \Upsilon, \alpha, \gamma} & \sum_{i' \neq j'} b_{i',j'}(\alpha_{i',i',j'} - \alpha_{j',i',j'}) - \sum_{(i,j) \in A} u_{i,j}\gamma_{i,j} & & \\ \text{s.t. } & \alpha_{i,i',j'} - \alpha_{j,i',j'} - \gamma_{i,j} \leq c_{i,j} + d_{i,j}Y_{i,j} & \forall \ \ (i,j) \in V, i', j' \in N & \\ & \gamma_{i,j} \geq 0 & \forall \ \ (i,j) \in A & \end{aligned} \tag{2.78}$$

where

$$\Upsilon = \begin{cases} \sum\limits_{(i,j) \in V} Y_{i,j} \leq maxattacks, & \\ Y_{i,j} \in \{0,1\}, & \forall (i,j) \in A. \end{cases} \tag{2.79}$$

and $\alpha_{i,i',j'}$ is the flow balance constraint dual variable, and $\gamma_{i,j}$ is the flow capacity constraint dual variable.

2.5.4 Multicommodity Network Interdiction

Two variants of multicommodity network flow are considered: multicommodity maximum flow and multicommodity minimum cost flow.

2.5.4.1 Multicommodity Maximum Flow

Akgun [3] developed a formulation in the context of maximizing flow among 3 or more groups of nodes. In other words, the flow from a given node (in one group) to all nodes in other groups was maximized. Since this is being done for all nodes, this can be modeled as a multicommodity problem.

This formulation is as follows: [3, p. 29]

$$
\begin{aligned}
\min_{x\in X} \max_{y} \quad & \sum_k \Big(\sum_{(i,j)\in A: i\in N'_k} y_{ijk} + \sum_{(i,j)\in A: j\in N'_k} y_{jik} \Big) & \\
\text{s.t.} \quad & \sum_{(i,j)\in A^+} y_{ijk} - \sum_{(j,i)\in A^+} y_{jik} = 0 & \forall \ \ k = 1, \ldots, K, i \in N - N' \\
& \sum_k (y_{ijk} + y_{jik}) \leq u_{ij}(1 - x_{ij}) & \forall \ \ (i,j) \in A \\
& y_{ijk\geq 0}, \quad y_{jik} \geq 0 & \forall \ \ k = 1, \ldots, K, (i,j) \in A
\end{aligned} \tag{2.80}
$$

$$
\begin{aligned}
& y_{ijk} \equiv 0 & \forall \ \ k = 1, \ldots, K, i \in N' - N'_k, (i,j) \in A \\
& y_{jik} \equiv 0 & \forall \ \ k = 1, \ldots, K, j \in N' - N'_k, (i,j) \in A \\
& y_{ijk} \equiv 0 & \forall \ \ k = 1, \ldots, K, j \in N'_k, (i,j) \in A \\
& y_{jik} \equiv 0 & \forall \ \ k = 1, \ldots, K, i \in N'_k, (i,j) \in A
\end{aligned} \tag{2.81}
$$

where $X = \{x \in \{0,1\}^{|A|} : \sum\limits_{(i,j)\in A} r_{ij}x_{ij} \leq R\}$, $i, j \in N$ are nodes in an undirected network, $(i,j) \in A$ are undirected arcs in the network, N_k^+ are the subset of "special nodes," $N' = \bigcup\limits_{k=1}^{K} N'_k$, u_{ij} is the nominal capacity of arc (i,j), r_{ij} is the interdiction resource required to interdict arc (i,j), R is the total interdiction resource, y_{ijk} is the amount of flow on arc (i,j) whose source is in N'_k, x_{ij} is 1 if arc (i,j) is interdicted and 0 otherwise. The objective is to minimize the maximum flow between the subsets N'_k.

> For $x \equiv 0$, the inner maximization is simply the multi-commodity maximum flow model . . . [which] models the enemy's potential transfers of material among the subsets N'_k using K single-commodity flow models linked by joint capacity constraints . . . each subset N'_k is a single-

commodity maximum flow model in which the N'_k are treated as sources and nodes in $\bigcup\limits_{k' \neq k} N'_{k'}$ are treated as sinks. [3, p. 30]

The "convexified" version of this formulation can be converted into a mixed-integer formulation by fixing x temporarily, taking the dual of the inner maximization and then releasing x which results in the following: [3, p. 33]

$$\begin{aligned}
\min_{x,\alpha,\beta} \quad & \sum_{(i,j)\in A} u_{ij}\beta_{ij} & & \\
\text{s.t.} \quad & -\alpha_{ik} + \alpha_{jk} + \beta_{ij} + x_{ij} \geq \delta_{ijk} & \forall\ \ k = 1,\ldots,K, (i,j) \in A & \\
& -\alpha_{jk} + \alpha_{ik} + \beta_{ij} + x_{ij} \geq \delta_{jik} & \forall\ \ k = 1,\ldots,K, (i,j) \in A & \qquad (2.82) \\
& \sum_{(i,j)\in A} r_{ij}x_{ij} \leq R & & \\
& x_{ij} \in \{0,1\}, \beta_{ij} \geq 0 & \forall\ \ (i,j) \in A & \\
& \alpha_{ik} \text{ unrestricted} & \forall\ \ k = 1,\ldots,K, i \in N & \\
& & & \qquad (2.83)
\end{aligned}$$

where δ_{ijk} is 1 if $i \in N'_k$ and 0 otherwise, x_{ij} is 1 if arc (i,j) is interdicted and 0 otherwise, α_{ik} are dual variables associated with flow-balance constraints, and β_{ij} are dual variables associated with the capacity constraints.

This formulation was solved directly and computation times were stated. [3, p. 36] Using Benders' partitioning and/or integer programming based cuts was left for future work which has not yet been found in the literature.

2.5.4.2 Multicommodity Minimum Cost Flow

Similarly, Lim and Smith developed a model for multicommodity minimum cost flow interdiction. [82, p. 20] Let K be the set of commodities where each commodity $k \in K$ may have multiple supply nodes, S^k, and demand nodes, D^k. The maximum supply at node l for commodity k is denoted s^k_l. Similarly, demand is denoted d^k_l. Lim and Smith used h to index the arcs, such that flow costs are represented by r^k_h with flow y^k_h being the flow assigned across arc h. Finally, x_h is 1 if arc h is interdicted, and 0 otherwise. With this notation, the following multicommoditiy minimum cost flow interdiction problem is given by

$$\begin{aligned}
\min_{x\in X}\max & \sum_{h\in A}\sum_{k\in K} r_h^k y_h^k \\
\text{s.t.} \ & \sum_{i\in FS(l)} y_i^k - \sum_{j\in RS(l)} y_j^k = 0 && \forall \ \ k\in K, \forall \ \ l\in N\backslash(s^k\cup D^k) \\
& \sum_{i\in FS(l)} y_i^k - \sum_{j\in RS(l)} y_j^k = s_l^k && \forall \ \ k\in K, \forall \ \ l\in S^k \\
& \sum_{i\in FS(l)} y_i^k - \sum_{j\in RS(l)} y_j^k = -d_l^k && \forall \ \ k\in K, \forall \ \ l\in D^k \\
& \sum_{k\in K} y_h^k \le u_h(1-x_h) && \forall \ \ h\in A \\
& y_h^k \ge 0 && \forall \ \ h\in A, \forall \ \ k\in K
\end{aligned} \tag{2.84}$$

where $x \in X_I \equiv \{x : \sum_{h\in A} b_h x_h \le B, x_h \in \{0,1\} \ \ \forall \ \ h \in A\}$, B is the interdiction budget, and b_n is the cost of interdicting arc n.

Since the inner problem is a multicommodity flow problem, the dual can be taken to form a mixed-integer bilinear formulation. Let π_l^k be the dual variable associated with the first three constraints, and ϕ_h be the dual associated with the forth constraint. This leads to the following formulation: [82, p. 22]

$$\begin{aligned}
\min & \sum_{k\in K}\sum_{l\in S^k} s_l^k\pi_l^k - \sum_{k\in K}\sum_{l\in D^k} d_l^k\pi_l^k + \sum_{h\in A} u_h\phi_h - \sum_{h\in A} u_h w_h \\
\text{s.t.} \ & x\in X_I \\
& (\pi,\phi)\in\theta \\
& w_h - \phi_h \le 0 && \forall h\in A \\
& w_h - \overline{\phi}_h x_h \le 0 && \forall h\in A
\end{aligned} \tag{2.85}$$

where $w = x_h\phi_h$ and θ is the dual feasible region. The last two constraints result from a linearization technique. An alternative (equivalent) formulation which incorporates their new cutting plane techniques is given in [110, p. 17-18].

Alternatively, the original formulation could be reformulated as a penalty function. [82, p. 23] The only new variable is M_h^k which is some large constant value. This leads to the following formulation: [82, p. 25]

$$\begin{aligned} \min & \sum_{k \in K} \sum_{l \in S^k} s_l^k \pi_l^k - \sum_{k \in K} \sum_{l \in D^k} d_l^k \pi_l^k + \sum_{h \in A} u_h \phi_h & \\ \text{s.t. } & \pi_{f(h)} - \pi_{t(h)}^k + \phi_h + M_h^k x_h \geq r_h^k & \forall k \in K, \forall h \in A \\ & \pi_l^k \text{ unrestricted} & \forall k \in K, \forall l \in N, \phi_h \geq 0 \forall h \in A \\ & x \in X_I & \end{aligned} \tag{2.86}$$

where the first constraint has the interpretation of "deactivating a dual constraint corresponding to an arc $h \in A$ when h has been interdicted." [82, p. 25] In addition, no linearization techniques are needed.

2.5.5 System Flow Interdiction

"While maximum flow approaches seek to identify interdiction schemes that reduce the capacity of a particular O-D [origin-destination] pair, system flow approaches focus on the interaction between all O-D pairs." [89, p. 106] In other words, instead of considering a single origin-destination pair, system flow interdiction considers total flow between all origin-destination pairs. This is similar to the multicommodity maximum flow interdiction problem discussed in Section 2.5.4.1.

Let k be the index of paths, j be the index of facilities, o be the index of origins, d be the index of destinations, N_{od} be the sets of paths enabling O-D flow, f_{od} be the flow observed between O-D, p be the number of facilities to remove, ϕ_k be the set of facilities along path k, x_j be 1 if facility j is interdicted and 0 otherwise, y_k be 1 if path k remains unaffected by interdiction and 0 otherwise, and z_{od} be 1 if no flow is possible between O-D and 0 otherwise. The formulation developed by Murray is then given by

$$
\begin{aligned}
\max or \min \sum_{o} \sum_{d} & f_{od} z_{od} & & \\
\text{s.t.} \quad \sum_{k \in N_{od}} & y_k + z_{od} \geq 1 & \forall \ o, d & \\
& z_{od} \leq 1 - y_k & \forall \ o, d, k \in N_{od}, k & \\
& y_k \geq 1 - \sum_{j \in \phi_k} x_j & \forall \ k & \\
& y_k \leq (1 - x_j) & \forall \ k, j \in \phi_k & \qquad (2.87) \\
& \sum_{j} x_j = p & & \\
& x_j \in \{0, 1\} & \forall \ j & \\
& y_k \in \{0, 1\} & \forall \ k & \\
& z_{od} \in \{0, 1\} & \forall \ o, d &
\end{aligned}
$$

Clearly, this single level model relies on the variable f_{od} which is the *observed* flow between O-D. This seems to require a steady state flow network to study. This is borne out in their example: "flow observed in network routers was collected . . . " [89, p. 109] In addition, it requires the enumeration of all paths connecting an O-D pair.

This concept can be extended using the bilevel approach explicitly modeling either maximum flow (without requiring an "observed" flow) or shortest paths. Specifically, this extension is discussed in the methodology chapter in the context of social networks.

2.5.6 Facility Location Interdiction and Fortification

The final network interdiction type problem that is discussed in this literature review is the facility location interdiction problem. For example, the p-median problem selects the location of p facilities in such a manner that the total weighted distance of supplying each demand from its closest facility is minimized. [40, p. 494] The facility location interdiction problem is the antithesis of this problem which

Church *et al.* define as follows: "Of the p different locations of supply, find the subset of r facilities, which when removed, yields the highest level of weighted distance." [40, p. 494] In some respects, this is similar to running the network synthesis problem in reverse.

To formulate this problem, let i be the index representing places of demand, j be the index representing existing facility locations, s_j be 1 if a facility located at j is interdicted and 0 otherwise, F be the set of existing facilities j, x_{ij} be 1 if demand i is assigned to a facility at j and 0 otherwise, a_i be a measure of demand needed at demand i, d_{ij} be the shortest distance between the supply/service facility at j and demand i, r be the number of facilities to be interdicted, and T_{ij} be the set of existing sites that are as far or farther than j from i ($= \{k \in F | k \neq j$ and $d_{ik} > d_{ij}\}$). The formulation is then: [40, p. 495]

$$
\begin{aligned}
\max Z = & \sum_i \sum_{j \in F} a_i d_{ij} x_{ij} & & \\
\text{s.t. } & \sum_{j \in F} x_{ij} = 1 & & \forall \ i, \ \forall \ j \in F \\
& \sum_{j \in F} s_j = r & & \\
& \sum_{k \in T_{ij}} x_{ik} \leq s_j & & \forall \ i, \ \forall \ j \in F \\
& x_{ij} \in \{0,1\} & & \forall i, \ \forall j \in F \\
& s_j \in \{0,1\} & & \forall j \in F
\end{aligned}
\tag{2.88}
$$

where the objective seeks to maximize the weighted distance impact due to interdiction of r-facilities, the first constraint assigns a facility to each demand after interdiction, the second constraint restricts the number of interdictions to r, the third constraint assigns a demand i to the closest remaining facility to i.

Scaparra and Church claim that "it is in principle a trilevel (defender-attacker-user) problem reduced to a bilevel minmax problem" [106, p. 1906] This is done by making the following observation: "At least one optimal fortification of q of the p facilities includes at least one site of the interdiction set U" where U is the set of optimal interdictions assuming no facilities have been fortified. [39, p. 133] With this observation, the heuristic developed by Brown *et al.* ([28]) as discussed in Section 2.5.7) can be formalized, and used recursively to find optimal solutions using branch and bound.

Fortification

Snyder *et al.* also reviewed methods to protect against worst-case losses for network models. [113, p. 251-252] The three-level shortest path interdiction problem with fortification is formulated as: [113, p. 251]

$$\begin{array}{lll} \min\limits_{z \in F} \max\limits_{s \in D} \min\limits_{Y} & \sum\limits_{(i,j) \in A} (d_{ij} + p_{ij} s_{ij}) y_{ij} & \\ \text{s.t.} & \sum\limits_{(j,i) \in A} y_{ji} - \sum\limits_{(i,j) \in A} y_{ij} = b_j & \forall \ j \in V \\ & s_{ij} \leq 1 - z_{ij} & \forall \ (i,j) \in A \\ & y_{ij} \geq 0 & \forall \ (i,j) \in A \end{array} \tag{2.89}$$

where $F = \{z \in \{0,1\}^n | \sum\limits_{(i,j) \in A} Z_{ij} = Q\}$ and $D = \{s \in \{0,1\}^n | \sum\limits_{(i,j) \in A} s_{ij} = R\}$, $b_0 = 1$, $b_d = -1$, $b_j = 0$ for all other nodes j in V. The objective function computes the minimum-cost path after the worst-case interdiction of R unprotected facilities which includes penalties associated with interdicted arcs (protected arcs cannot be interdicted).

Similarly, the maximum flow interdiction problem with fortification is formulated as [113, p. 252]

$$
\begin{array}{llr}
\max\limits_{z\in F}\min\limits_{s\in D}\max\limits_{y\geq 0} \;\; W & & \\
\text{s.t.} \;\; \sum\limits_{(j,i)\in A} y_{ji} - \sum\limits_{(i,j)\in A} y_{ij} = W & j = o & \\
\sum\limits_{(j,i)\in A} y_{ji} - \sum\limits_{(i,j)\in A} = 0 & \forall \;\; j \in V\backslash\{o,d\} & \\
\sum\limits_{(j,i)\in A} y_{ji} - \sum\limits_{(i,j)\in A} y_{ij} = -W & j = d & (2.90) \\
y_{ij} \leq k_{ij}(1 - r_{ij}s_{ij}) & \forall \;\; (i,j) \in A & \\
s_{ij} \leq 1 - z_{ij} & \forall \;\; (i,j) \in A & \\
y_{ij} \geq 0 & \forall \;\; (i,j) \in A &
\end{array}
$$

where the objective function maximizes the total flow W after the worst-case interdiction of the capacities of R arcs.

Snyder *et al.* state that these trilevel problems can be reduced to bilevel programs by taking the dual of the inner network flow problems. Citing a work by Scaparra and Cappanera ([105]), they state "the resulting bilevel problem can be solved efficiently through an implicit enumeration scheme that incorporates network optimization techniques." [113, p. 252]

2.5.7 Protection and Trilevel Models

2.5.7.1 Trilevel

Although trilevel problems are complex and difficult to solve exactly, heuristics have also proven problematic. As Yao *et al.* explained, "the presence of multiple optima and ε-optimal solutions render heuristic determination of high-quality defense plans difficult and/or time consuming." [138, p. 712]

To identify the optimal defense plan given a limited budget (for defense), bilevel/network interdiction problems can be extended to trilevel problems. While defender-attack methods defend against the single most damaging attack, the "trilevel model provides a robust defense strategy against the n most damaging attack plans." [138, p. 712]

To formulate this problem, let $\mathbf{w}$ be a binary vector of defensive decisions (e.g. $w_k = 1$ if asset k is protected and 0 otherwise, where $\mathbf{w} \in W$ denotes the binary restrictions on $\mathbf{w}$ together with budgetary constraints. The following formulation results: [28, p. 536]

$$\begin{aligned} \min_{\mathbf{w}\in W} \max_{x\in X} \min_{y\in Y} \quad & \mathbf{cy} \\ \text{s.t.} \quad & A\mathbf{y} = \mathbf{b} \\ & 0 \leq \mathbf{y} \leq U(1 - (\mathbf{x} - \mathbf{w})^+) \end{aligned} \tag{2.91}$$

where it is assumed that if asset k is defended, then the asset is invulnerable. In addition, $h^+ \equiv \max\{0, h\}$ so that $(\mathbf{x} - \mathbf{w})^+$ is the "net attack plan" where plan $\mathbf{x}$ is implemented against plan $\mathbf{w}$.

As with bilevel problems, the dual of the inner minimization can be taken resulting in: [28, p. 536]

$$\begin{aligned} \min_{\mathbf{w}\in W} \max_{\mathbf{x}\in X} \max_{\alpha,\beta} \quad & \alpha\mathbf{b}^T + \beta U(1 - (\mathbf{x} - \mathbf{w})^+) \\ \text{s.t.} \quad & \alpha A + \beta I \leq \mathbf{c} \\ & \beta \leq \mathbf{0} \end{aligned} \tag{2.92}$$

or

$$\begin{aligned} \min_{\mathbf{w}\in W, z} \quad & z \\ \text{s.t.} \quad & z \geq \hat{\alpha}_l \mathbf{b}^T + \hat{\beta}_l U(1 - (\hat{\mathbf{x}}_l - \mathbf{w})^+) \\ & l \in L \end{aligned} \tag{2.93}$$

where L enumerates all combinations of maximal attack plans $\hat{\mathbf{x}} \in X$ and extreme points $(\hat{\alpha}, \hat{\beta})$ from the first two constraint sets.

Brown *et al.* point out that this formulation could be solved with Benders' partitioning where the subproblems will be instances of attacker-defender problems and the master problem will require constructs to handle the "+" operator. [28, p. 536] In addition, they have done research into using "super-valid inequalities" to help find solutions. In fact, as they point out, one could use a master problem "whose constraints consist only of super-valid inequalities, and with an objective function that represents any of the lower bounding functions." [28, p. 536]

Brown considered a slight variation of this formulation. Instead of only considering fortification, his formulation also allowed network synthesis (i.e. additional arcs to increase maximum flow). [29, p. 22] This formulation was not solved exactly, but by a heuristic approach through an indirect decomposition.

2.5.7.2 Defender-Attacker Models

Some defense problems lend themselves to bilevel, defender-attacker models. In these formulations, "the defender becomes the leader in this new Stackelberg game, so we essentially reverse the meanings of $\mathbf{x}$ and $\mathbf{y}$, and make the following definitions." [28, p. 535] Let k be an asset the defender wishes to defend and the attacker may wish to attack, c_k be the value to the attacker of attacking undefended asset k, p_k be the reduction in value if that asset is defended, where x_k is 1 if the defender defends k and 0 otherwise, and y_k is 1 if the defender's kth asset is attacked and 0 otherwise. Here, $\mathbf{x} \in X$ are the resource constraints and binary restrictions on the defender's defense plan ($X = \{\mathbf{x} \in \{0,1\}^n | G\mathbf{x} \leq \mathbf{f}\}$). In addition, $\mathbf{y} \in Y$ are resource constraints and binary restrictions on the attacker's attack plan ($Y = \{y \in \{0,1\}^n | A\mathbf{y} = b\}$). Given these definitions, the formulation is

$$\min_{\mathbf{x} \in X} \max_{\mathbf{y} \in Y} (\mathbf{c} + \mathbf{x}^T P)\mathbf{y} \tag{2.94}$$

This technique works if evaluating the attackers objective does not require solution of an optimization problem. Therefore, this model can not incorporate a detailed operational model of the defender's system. [28, p. 535] In addition, in general, these problems are difficult to solve because the inner maximization problem is not a linear program. Brown *et al.* offer three cases to resolve this problem. First, assume that a continuous attack represents a reasonable approximation of reality. Second, if the linear programming relaxation yields binary solutions, then conversion to a mixed-integer linear program is possible. This is the case when it corresponds to a network-flow problem. Finally, if neither of the first cases permits, then the restriction must be included. [28, p. 535]

This was done with border patrol [28, p. 537], the DC-metro system [27, p. 129], Los Angeles airport security [27, p. 132], and supply chains [28, p. 541] In addition, this technique was used on a variation of formulation (2.94) as well. Specifically, instead of the "defender" simply fortifying specific assets, the defender also has the option to add additional links to increase maximum flow. [29, p. 14]

2.5.7.3 Heuristic

Brown suggests using the defender–attacker as a heuristic to find good solutions to the trilevel protection problem. The idea is to find the optimal solution to the attacker-defender problem to determine vulnerabilities. Using the limited fortification budget, the next step is to protect some/all these vulnerabilities identified in the defender–attacker routine. With these assets protected, the defender–attacker problem is re-run. Iterations between the defender–attacker problem and the protection problem are continued until some stopping criteria is reached. This was done with the Strategic Petroleum Reserve [28, p. 537], the electric grid [28, p. 539], and oil pipelines in Saudi Arabia [27, p. 126]

As mentioned in Section 2.5.6, this method could be formalized and used to find an optimal solution. Using branch-and-bound, subsets of protection assets could be separately examined to find the optimal solution.

2.6 Benders' Partitioning

A final area included in this literature review is Benders' Partitioning. Although this is not included in the foundational research, Benders' partitioning is reviewed as a method to solve some of the formulations developed.

The networks considered for targeting in this research can be extremely large. For example, infrastructure networks span entire continents, and contain tens of thousands of nodes and links (or more). One method for dealing with "large" problems is to decompose them into smaller more manageable parts. These subproblems are then managed by a "master problem" which combines the subproblem solutions to find an optimal solution to the overall problem. Ideally, the large problem decomposes into subproblems which have structure which can be exploited to quickly generate solutions.

Specifically, Benders, in his 1962 paper, presented a procedure for solving problems involving models where the variables can be partitioned into two distinct sets. [14] Benders' partitioning takes a problem of the form $\max\{c^T + f(y) \leq b; x \in \mathcal{R}, y \in S\}$ (where S may be a polyhedron, a set with discrete variables, a set with nonlinearities, or so on); and partitions it into two mutually exclusive subsets which are solved separately. One set consists only of the continuous variables with all the other variables fixed at some value. Either this problem is solved directly, or the dual is solved, and the dual variable values are passed to the "master" problem. This master problem then solves for the new y's which are again fixed for the other partition(s).

Visually, an original formulation for networks may be as presented in Figure 2.5. Benders' partitioning will result in a master problem and a subproblem similar to that depicted in Figure 2.6.

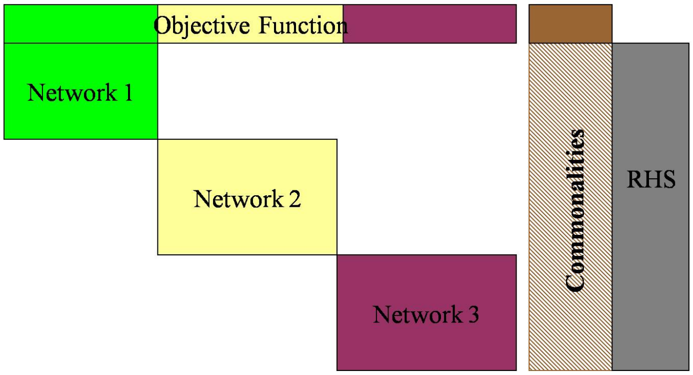


Figure 2.5: Original Formulation

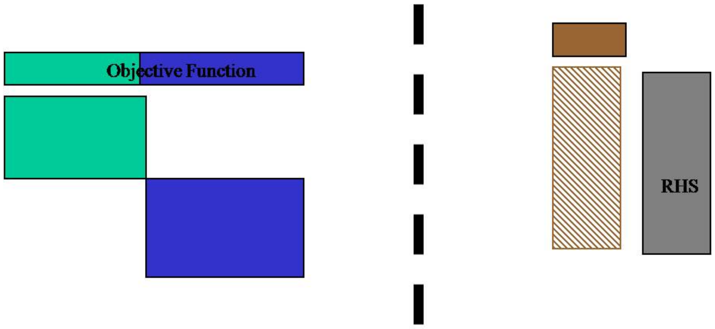


Figure 2.6: Benders' Partition

2.7 Summary and Layout

This chapter reviewed the relevant current literature in network research. This discussion of multiobjective programming, network optimization, and multilevel programming provides the foundation for extensions which are developed in the next chapter. Two areas which already build on this foundation (multilevel network optimization and network interdiction) are also included in the review as they are further developed and extended in the next chapter.

The overall goal of this research was to formulate and develop techniques to analyze a new problem: the synthesis/vulnerability of layered networks. In addition, this is combined with multiobjective techniques to consider cost, and consider the potential impact of coalitions/multiple optimal solutions. Based on historical techniques, solution approaches will be developed for this new problem set.

This research follows the guidance and principles of the Committee on Algorithms (COAL) concerning reporting results, claims, and conclusions. Two important relevant principles are "the results presented must be sufficient to justify the claims" and "there must be sufficient detail to allow reproducibility of the results." [21, p. 414] To accomplish this, the claims are specifically stated. To facilitate progress toward a unified analysis of layered networks, this research provides the following theoretical contributions: a new formulation and initial solution methodology for

- multilayer network interdiction
- cost & robustness tradeoffs in layered networks
- human network interdiction
- impact determination of coalitions/multiple optimal solutions in interdictions

To demonstrate the formulation and solution techniques developed as part of this research, they are implemented in GAMS.

As shown in Figure 1.2, several formulational developments were pursued leading to the goal of a unified formulation/technique to analyze the synthesis, interdiction, and protection of layered networks. First, a single layer multiobjective network formulation was developed based on multiobjective optimization and network optimization. This was then combined with multilayer network formulations to develop a new formulation for robust network design so that layered networks with interconnected effects can be developed. This allows such items as holistic analysis of infrastructure networks vice single infrastructure networks in isolation.

A second formulation development combines multilayered network programming with network interdiction to form multilayer interdiction. Next, a new formulation is developed to show how network interdiction techniques can be extended to traditional social network analysis. Using measures developed by sociologists and others, this piece shows how Operations Research techniques can be used to maximally disrupt a human network. Finally, two formulation extensions (which have not yet been developed) are discussed: nodal interdiction and coalitions and/or multiple optimal solutions.

The remainder of this dissertation is organized as follows: the next chapter, Chapter 3, discusses the extension of network interdiction to directly include nodes. Traditionally, network interdiction focuses on edges, and this extension allows direct modeling of nodes. Chapter 4 discusses the extension of network interdiction to social networks. This allows the use of social network measures to determine key relationships for disrupting the social network. Chapter 5 extends network interdiction to allow for modeling of multiple layered networks (with interdependencies). Chapter 6 discusses the cost versus robustness in designing/expanding networks. This is further extended to include multiple layered networks, as well. Chapter 7 discusses the importance of considering multiple optimal solutions when studying network vulnerabilities/attacks/etc. Finally, Chapter 8 shows how these tools can be combined and applied to analyze infrastructures and allocation of resources.

III. Nodal Interdiction

3.1 Overview

This chapter extends network interdiction to directly include node interdiction. Current interdiction literature focuses primarily on arcs/edges. However, the military often targets nodes. In addition, social networks revolve around people as nodes connected by relationships. Conceptually, it is often easier to think of targeting/influencing people as opposed to nodes. Traditional network interdiction generally incorporates nodes by replacing each node with two artificial nodes and an artificial link and then uses a links interdiction approach. However, this increases the size of the network (in fact, it would double the the number of nodes in the network if every node is targetable), and in some cases, may not be intuitive to the user. To more transparently and directly represent nodes to target/protect, a formulation that explicitly considers nodes is developed and discussed.

3.2 Background

Joint doctrine defines interdiction as follows:

> Interdiction operations are actions to divert, disrupt, delay, or destroy an enemy's surface capabilities before they can be used effectively against friendly forces, or to otherwise achieve objectives. [97, p. I-1]

Diversion is defined as making the enemy "consume resources or capabilities critical to enemy operations in a way that is advantageous to friendly operations." [97, p. I-2] Disruption involves "upsetting the flow of information, operational tempo, effective interaction, or cohesion of the enemy force or those systems." [97, p. I-1] Delay involves "alter the ability of the enemy or adversary to project forces or capabilities" in a timely manner. [97, p. I-3] Finally, destroy refers to "damage the structure, function, or condition of a target so that it can neither perform as intended nor be restored to a usable condition, rendering it ineffective or useless." [97, p. I-4]

Since the term "interdiction" incorporates all these actions, some confusion may result if the terms are not used carefully. For example, if a node is destroyed, then all flow through arcs to and from the destroyed node are disrupted. Technically, both the node and the arcs are "interdicted," but using the specific terms such as destroyed and/or disrupted clarifies how the interdiction has occurred (i.e. first-order versus second-order effects). These distinctions are rarely made in the operations research literature, which makes interpretation of results difficult.

Interdiction doctrine also recognizes that interdiction may be potentially used for homeland security. "Interdiction also can be used to prevent an enemy from achieving a variety of objectives affecting the US populace, economy, or national interests." [97, p. I-1] In addition, doctrine also recognizes that information operations may play a unique role in interdiction stating, "interdiction may support, be supported by, or include aspects of information operations." [97, p. I-1]

> The nonlethal nature of many IO capabilities allows their use prior to and after hostilities, extending contact across time, thereby giving the friendly force greater opportunity to influence events and outcomes favorably. [97, p. II-13]

This chapter focuses on extensions to maximum flow interdiction to explicitly allow nodal interdiction. Other network flow formulation could also be used. For example, social network analysis relies heavily on shortest paths, so a shortest-path nodal interdiction is desired. Later chapters, Social Network Interdiction and Multiple Layered Network Interdiction, demonstrate how nodal interdiction can be modified and incorporated to extend these formulations. However, for ease of illustration, the primary focus of this chapter is maximum flow nodal interdiction.

Sections 2.2.3 and Section 2.5.2 discussed maximum flow and maximum flow arc-interdiction, respectively. The maximum flow problem can be written as

$$
\begin{aligned}
\max \quad & x_{ts} \\
\text{s.t.} \quad & \sum_j x_{sj} - \sum_j x_{js} - x_{ts} = 0 \\
& \sum_j x_{ij} - \sum_j x_{ji} = 0, \quad \forall \ \ (i,j) \in A \\
& \sum_j x_{tj} - \sum_j x_{jt} + x_{ts} = 0 \\
& 0 \leq x_{ij} \leq u_{ij}, \quad \forall \ \ (i,j) \in A \\
& x_{ij} \geq 0
\end{aligned}
\tag{3.1}
$$

where A is a set of directed arcs, x_{ij} represents flow across arc (i,j), x_{ts} is an artificial arc from t to s, and u_{ij} is an upper bound on the flow across arc (i,j).

An attacker on this maximum flow network would seek to minimize the amount of flow from s to t. If the attacker has sufficient resources, then he may determine a (node or mixed) cut-set (see Section 2.2.5) and prevent all flow from s to t. However, if the attacker has limited resources, then he or she must determine which subset of elements to attack which would maximally limit the amount of flow. Wood has shown that interdiction of this maximum flow can be expressed as the following bilevel program: [137, p. 5]

$$
\begin{aligned}
\min_{\gamma \in \Gamma} \max_{x} \quad & x_{ts} \\
\text{s.t.} \quad & \sum_j x_{sj} - \sum_j x_{js} - x_{ts} = 0 \\
& \sum_j x_{ij} - \sum_j x_{ji} = 0, && \forall \ \ i \in \text{N-\{s,t\}} \\
& \sum_j x_{tj} - \sum_j x_{jt} + x_{ts} = 0 \\
& x_{ij} - u_{ij}(1 - \gamma_{ij}) \leq 0, && \forall \ \ (i,j) \in A \\
& x_{ij} \geq 0, && \forall (i,j) \in A \cup \{(t,s)\}
\end{aligned}
\tag{3.2}
$$

where $\Gamma \equiv \{\gamma_{ij} | \gamma_{ij} \in \{0,1\} \forall (i,j) \in A, \sum_{(i,j) \in A} r_{ij}\gamma_{ij} \leq R\}$.

This formulation determines the set of arcs (identified by γ_{ij}) whose disruptions will maximally reduces the amount of resulting flow through the network. When $\gamma_{ij} = 1$ then the 4th constraint of (3.2) forces the upper bound of flow through that arc to 0; whereas when $\gamma_{ij} = 0$ then flow (and upper bounds) on an arc are unaffected.

To facilitate solution algorithms, Wood proves the model given in (3.2) can be transformed into a single level integer program. He shows that for fixed γ_{ij}, the inner (follower's) problem is a traditional maximum flow problem. Since maximum flow problems are unimodular, the integrality requirements are non-restrictive, allowing the dual of this inner problem (with fixed γ_{ij}) to be taken. Therefore, the inner problem can be replaced with its dual, and γ_{ij} is "released." With this transformation, both objective functions are minimizations, but over differing sets of variables (the leader's primary variables, and the follower's dual variables). The objective functions can, therefore, be combined resulting in the following mixed-integer program: [137, p. 7]

$$\begin{aligned}
\min \; & \sum_{(i,j)\in A} u_{ij}\beta_{ij} & & \\
\text{s.t. } & \alpha_i - \alpha_j + \beta_{ij} + \gamma_{ij} \geq 0, & & \forall (i,j) \in A \\
& \alpha_t - \alpha_s \geq 1 & & \\
& \sum_{(i,j)\in A} r_{ij}\gamma_{ij} \leq R & & \\
& \alpha_i \in \{0,1\}, & & \forall i \in N \\
& \beta_{ij}, \gamma_{ij} \in \{0,1\}, & & \forall (i,j) \in A
\end{aligned} \tag{3.3}$$

where $\alpha_i = 1$ for i on the t side of the cut and $\alpha_i = 0$ for i on the s side of the cut; γ_{ij} is 1 if (i,j) is a forward arc across the cut which is to be disrupted; β_{ij} is 1 if (i,j) is a forward arc across the cut, but is not to be disrupted; and all other β_{ij} and γ_{ij} are zero.

As with formulation (3.2), the solution of Wood's model in (3.3) indicates which arcs (identified by $\gamma_{ij} = 1$) should be denied to minimize the maximum flow through the network.

3.3 Node Only Formulation

However, formulation (3.3) does not allow for the direct interdiction of nodes. Traditionally, this problem is solved via "node splitting." Node splitting replaces each candidate target node in the original network, i, with 2 artificial nodes, i' and i'', and a link from i' to i''. With this transformation, interdiction of arc (i', i'') is interpreted as an interdiction of node i in the original network. Unfortunately, this approach, in a worst case, could double the number of nodes and adds an additional n arcs if all n nodes are candidates. While allowing nodes to be interdicted allows more realistic modeling, "computations can be hindered by the larger size of the transformed network." [46, p. 8] It should be noted that if an attacker's resources are not a limitation, then numerous cut-set identification methods exist in the literature which include nodes/vertices: see [55, 93]. However, these approaches may not be appropriate for interdiction where attacker resources are limited (making a full cut-set impossible).

A node-only interdiction formulation was provided by Whiteman [132]. In this article, he showed that Wood's edge formulations could be converted to a node interdiction formulation by simply using nodal subscripts instead of edge subscripts in the dual formulation of the maximum flow problem. [132, p. 21] This formulation is useful if only node information is available. However, it may be difficult to extend this formulation to include edge information (and interdictions). Therefore, a more general version of this formulation is developed.

Similar to Whiteman's approach, this development begins with Wood's original formulation and modifies it. A node-only interdiction formulation is developed as a bilevel program.

$$\begin{aligned}
\min_{\gamma\in\Gamma}\max_{x} \quad & x_{ts} & & \\
\text{s.t.} \quad & \sum_j x_{sj} - \sum_j x_{js} - x_{ts} = 0 & & \\
& \sum_j x_{ij} - \sum_j x_{ji} = 0, & & \forall \;\; i \in \text{N-\{s,t\}} \\
& \sum_j x_{tj} - \sum_j x_{jt} + x_{ts} = 0 & & \qquad (3.4) \\
& x_{ij} - u_{ij}(1-\gamma_{ij}) \leq 0, & & \forall \;\; (i,j) \in A \\
& \gamma_{ij} = \gamma_i & & \forall \;\; i \in N \text{ and } \forall \;\; (i,j) \in A \\
& x_{ts} \geq 0, & & \forall (i,j) \in A \cup \{(t,s)\}
\end{aligned}$$

where γ_i is 1 if node i is interdicted (and 0 otherwise), r_i is the cost of interdicting node i, and $\Gamma \equiv \{\gamma_i | \gamma_i \in \{0,1\} \forall (i) \in N, \sum_i r_i \gamma_i \leq R\}$.

There are two differences between the formulation in (3.4) and that of (3.2). First, the follower's problem has an additional constraint $\gamma_{ij} = \gamma_i$. This constraint states that outgoing edges of a node are denied/disrupted if and only if the node itself is interdicted. In other words, if a node is interdicted, all (outgoing) edges incident to that node are denied/disrupted (have their upper bound capacities reduced to zero). Of course, if a node is interdicted all incoming edges are denied/disrupted as well, but this constraint is redundant so it is not included. The second difference between (3.4) and (3.2) is that the attacker/leader is restricted to attacking nodes, γ_i, (instead of arcs) with associated interdiction costs, r_i.

Similar to Wood's development for traditional maximum flow interdiction, formulation (3.4) is transformed into a (single level) mixed-integer program. As with Wood's formulation, for fixed interdiction, γ_i, the follower's problem is a traditional maximum flow problem. Therefore, interdiction variables (γ_i and therefore γ_{ij}) are temporarily fixed, the dual of the follower's problem is taken, and then interdiction variables are released. This results in the following formulation:

$$
\begin{aligned}
\min \; & \sum_{(i,j)\in A} u_{ij}\beta_{ij} && \\
\text{s.t.} \; & \alpha_i - \alpha_j + \beta_{ij} + \gamma_{ij} \geq 0, && \forall (i,j) \in A \\
& \alpha_t - \alpha_s \geq 1 && \\
& \sum_{i\in N} r_i \gamma_i \leq R && \\
& \gamma_{ij} = \gamma_i && \forall i \in N \\
& \alpha_i \in \{0,1\}, && \forall i \in N \\
& \beta_{ij}, \gamma_{ij} \in \{0,1\}, && \forall (i,j) \in A
\end{aligned}
\tag{3.5}
$$

where $\alpha_i = 1$ for i on the t side of the cut and $\alpha_i = 0$ for i on the s side of the cut; γ_{ij} is 1 if (i,j) is a forward arc across the cut which is to be broken; β_{ij} is 1 if (i,j) is a forward arc across the cut, but arc (i,j) is not to be broken; and all other β_{ij} and γ_{ij} are zero.

As with formulation (3.4), the solution of this program indicates which nodes (identified by $\gamma_i = 1$) should be interdicted to minimize the maximum flow through the network.

3.4 Nodes and Edges Formulations

In addition to arc-only and node-only interdiction, some circumstances may exist where both nodes and edges are targetable. Due to the differences between targeting a node and targeting an edge, targeting edges and nodes, a mixed interdiction model, may require the use of different resources. Therefore, in these cases, separate resource constraints could be used. Following this approach, Subsection 3.4.1 discusses mixed interdiction where the components do not share resources in targeting. This is followed by Subsection 3.4.2 where all resources are shared. Of course, other variations are possible, such as having some resources shared and some not. The developments presented here allow for an exhaustive choice of combinations.

3.4.1 Different Node and Edge Interdiction Resources

Continuing the development shown in the previous section, a bilevel formulation of maximum flow interdiction for both targeted nodes and edges is developed.

$$
\begin{aligned}
\min_{\gamma\in\Gamma}\max_{x}\quad & x_{ts} & & \\
\text{s.t.}\quad & \sum_j x_{s,j} - \sum_j x_{js} - x_{ts} = 0 & & \\
& \sum_j x_{ij} - \sum_j x_{ji} = 0, & \forall\ \ & i \in \text{N-\{s,t\}} \\
& \sum_j x_{tj} - \sum_j x_{jt} + x_{ts} = 0 & & \qquad (3.6) \\
& x_{ij} - u_{ij}(1-\gamma_{ij}) \leq 0, & \forall\ \ & (i,j) \in A \\
& \gamma_{ij} \geq \gamma_i & \forall\ \ & i \\
& x_{ts} \geq 0, & & \forall (i,j) \in A \cup \{(t,s)\}
\end{aligned}
$$

where γ_i is 0 if node i is interdicted (and 0 otherwise), r_i is the cost of interdicting node i, r_{ij} is the cost of interdicting edge (i,j), and $\Gamma \equiv \{\gamma_i, \gamma_{i,j} \in \{0,1\}, \sum_i r_i\gamma_i \leq R_1, \sum_{(i,j)\in A} r_{ij}\gamma_{ij} - \sum_{(i,j)\in A} r_{ij}\gamma_i \leq R_2\}$.

The difference between the follower's problem in (3.6) and (3.4) is that the constraint $\gamma_{ij} = \gamma_i$ is transformed into a "greater than or equal to" constraint. This constraint still forces all outgoing arcs from an interdicted node to also be disrupted, but it also allows for arcs to be interdicted which are not part of a nodal interdiction. Therefore, care must be taken to account for interdiction resources in the leader's/attacker's problem. For nodes selected for interdiction, the resource constraint is the same as in (3.4). However, for edges selected for interdiction, only those edges which are not part of nodes selected for interdiction have their costs included. In other words, whenever a node is interdicted, its associated outgoing edges are also considered to be denied. However, this cost is already accounted for in the node resource constraint; in order to prevent these costs from being counted again in the edge resource constraint, the costs associated with edges disrupted due to nodal interdiction are subtracted from the constraint. This is done with the following constraint on the leader: $\sum_{(i,j)\in A} r_{ij}\gamma_{ij} - \sum_{(i,j)\in A} r_{ij}\gamma_i \leq R_2$. The first term is the direct arc interdiction cost. If a node is interdicted, all associated arcs are automatically disrupted, but this cost is already included in the node interdiction, so the cost of disrupting these arcs is subtracted from the direct arc interdiction cost.

Following the same approach developed by Wood, this formulation is transformed into a single level mixed-integer formulation. This is done by fixing the interdiction γ_{ij} and γ_i, replacing the follower's problem with its dual, and releasing the interdiction. This results in the following program:

$$\begin{aligned}
\min \; & \sum_{(i,j)\in A} u_{ij}\beta_{ij} & \\
\text{s.t. } & \alpha_i - \alpha_j + \beta_{ij} + \gamma_{ij} \geq 0, & \forall (i,j) \in A \\
& \alpha_t - \alpha_s \geq 1 & \\
& \gamma_{ij} \geq \gamma_i & \forall i \in N \\
& \sum_{i\in N} r_i\gamma_i \leq R_1 & \\
& \sum_{(i,j)\in A} r_{ij}\gamma_{ij} - \sum_{(i,j)\in A} r_{ij}\gamma_i \leq R_2 & \\
& \alpha_i \in \{0,1\}, & \forall i \in N \\
& \beta_{ij}, \gamma_{ij} \in \{0,1\}, & \forall (i,j) \in A
\end{aligned} \tag{3.7}$$

where all variables are as previously defined.

The solution of this program indicates which nodes (identified by $\gamma_i = 1$) and edges (identified by γ_{ij}) should be interdicted to minimize the maximum flow through the network. In this formulation, node and edge interdiction do not share interdiction resources.

3.4.2 Shared Node and Edge Interdiction Resources

Finally, the case where both nodes and edges can be interdicted, and doing so consumes the same resources, is considered. In the bilevel formulation of the problem, the follower has the same problem as in (3.6). Only the leader's resource constraints are different. Therefore, the development of the single level mixed integer program is the same and therefore is not repeated here. This results in the following formulation:

$$
\begin{aligned}
\min \; & \sum_{(i,j)\in A} u_{ij}\beta_{ij} & & \\
\text{s.t. } \; & \alpha_i - \alpha_j + \beta_{ij} + \gamma_{ij} \geq 0, & & \forall (i,j) \in A \\
& \alpha_t - \alpha_s \geq 1 & & \\
& \sum_{(i,j)\in A} r_{ij}\gamma_{ij} - \sum_{(i,j)\in A} r_{ij}\gamma_i + \sum_{i\in N} r_i\gamma_i \leq R & & \\
& \gamma_{ij} \geq \gamma_i & & \forall i \in N \\
& \alpha_i \in \{0,1\}, & & \forall i \in N \\
& \beta_{ij}, \gamma_{ij} \in \{0,1\}, & & \forall (i,j) \in A
\end{aligned}
\tag{3.8}
$$

The difference between (3.8) and (3.7) is that the resource constraints are combined. The solution of this program indicates which nodes (identified by $\gamma_i = 1$) and edges (identified by γ_{ij}) should be interdicted to minimize the maximum flow through the network. In this formulation, node and edge interdiction share interdiction resources.

3.5 Notional Example

In this section, the formulations developed in (3.8) are demonstrated on the notional network in Figure 3.1.

Unless stated otherwise, the cost to interdict an edge or node (other than node 1 and node 6) is one unit (to simplify the illustration). Of course, in real world

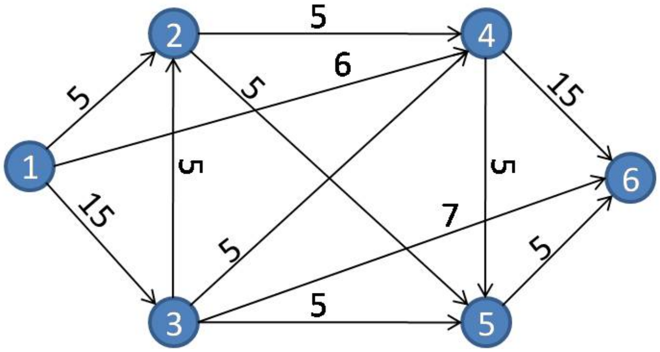


Figure 3.1: Notional Network

networks, these costs will vary. In this example, it is assumed that neither the source nor sink may be interdicted. Therefore, they are assigned an arbitrarily high interdiction cost, M. The uninterdicted maximum flow of this network is 26.

3.5.1 Node Only Interdiction

Using formulation (3.8), the following mathematical program results with a resource constraint of one (i.e. since the cost of interdicting transhipment nodes is also one, this means one node can be interdicted):

$$
\begin{aligned}
\min\ & 5\beta_{(1,2)} + 15\beta_{(1,3)} + 6\beta_{(1,4)} + 5\beta_{(2,4)} + 5\beta_{(2,5)} + 5\beta_{(3,2)} \\
& + 5\beta_{(3,4)} + 5\beta_{(3,5)} + 7\beta_{(3,6)} + 5\beta_{(4,5)} + 15\beta_{(4,6)} + 5\beta_{(5,6)} \\
\text{s.t. } & \alpha_1 - \alpha_2 + \beta_{(1,2)} + \gamma_{(1,2)} \geq 0 \\
& \alpha_1 - \alpha_3 + \beta_{(1,3)} + \gamma_{(1,3)} \geq 0 \\
& \alpha_1 - \alpha_4 + \beta_{(1,4)} + \gamma_{(1,4)} \geq 0 \\
& \alpha_2 - \alpha_4 + \beta_{(2,4)} + \gamma_{(2,4)} \geq 0 \\
& \alpha_2 - \alpha_5 + \beta_{(2,5)} + \gamma_{(2,5)} \geq 0 \\
& -\alpha_2 + \alpha_3 + \beta_{(3,2)} + \gamma_{(3,2)} \geq 0 \\
& \alpha_3 - \alpha_4 + \beta_{(3,4)} + \gamma_{(3,4)} \geq 0 \\
& \alpha_3 - \alpha_5 + \beta_{(3,5)} + \gamma_{(3,5)} \geq 0 \\
& \alpha_3 - \alpha_6 + \beta_{(3,6)} + \gamma_{(3,6)} \geq 0 \\
& \alpha_4 - \alpha_5 + \beta_{(4,5)} + \gamma_{(4,5)} \geq 0 \\
& \alpha_4 - \alpha_6 + \beta_{(4,6)} + \gamma_{(4,6)} \geq 0 \\
& \alpha_5 - \alpha_6 + \beta_{(5,6)} + \gamma_{(5,6)} \geq 0 \\
& \gamma_{(1,2)} - \gamma_1 = 0 \\
& \gamma_{(1,3)} - \gamma_1 = 0 \\
& \gamma_{(1,4)} - \gamma_1 = 0 \\
& \gamma_{(2,4)} - \gamma_2 = 0 \\
& \gamma_{(2,5)} - \gamma_2 = 0 \\
& \gamma_{(3,2)} - \gamma_3 = 0 \\
& \gamma_{(3,4)} - \gamma_3 = 0 \\
& \gamma_{(3,5)} - \gamma_3 = 0 \\
& \gamma_{(3,6)} - \gamma_3 = 0 \\
& \gamma_{(4,5)} - \gamma_4 = 0 \\
& \gamma_{(4,6)} - \gamma_4 = 0 \\
& \gamma_{(5,6)} - \gamma_5 = 0 \\
& M\gamma_1 + \gamma_2 + \gamma_3 + \gamma_4 + \gamma_5 + M\gamma_6 \leq 1
\end{aligned}
\tag{3.9}
$$

where all variables are binary.

Solving this model results in $\gamma_3 = 1$, indicating node 3 is interdicted, which denies the use of edges $(3, 2)$, $(3, 4)$, $(3, 6)$, $(3, 5)$, and $(1, 3)$. This results in a remaining maximum possible flow of 11. If the resource constraint is raised to 2, then nodes 3 and 4 are interdicted, leaving a maximum possible flow of 5.

3.5.2 Nodes and Edges Formulations with Different Resources

In this extension of the example, nodes and edges can be interdicted, and the interdiction consumes different resources. Again, for illustration/simplicity, it is assumed that the node and edge interdiction is limited to a resource constraint of 1 each. The formulation remains very similar to the previous section with minor changes. The node to edge constraints are changed from equality constraints to "greater than or equal to" constraints. In addition, the following constraint is added to allow for edge interdiction: $\gamma_{(1,2)} + \gamma_{(1,3)} + \gamma_{(1,4)} + \gamma_{(2,4)} + \gamma_{(2,5)} + \gamma_{(3,2)} + \gamma_{(3,4)} + \gamma_{(3,5)} + \gamma_{(3,6)} + \gamma_{(4,5)} + \gamma_{(4,6)} + \gamma_{(5,6)} - 3\gamma_1 - 2\gamma_2 - 4\gamma_3 - 2\gamma_4 - \gamma_5 \leq 1$

This model results in an interdiction selection of node 3 and edge $(4, 6)$, each interdiction consuming one unit of their respective resources. This interdiction results in a maximum possible flow of 5.

3.5.3 Nodes and Edges Formulations with Shared Resources

Finally, we consider the case where both nodes and edges can be interdicted, and the interdiction resources are shared. This formulation is the same as the previous section, except that the resource constraints are combined. In the first extension of the case, a resource constraint of 1 unit is initially assigned. Since all edges and transhipment nodes have a cost of 1, this will limit the interdiction to 1 node or edge. This leads to an interdiction selection of edge $(1, 3)$ which results in a possible uninterdicted flow of 11. It should be noted that there are multiple optimal solutions. The interdiction of node 3 would also result in the same flow. However, a more extensive discussion of multiple optimal solutions is delayed until Chapter VII.

If the interdiction resource is increased to 2, then an optimal strategy is interdicting edges $(3, 6)$ and $(4, 6)$; leaving a possible uninterdicted flow of 5. If the cost to interdict edge $(3, 6)$ is increased to 2, and the resource constraint remains at 2, then an optimal solution is to interdict edge $(4, 6)$ and node 3; again leaving a possible flow of 5.

3.6 Computational Experiments

3.6.1 Comparison

There are no previous studies in the open literature which explicitly consider nodal interdiction. Instead, authors generally use interdiction on edges, and point out that nodes can be converted to edges if desired (using the node splitting technique discussed in Section 3.3). In Wood's seminal article on network interdiction, he presents the example illustrated in Figure 3.2 and Table 3.1. [137, p. 16]

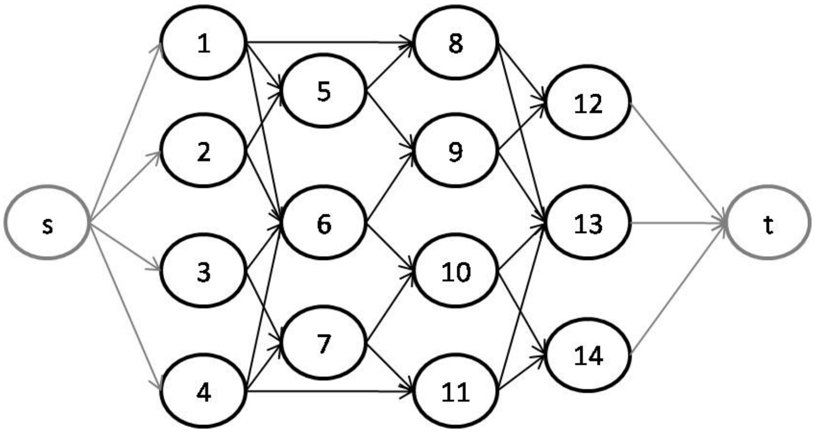


Figure 3.2: Notional Network [137, p. 16]

To facilitate a nodal interdiction problem, cost to interdict nodes was assigned to interdictable nodes. For this problem, Wood assumed the following nodes could not be interdicted: 1 through 4, and 12 through 14 (as well as the artificial source and sink nodes). The remaining nodes were assigned interdiction costs as shown in Table 3.2.

Table 3.1: Wood's Arc Interdiction Resources [137, p. 16]

Arc	Capacity	Resource	Arc	Capacity	Resource
(1,5)	60	5	(6,9)	120	4
(1,6)	60	5	(6,10)	150	6
(1,8)	70	4	(7,10)	120	6
(2,5)	50	3	(7,11)	80	4
(2,6)	50	3	(8,12)	80	4
(2,7)	60	5	(8,13)	50	5
(3,6)	100	3	(9,12)	100	5
(3,7)	80	5	(9,13)	80	4
(4,6)	50	5	(10,13)	180	6
(4,7)	100	5	(10,14)	100	4
(4,11)	80	4	(11,13)	80	5
(5,8)	60	4	(11,14)	100	6
(5,9)	60	7			

Table 3.2: Node Interdiction Costs

Node	Resource
5	8
6	16
7	15
8	8
9	11
10	12
11	8

The optimal node interdiction solution can be found directly using the formulation in Section 3.3. However, to use the traditional edge-only formulation, node-splitting must be used. This results in the network in Figure 3.3.

With this modified example, the edges between the "split" nodes are assigned the interdiction cost associated with the node as in Figure 3.3.

The example in Figure 3.2 was solved with the node only formulation, and the modified example in Figure 3.3 was solved using traditional edge-only interdiction formulation. These formulations were solved using GAMS with the BARON/CPLEX

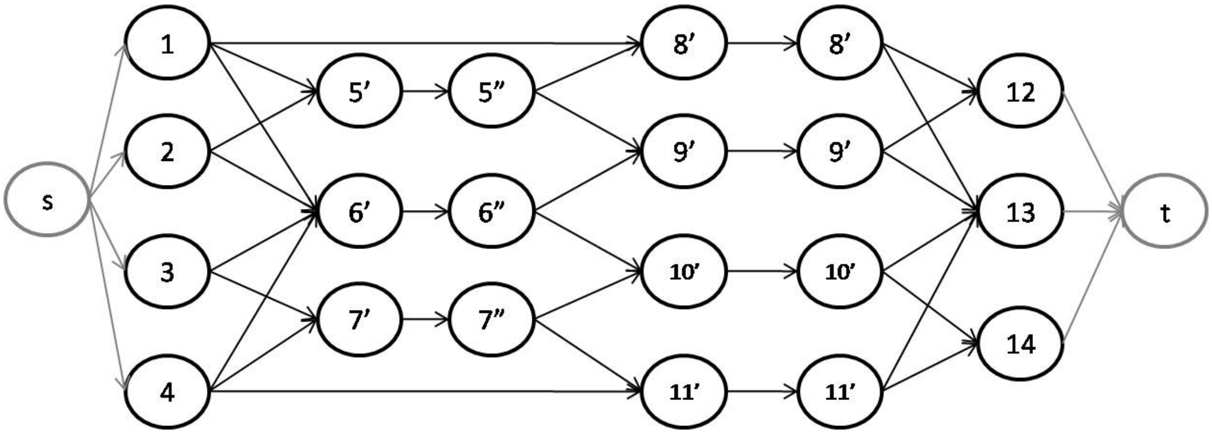


Figure 3.3: Wood Example with Node Splitting

solvers and on a Dell Precision M6300 with 2.50 gigahertz Intel Core2 Duo processor and 4094 megabytes of RAM. The results are presented in Table 3.3.

Table 3.3: Node Interdiction Costs

	Modified Edge Interdiction	Direct Node Interdiction
Rows	33	26
Columns	60	46
Non-zeros	110	96
Binaries	60	46
Solution Time	0.115 sec	0.113 sec

As this table shows, using the node interdiction formulation resulted in a smaller problem size and a virtually indistinguishable run time. However, as problem sizes grow, it is anticipated that the smaller formulation sizes would result in a more significant solution time difference. To test this algorithm against larger problem sizes, the next section considers progressively larger networks.

3.6.2 Larger Networks

To construct larger networks and to be consistent with traditional interdiction literature, GRIDGEN was used to generate random grid networks. GRIDGEN was developed by Dimitri Bertsekas and is available in his text *Linear Network Optimization.* [15, p. 254-259]. Networks were generated by providing a specified length and width of the network to be generated. In this generator, all edges are directed and all nodes are connected to their adjacent nodes. In addition, each node is also connected to two randomly chosen additional nodes. Unless specified otherwise, all networks generated in this section have the following properties: all edges have capacities and edge interdiction costs randomly (uniformly) assigned between 10 and 100, and node interdiction costs are randomly assigned uniformly between 10 and 100 (and rounded to the nearest integer). Initially, all nodes down the first column of the grid are source nodes, and all nodes down the last column are demand nodes. A "super-source" and "super-sink" are added to transform the network into a traditional maximum flow form. These super-nodes (and associated edges) can not be interdicted.

To be consistent with previous network interdiction literature ([46], [71], [102]), the following network grids were included: 10 nodes by 10 nodes, 30 nodes by 30 nodes, and 40 nodes by 80 nodes. In addition, one goal of this research is to provide analysis techniques for large networks, such as infrastructure networks. Therefore, an additional network was considered: 100 nodes by 150 nodes. This network has 15,000 (including 100 source and 100 sink) nodes and 90,000 edges.

In order to make a direct comparison, the same set of networks were used across all the testing. For example, the same "10x10" network was used both within each model type (i.e. various interdiction resources of 10, 50, and 100) and across model types (i.e. node only, node and edges, and so forth).

As the results in Table 3.4 demonstrate, these formulations can indeed be applied to a range of networks, including larger sizes, with reasonable solution times. In all cases, as the the size of the network grows, the number of iterations and time required grows proportionately, as expected. In addition, as the amount of resources available for interdiction grows, the number of iterations and time usually did as well. Interestingly, the models with shared resources require the most iterations and time of all three models. This is due to the increased flexibility of using resources

Table 3.4: Computational Results of Interdiction Formulations

Formulation	Grid	Interdiction Resources	Iterations	Total Clock (H:MM:SS)
Node Only	10x10	10	185	0:00:00.415
		50	219	0:00:00.391
		100	325	0:00:00.445
	30x30	10	1265	0:00:01.242
		50	1435	0:00:01.507
		100	1479	0:00:01.650
	40x80	10	4269	0:00:04.409
		50	5135	0:00:06.192
		100	5334	0:00:07.253
	100x150	10	17556	0:00:27.138
		50	21007	0:00:31.114
		100	21422	0:00:39.388
Nodes and Edges with Shared Resources	10x10	10	1265	0:00:00.753
		50	1108	0:00:00.542
		100	1080	0:00:01.056
	30x30	10	10394	0:00:09.032
		50	9996	0:00:08.128
		100	9808	0:00:16.187
	40x80	10	39349	0:02:07.302
		50	39062	0:02:49.359
		100	36239	0:01:52.462
	100x150	10	123617	0:29:44.409
		50	163617	0:35:07.158
		100	173480	0:49:44.409
Nodes and Edges with Different Resources	10x10	10	804	0:00:00.510
		50	937	0:00:00.662
		100	1193	0:00:00.728
	30x30	10	3441	0:00:02.476
		50	7095	0:00:07.770
		100	9611	0:00:11.729
	40x80	10	18374	0:00:47.728
		50	22704	0:01:19.714
		100	28705	0:02:07.437
	100x150	10	67188	0:12:09.003
		50	91899	0:22:21.841
		100	107584	0:34:12.400

for nodes or edges, so many more combinations are possible and these tradeoffs need to be considered.

This table also shows some seemingly counterintuitive results. For example, in some cases (i.e. 30x30 grid in shared resources model) some variations required more computational time, but fewer iterations. In another example, in a few cases the model solved faster with more resources available than with less interdiction resources. These outcomes are a result of using the default settings of GAMS. In some cases, GAMS was able to create generalized upper bounding constraints and/or cover constraints during preprocessing. In these cases, the number of iterations and time required to converge to an optimal solution is skewed by the presence of these additional constraints.

3.6.2.1 Comparison

For the prospposed technique to be useful, its effectiveness must be at least equivalent to current methods, and preferably faster. The current algorithms for node interdiction use the node splitting technique discussed in Section 3.3. To facilitate a comparison of the node interdiction developed here with the previous methods of node-splitting, the grid networks from the previous section were used. Specifically, the 40x80 grid was used directly in the node-only formulation. For the node-splitting algorithm, the grid was modified as follows: the first and last columns remain unsplit as they are the source and sinks, respectively; all remaining nodes were split using to the node splitting technique. All interdictable arcs (i.e. those between split nodes) were assigned an interdiction cost of one unit.

Both grid networks were used in the respective formulations in GAMS with the results in Table 3.5.

As Table 3.5 shows, the node splitting technique, as expected, increases the size of the network. The 40x80 network is transformed into a 78x80 network. However, as only edges can be interdicted between split nodes, the computational times between

Table 3.5: Direct Nodal versus Node Split Formulations

Formulation	Network	Interdiction Resources	Iterations	Total Clock (H:MM:SS)
Nodal Formulation	40x80	10	4269	0:00:04.409
		50	5135	0:00:06.192
		100	5334	0:00:07.253
Link (split nodes) Formulation	40x80 \ 78x80	10	8209	0:00:06.066
		50	8572	0:00:06.344
		100	13215	0:00:13.993

the two methods is comparable, with the direct nodal interdiction method requiring slightly fewer iterations and less time in this illustration.

The remaining nodal interdiction methods developed (which include edge interdiction) were not compared in this study. For node-split networks, all network edges (those from nodes and the original edges) are indistinguishable to the traditional edge interdiction algorithm. Although extensions to the traditional interdiction method to allow a comparison (such as adding appropriate subscript to the edges) are straightforward, the various potential methods would be arbitrary and/or specific the the particular application. As any comparison would be dependent to the method chosen, a general comparison of the two methods would be impossible.

3.7 Application

In order to demonstrate the potential of the models developed, they have been applied to a realistic communications network. One such notional communications network is depicted in Figure 3.4.

Edge capacities are given by Pinkstaff in Appendix C. [95, p. 182-208] In addition, Pinkstaff also provided “node reconstruction costs.” For this example, these costs are assigned to the nodes as node destruction costs. In addition, for the formulations in which edges can be interdicted, these costs were assigned as one-half of the average of the two nodes which that particular edge connects. To use

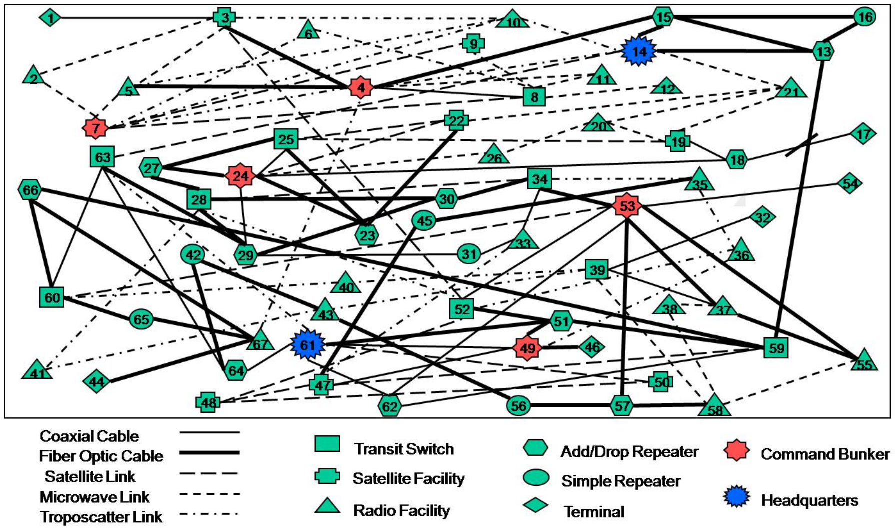


Figure 3.4: Notional Communications Network [95, p. 55]

this notional network as a maximum flow network, nodes marked “Headquarters” are assigned as sources, and “Command Bunkers” are assigned demand nodes. To simplify the formulation, a super-source and super-sink were also added.

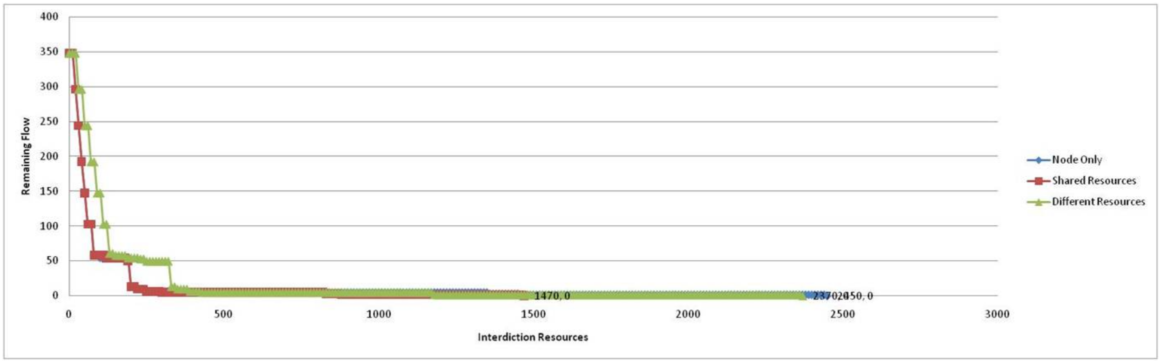


Figure 3.5: Interdiction Results

The results of the three formulations when applied to this notional network are depicted in Figure 3.5. As the graph shows, “node only” and “node and edges with shared resources” follow a very similar path; with “node only” reaching zero flow slightly faster due to its increased flexibility. In all cases, there is a dramatic drop as resources are increased for interdiction, followed by a plateau.

Table 3.6: Interdiction Results

	Node Only	Different Resources	Shared Resources
0	347.63	347.63	347.63
100	54.54	147.38	57.9
200	13.17	54.54	13.17
300	6.45	49.64	4.9
400	4.9	6.45	4.9
500	4.9	4.9	4.9
600	4.9	4.9	4.9
700	4.9	4.9	4.9
800	4.9	4.9	4.9
900	4.9	4.9	1.54
1000	4.9	4.9	1.54
1100	4.9	4.9	1.54
1200	4.9	1.54	1.54
1300	4.9	1.54	1.54
1400	3.09	1.54	1.54
1500	1.54	1.54	0
1600	1.54	1.54	0
1700	1.54	1.54	0
1800	1.54	1.54	0
1900	1.54	1.54	0
2000	1.54	1.54	0
2100	1.54	1.54	0
2200	1.54	1.54	0
2300	1.54	1.54	0
2400	1.54	0	0
2450	0	0	0

Specifically, the graph shows that with 80 units of interdiction resources, all three methods decreased the total flow from 350 to 50. However, it would take an additional 110 to 180 units of interdiction resources to decease the flow below 50.

3.8 Conclusion

This chapter extended network interdiction to include node interdiction. Three different formulations have been developed and these algorithms were demonstrated on small notional examples, small/medium/larger grid networks, and a realistic notional communications network. These tests show that these formulations allow for a potentially more intuitive formulation (because nodes are directly represented and do not require a split), a smaller formulation (than traditional node-splitting), and formulations that can be used to solve larger networks (including the 15,000 node network demonstrated). These approaches supplement the currently available approaches and give the analysts a wider set of options to directly model node and arc interdiction.

In addition, this formulation can be more easily extended than previous formulations. For example, in Chapters IV and V, the developed formulations are extended to nodal interdiction using the methods discussed in this chapter. In addition, this method could potentially be advantageous for sensitivity analysis because the node potentials are readily available in the given formulations.

IV. Social Network Interdiction

4.1 Overview

As shown in Figure 4.1, social network interdiction builds on traditional network interdiction. This is combined with traditional social network analysis techniques/metrics.

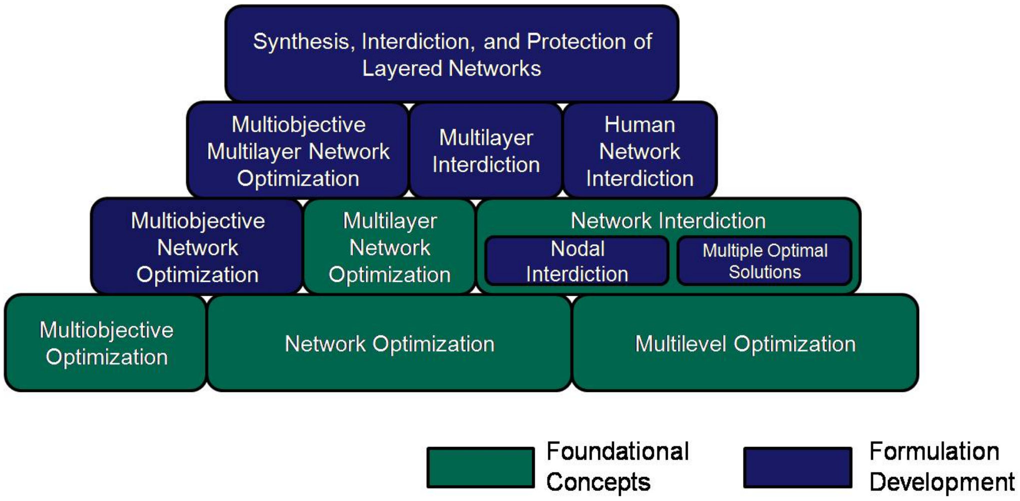

Synthesis, Interdiction, and Protection of Layered Networks
Multiobjective Multilayer Network Optimization
Multilayer Interdiction
Human Network Interdiction
Multiobjective Network Optimization
Multilayer Network Optimization
Network Interdiction
Nodal Interdiction
Multiple Optimal Solutions
Multiobjective Optimization
Network Optimization
Multilevel Optimization
Foundational Concepts
Formulation Development

Figure 4.1: Human Network Interdiction Formulation Development

This research developed a technique which combines network interdiction with traditional social network measures such as centrality. Previous research into disrupting social networks has not taken full advantage of network interdiction techniques. Herbranson [68] and Hamill [66] consider disruption of social networks, but they do not explicitly account for limited resources that could prevent various isolation strategies. Nesbitt [90] develops the idea of limited resources in network interdiction of social networks, but his method relies on an abstract (undefined) “flow” through the social network. One approach might be to combine this technique with flow measures developed by Renfro [99] or Hamill [66]; however, this research will combine the network interdiction programming techniques with closeness centrality measures currently used in SNA of covert organizations.

4.1.1 Assumptions

In this research, it is assumed that all actors and links of a social network are known. Although this is an unrealistic assumption in a covert network, the theory and technique developed here identify the optimal actor/relationship to target *given the information available*. Of course, this is the underlying assumption of all SNA measures, as they were not developed for use with missing information in mind. In addition, Carley noted that "any isolation is better than none, assuming our goal is to degrade the performance and that we don't need perfect information to be quite effective." [32, p. 10]

Borgatti *et al.* showed that centrality measures are robust to missing information in random graphs. [24] In random networks, he showed that errors in centrality measures increased linearly with the amount of missing information. However, Borgatti noted the degradation in estimation appears faster for cellular networks (which have little/no hierarchial structure and little communications between clusters of groups within the network) and may not be linear. Borgatti suggests that in random networks, all destabilization tactics (such as isolation of the individual that is highest in centrality) have approximately the same effect [25, p. 128]; but for cellular networks this may not be true. While outside the scope of this effort, additional study needs to be done to determine the impact of missing information on cellular structured networks.

In this chapter, it is assumed that resource constraints prevent complete isolation/cuts across a social network. These resource constraints could be monetary, political, ability to reach certain conduits, or any other constraint that would limit the ability to disrupt a social relationship. If resource constraints are not an issue, then methods discussed previously (such as isolation of key-players developed in Bellmore *et al.* [12] and expanded by Herbranson [68]) could be used.

In other cases, either the attacker does not have sufficient resources, a target is inaccessible, or completely "cutting" an arc may be impractical. For example, it may not be possible to break the ties between a parent and child or two life long friends. In the case of social network analysis, one may disrupt an actor directly (i.e. kill or detain the actor), or one may influence a relationship. For example, if one wished to disrupt a relationship, he/she could provide information that would cause actors not to trust each other. This would cause an increase in the centrality measures in a social network. Finally, one could weaken/cut a relationship by denying, disrupting, or degrading all means of communication between the actors.

4.2 Targeting

Some techniques in counter-terrorism involve identifying and isolating key members of a social network. Given social network structural information, the military and/or law enforcement can target those individuals and/or relationships to isolate "key players" in the network in an attempt to fragment the network and make the network less effective. [23, 68]. In addition, if a network "flow" measure is defined (such as in Renfro's work [99]) then a method developed by Nesbitt [90] can be used which applies network interdiction techniques to reduce this flow as much as possible in a covert network. However, there are no known studies in the open literature which examine the network interdiction techniques and social network measures directly.

Tsvetovat suggests that targeting those actors with high betweenness centrality may temporarily separate a cellular network into disconnected cells; however, the network uses latent resources to quickly recover. [119] In addition, targeting leadership in a cellular organization does not seem to disrupt the activities of the organization itself. This is made clear by Carley who asserts that isolating a key actor may not destabilize the network; in fact, this isolation "may have the same effect as cutting off the Hydras head; many new key actors or leaders may emerge." [33, p. 2]

Since the leadership and gatekeepers do not always seem to be attractive targets to disrupt a cellular organization, an alternative is to disrupt individual cells. Tsvetovat [119] suggests a search for highly central individuals is likely to turn up members of a densely connected cell. To disrupt this cell the members of the cell need to be forced to be less “connected.” However, thus far, connectedness has only been defined in terms of individuals; therefore a measure for the overall connectedness of the cell is needed.

4.2.1 Influence Operations

“Targeting is a comprehensive and involved process of matching a target within the cognitive, information, or physical domain with kinetic weapons or nonkinetic capabilities.” [1, p. 31] When dealing with people (as nodes), targeting can refer to a physical attack, or it can refer to an influence operation. Influence operations are focused on affecting the perceptions and behaviors of leaders, groups, or entire populations. [1, p. 43]

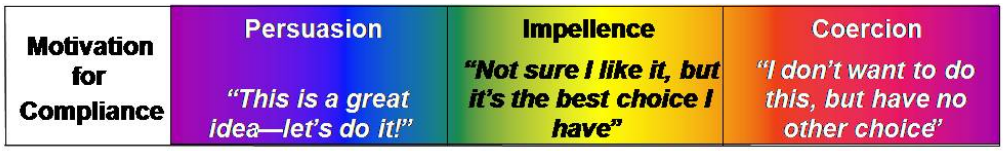
“The military capabilities of influence operations are psychological operations (PSYOP), military deception (MILDEC), operations security (OPSEC), counter-intelligence (CI) operations, counter-propaganda operations and public affairs (PA) operations.” [1, p. 5] For example, to disrupt a relationship between two individuals (the target), the military can employ PSYOP. “PSYOP seeks to induce, influence, or reinforce the perceptions, attitudes, reasoning, and behavior of foreign leaders, groups, and organizations in a manner favorable to friendly national and military objectives.” [1, p. 9] Additional information about PSYOPs can be found in Air Force Doctrine Document 2-5.2, *Psychological Operations*; and Joint Publication 3-53, *Doctrine for Joint Psychological Operations.*

Another example technique to target a relationship is military deception. “Military deception misleads or manages the perception of adversaries, causing them to act in accordance with friendly objectives.” [1, p. 11] Additional information about

MILDEC can be found in Joint Publication 3-13.4, *Military Deception*; and Joint Publication 3-58, *Joint Doctrine for Deception Operations.*

> There are a variety of ways to influence a person. At one extreme, an individual may be killed. It is fair to say that the deceased will no longer be opposed to our policy–but they will not help further it, either. At the other end of the spectrum, that same individual can be persuaded to accept, and hopefully embrace our policy–changing the individual's attitudes, behavior, opinions, and ultimately actions . . . In between killing someone and convincing them there is a spectrum of options–coerce, deter, compel, and persuade . . . [111, p. 6]

The influence weapon used will determine the motivation of the target. Figure 4.2 provides an example representation of the motivation spectrum.



Motivation for Compliance	Persuasion	Impellence	Coercion
	"This is a great idea—let's do it!"	*"Not sure I like it, but it's the best choice I have"*	*"I don't want to do this, but have no other choice"*

Figure 4.2: Spectrum of Motivation [111, p. 57]

To develop a measurable definition of influence, Clark defined influence as a function of prestige and connectivity. [41, p. I-14] Non-network prestige was determined from a discriminate analysis, while network interpersonal influence was calculated via information centrality. [41, p. 1-16] Section 2.2.7.4 discusses Clark's analysis technique in detail.

Of course, the influence weapon used should be tailored to the target audience. Targets also comprise a spectrum from one key decision maker, a small group (leadership coterie), or a direct appeal to the masses. [111, p. 7]

> As the population of the target audience increases, the requirement for precision intelligence on individual hopes and fears decreases, but the requirement for understanding the underlying social structure and values increases, along with the emotional inertia–and hence the time or magnitude of event required to inculcate real change. [111, p. 61]

This is shown in Figure 4.3. For example, in planning for influence operations, targets must be accessed for susceptibility to influence and expected benefit from influence. [111, p. 67]

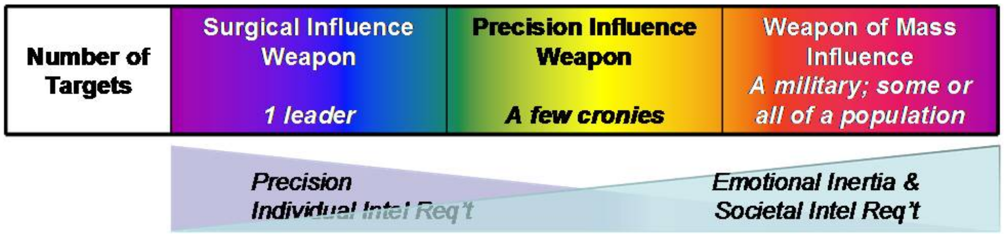


Figure 4.3: Number of Targets [111, p. 61]

In determining *how* to influence individuals, Kimminau merged rational decision making with prospect theory. According to Kimminau, most theories of influence rely on adversaries making rational choices. In other words, they will understand and weigh costs and benefits of alternatives, and they will chose the value-maximizing alternative. [75, p. 10] Kimminau argues that prospect theory is a more appropriate model for influence because it is based on individual decision making under risk. [75, p. 12] Prospect theory suggests that people frame their decisions based on their perception of their situation, and then evaluate alternatives differently depending upon their frame. Therefore, to apply a decision model to influence, "the alternatives must be defined in terms of costs, benefits, and uncertainty, and frame of the decision maker must be identified." [75, p. 29]

4.2.2 Cell Closeness

In this study, an approach for using network interdiction with SNA measures is developed using closeness centrality. Closeness centrality was discussed in Section 2.2.7.2, and was selected for several reasons. One, centrality is often used in SNA to determine "who" is important. In the context of covert organizations, members with high "closeness" measures are often found to be members of a terrorist cell when analyzing terrorist data. [119, p. 6] Therefore, although this measure may not identify overall core leadership of the organization, disrupting individual terrorist "cells" will disrupt or delay a covert organizations objectives.

Second, like many SNA measures, closeness centrality is built from shortest-path calculations. Closeness centrality is a measure of how close an individual is to everyone else in a connected network. The shortest paths "linking the central nodes to the other nodes must be as short as possible." [124, p. 183] As another example, betweenness centrality is a measure of how many times an actor appears on the shortest-path of other actors. [124, p. 189] Therefore, with the theory developed for closeness centrality, the approach developed here could be extended to other SNA measures that are based on shortest path calculations.

To disrupt a social network, one may wish to maximize the shortest distance from a set of central nodes to other nodes. Any method used to disrupt a relationship will incur a (not necessarily monetary) cost. Therefore, the goal is to identify the links which, if disrupted, would maximize the distance from central actors to all other actors subject to limited resources. If limited resources are not a (mathematical) constraint, then one could identify cut-sets which would completely disconnect the central nodes from other nodes. [12, 68].

Closeness centrality calculates the inverse of the sum of the shortest paths to all other nodes in the network. It is a measure of how "key" members are to network communication; reach; and reachibility. [57] Closeness centrality is defined in terms of the distance from an individual to all other nodes. For example, a node who "has the shortest possible paths to all the other actors ... has maximum closeness. [124, p. 184] Mathematically, closeness centrality is given as:

$$C(n_i) = \left[\sum_{j=1}^{g} d(n_i, n_j) \right]^{-1} \tag{4.1}$$

where n_i is the node for which centrality is being calculated, $d(n_i, n_j)$ is the distance from node i to node j, and g is the total number of nodes. This measure sums the shortest path from a node to all other nodes and takes the inverse. If no arc length is given to each relation, the distance between a pair of connected nodes is assumed to be one.

A network attacker is interested in decreasing "closeness" between multiple agents, not just a single agent. Therefore, this research extends the previously discussed individual centrality to a measure of "group closeness." A measure for this developed in this research is to sum the individual closeness measure across all members in the cell. In other words, define the members of the cell one wishes to disrupt, $l \in N$.

$$G = \sum_{l \in N} \left[\sum_{j=1}^{g} d(n_i, n_j) \right]^{-1} \tag{4.2}$$

where G is the "group closeness."

It should be noted that the inner sum is a sum of shortest paths. Therefore, to decrease closeness, the links to target to maximally increase the sum across the shortest paths need to determined. Let g be the the sum of the all-pairs shortest paths in the network. In order to minimize closeness, g needs to be maximized by interdicting arcs subject to the resource constraints.

This measure is very similar to the measure ${}^{D}TF$ developed by Herbranson. [68] However, this measure of group closeness does not use a predefined subset of nodes to isolate, T. In this study, the cell is degraded by disrupting closeness, not isolation.

4.3 Human Network Interdiction Model

4.3.1 Individuals

As shown in Section 4.1, closeness centrality is defined in terms of shortest paths. Recall from Section 2.2.2 that the shortest path from one individual, s to another individual t can be found by

$$\begin{aligned}
&\min \sum_{(i,j)\in A} c_{ij} x_{i,j} \\
&\sum_{j:(i,j)\in A} x_{ij} - \sum_{j:(j,i)\in A} x_{ji} = \begin{cases} 1, & \text{for } i = s; \\ 0, & 0\ \forall i \in N\ \{s,t\}; \\ -1, & \text{for } i = t. \end{cases} \\
&x_{ij} \geq 0 \ \ \forall (i,j) \in A
\end{aligned} \tag{4.3}$$

where c_{ij} be the length of arc (i,j) (i.e. the strength of the relationship), and let x_{ij} be 1 if arc (i,j) is chosen, and zero otherwise. In addition, A is an adjacency matrix where each entry a_{ij} is one if there exists an edge from vertex i to vertex j, and zero otherwise.

To target an individual's closeness, an attacker would maximize the length of that individual's shortest path to all others in the network. This can be shown mathematically using the shortest path network interdiction formulation and algorithm discussed in Section 2.5.1. The following formulation maximizes the shortest $s-t$ path length in a directed network by interdicting arcs. This approach is based on the work of Israeli and Wood [71].

$$\begin{aligned}
&\max_{\delta\in\Delta} \min \sum (c_{ij} + \delta_{ij} d_{ij}) x_{ij} \\
&\sum_{j:(i,j)\in A} x_{ij} - \sum_{j:(j,i)\in A} x_{ji} = \begin{cases} 1, & \text{for } i = s; \\ 0, & 0\ \forall i \in N\{s,t\}; \\ -1, & \text{for } i = t. \end{cases} \\
&x_{ij} \geq 0 \ \ \forall k \in A
\end{aligned} \tag{4.4}$$

where $\Delta = \{\delta \in \{0,1\}^{|A|} | r^T \delta \leq D\}$; c_{ij} is the nominal integer length of arc (i,j); d_{ij} is the added integer delay if arc (i,j) is interdicted, $x_{ij} = 1$ if arc ij is traversed in the shortest path $(= 0$ otherwise); $\delta = 1$ if arc (i,j) is interdicted; r_{ij} is the resource required to interdict arc (i, j), and D is the total amount of interdiction resource available.

By first fixing x, taking the dual of the inner minimization problem, making "a few simple modifications" and releasing x, the following formulation results: [70, p. 19]

$$\begin{aligned} \max_{\delta,\pi} \;\; & \pi_t - \pi_s \\ \text{s.t.} \;\; & \pi_j - \pi_i - d_{ij}\delta_{ij} \leq c_{ij} \\ & \pi_s = 0 \\ & \delta \in \Delta \end{aligned} \tag{4.5}$$

This formulation gives the optimal attack against the shortest path between two individuals subject to an interdictor's resources. However, an individual's closeness is defined in terms of shortest paths between an individual and all other members of the network. Therefore, an attacker wishes to interdict all paths from an individual to all other members. The formulation in (4.5) must be modified to sum interdictions across these paths, subject to the interdictor's resource constraints. Since the source does not change in any of the shortest-paths, the constraints remain unchanged. It follows then that the objective functions can be modified to sum across all sinks/members of the network. This is done as follows:

$$\begin{aligned} \max_{\delta} \;\; & \sum_{i \in N \setminus s} \pi_i \\ \text{s.t.} \;\; & \pi_j - \pi_i - d_{ij}\delta_{ij} \leq c_{ij} \\ & \pi_s = 0 \\ & \delta \in \Delta \end{aligned} \tag{4.6}$$

It is noted, however, that this formulation is for directed arcs, while social networks are usually depicted with undirected arcs. To facilitate modeling, each undirected arc in the social network is replaced with 2 directed arcs in opposing directions. If the distance/strength of relationship and cost to interdict is truly symmetric (the same going both directions), then the following constraint can be used:

$$\delta_{i,j} = \delta_{j,i} \tag{4.7}$$

This constraint forces an interdictor to interdict both directions of an interdicted arc, or neither of them. Note: interdiction costs and resources should be matched accordingly. In this special case, a modification of Dijkstra's algorithm developed by Khachiyan *et al.* could be used to solve this problem. The algorithm developed by Khachiyan *et al.* maximizes the shortest path from all nodes to a particular node. [74, p. 4] However, if the paths are not symmetric, then the paths from a node to all others would be different than paths from all nodes to a particular node. In this case (or in any other variation discussed in this chapter), the modified Dijkstra's algorithm could not be used.

If an interdictor has the option or desire (and ability) to interdict one direction of a relationship, but not the other, then constraint (4.7) can be dropped. The model is flexible enough to handle either situation. In any case, the formulation determines which relationships should be targeted to maximally disrupt an individuals closeness centrality. To illustrate this, an example is provided in Section 4.4.

4.3.2 Cells

If an attacker wished to maximize the disruption to an entire cell/network, then the goal would be to maximize G as defined in (4.2). This is done by maximizing the shortest distances between all pairs of nodes. Therefore, the formulation developed in (4.6) will be modified. The formulation, (4.6), finds the interdiction across all paths from a specified node to all other nodes. Since this formulation already sums across all paths from a specified node, the modification simply needs to sum these paths/interdications across all nodes (as the specified node). Specifically, the following formulation results:

$$
\begin{aligned}
\max_{\delta} & \sum_{k\in N} \sum_{i\in N\setminus s} \pi_{ki} \\
\text{s.t. } & \pi_{kj} - \pi_{ki} - d_{ij}\delta_{ij} \le c_{ij} \\
& \pi_{k,i_s} = 0 \\
& \delta \in \Delta
\end{aligned}
\tag{4.8}
$$

Since G relies on an all-pairs shortest path, this interdiction formulation is also an all-pairs shortest path formulation. In any case, this formulation provides the relationships which should be targeted to maximally disrupt the group's closeness centrality, G. To illustrate this, an example is provided in Section 4.4.

4.3.3 Nodal Extensions

The closeness interdiction models above determine the optimal relationships/edges to influence/interdict to optimally disrupt the social networks (as measured by closeness centrality). As discussed in Chapter III, it may also be desired/necessary to model nodal interdiction (which in this case would be individuals in the social network), instead of / in addition to, edge interdiction (in this case the relationships).

Recall from Chapter III that a node-only max flow interdiction modifies a traditional edge interdiction formulation by forcing interdiction of a node to disrupt flow in (all/some) associated edges. In addition, the resource constraint must be modified to allow node interdiction as well. If only nodal interdiction is allowed, then the resource constraint is summed over all interdictable nodes. However, if nodes and edges are allowed, then modifications must be made to ensure edge disruptions associated with interdicted nodes are not double counted.

Although Chapter III focused on maximum flow nodal interdiction, the same idea is applicable in shortest-path nodal interdiction. In this case, instead of a nodal interdiction disrupting flow in associated arcs, it extends the length of the path of associated relationships. In other words, if an individual is selected for targeting/influence, then all (or a selected subset of) connected relationships are disrupted by having the relationship weakened (by being lengthened). Specifically, for individual (node only) targeting, the model is as follows:

$$
\begin{aligned}
\max_{\delta \in \Delta} & \sum_{i \in N \setminus s} \pi_i \\
\text{s.t. } & \pi_j - \pi_i - d_{ij}\delta_{ij} \leq c_{ij} \\
& \sum_i r_i \delta_i \leq D \\
& \delta_{ij} = \delta_i \\
& \delta_{ji} = \delta_i \\
& \delta_{ij} \in \{0,1\} \quad \forall (i,j) \in E \\
& \delta_i \in \{0,1\} \quad \forall i \in N
\end{aligned}
\tag{4.9}
$$

where all variables are as defined previously.

Formulation (4.9) is a model in which only nodes can be targeted. This formulation includes the additional constraints $\delta_{ij} = \delta_i$ and $\delta_{ji} = \delta_i$ which force all edges associated with an interdicted node to also be disrupted. The resource constraint is also appropriately modified. By modifing these constraints, one could also model the case where only a subset of associated edges are disrupted when a node is disrupted. Of course, additional variations are possible. As discussed in Chapter III, this could include formulations where nodes and edges are targetable, with the same or different resources.

Similarly, modifications can be made to the cell closeness interdiction formulation to allow interdiction of nodes. With the same modifications, the (node only) model for targeting the closeness of cellular social networks is as follows:

$$
\begin{aligned}
\max_{\delta\in\Delta} & \sum_{k\in N}\sum_{i\in N\setminus s}\pi_i \\
\text{s.t. } & \pi_j - \pi_i - d_{ij}\delta_{ij} \le c_{ij} \\
& \sum_i r_i\delta_i \le D \\
& \delta_{ij} = \delta_i \\
& \delta_{ji} = \delta_i \\
& \delta_{ij} \in \{0,1\} \quad \forall (i,j)\in E \\
& \delta_i \in \{0,1\} \quad \forall i \in N
\end{aligned}
\tag{4.10}
$$

where, again, all variables are as defined previously. In addition, straightforward modifications can be made to make nodes and edges targetable (either with shared or different resources).

4.3.4 Solution methodologies

Formulations such as (4.6) and (4.8) can be solved directly as mixed-integer problems. However, because of their special structures, it may be beneficial to consider decomposing or partitioning the formulations to take advantage of their structures. This section demonstrates how this could be done using Benders' partitioning.

4.3.4.1 Individuals

For fixed interdictions, δ_{ij}, the linear relaxation of (4.6) is a dual of a (sum across) shortest path(s) which has an intrinsically integer solution. Therefore, the dual can be taken which results in the following program:

$$\begin{gathered}\max_{\delta\in\Delta}\min\sum_{l\in N\setminus s}\sum(c_{ij}+\delta_{ij}d_{ij})x_{lij}\\ \sum x_{lij}-\sum x_{lji}=\begin{cases}1, & \text{for } i=s;\\ 0, & 0\ \forall i\in N\{s,t\};\\ -1, & \text{for } i=t.\end{cases}\\ x_{lij}\geq 0\ \ \forall k\in A\\ y_{ij}\in\{0,1\}\ \ \forall i\end{gathered}\tag{4.11}$$

where l is the set of all nodes, excluding the source node (i.e. the inner objective function sums across all shortest paths from the source to all other nodes). This inner formulation can be used to form the subproblem in a Benders' partitioning. For fixed interdictions, $\hat{\delta_{ij}}$, the subproblem is then

$$\begin{gathered}\min\sum_{l\in N\setminus s}\sum_{(i,j)\in A}(c_{ij}+\hat{\delta_{ij}}d_{ij})x_{lij}\\ \sum_{j:(ij)\in A}x_{lij}-\sum_{j:(ji)\in A}x_{lji}=\begin{cases}1, & \text{for } i=s;\\ 0, & 0\ \forall i\in N\{s,t\};\\ -1, & \text{for } i=t.\end{cases}\\ x_{lij}\geq 0\ \ \forall k\in A\\ y_{ij}\in\text{binary}\end{gathered}\tag{4.12}$$

Therefore, the associated master problems becomes:

$$\begin{gathered}\max_{\delta\in\Delta} z\\ z\leq\sum_{l\in N\setminus s}\delta_{ij}d_{ij}\hat{x_{lij}}\\ \sum_{(i,j)\in A}r_{ij}\delta_{ij}\leq R\\ y_{ij}=y_{ji}\\ y_{ij}\in\text{binary}\end{gathered}\tag{4.13}$$

With this subproblem and master problem, Benders' partitioning can be used as follows:

1. Solve Benders' subproblem as an LP with fixed interdiction (with solution z)
2. Set UB = max(UB, z)
3. Fix flow x_{lij}, and add cut to master problem
4. Solve master problem with solution z
5. Update LB = z
6. Check for convergence: (UB-LB) $\leq \epsilon$
7. If not converged, fix interdiction from master problem and resolve subproblem

4.3.4.2 Cells

In a similar manner, the formulation for all-pairs shortest path interdiction (as used for group closeness interdiction) can be decomposed using Benders' partitioning. The primary addition is that each node is also a source. Therefore, the objective function of the subproblem sums across all shortest paths from each node to all other nodes. This is formulated as follows:

$$\min \sum_{s \in N} \sum_{l \in N \setminus s} \sum (c_{ij} + \hat{\delta_{ij}} d_{ij}) x_{slij} \tag{4.14}$$

$$\sum x_{slij} - \sum x_{slji} = \begin{cases} 1, & \text{for } i = s; \\ 0, & 0 \ \forall i \in N\{s,t\}; \\ -1, & \text{for } i = t. \end{cases}$$

$$x_{slij} \geq 0 \quad \forall k \in A$$

$$y_{ij} \in \text{binary}$$

In addition, the master problem is similarly modified to include this summation across all nodes as sources:

$$
\begin{aligned}
&\max_{\delta \in \Delta} z \\
&z \leq \sum_{n=s \in N} \sum_{l \in N \setminus s} \delta_{ij} d_{ij} \hat{x_{lij}} \\
&\sum_{(i,j) \in A} r_{ij} \delta_{ij} \leq R \\
&y_{ij} = y_{ji} \\
&y_{ij} \in \{0, 1\}
\end{aligned}
\tag{4.15}
$$

With this modified subproblem and master problem, the Benders' algorithm described in the previous section can be used to find the optimal solution.

4.4 Notional Examples

In this section, the formulations developed in the chapter are demonstrated on the notional network in Figure 4.4 which contains is an example of a very small social network. The first number on each arc is c_{ij}, the length of the path (i.e. the strength of the relationship where lower numbers are closer/better); the second number is r_{ij}, the cost to disrupt that relationship. These “costs” can be any resource required to disrupt a relationship; for example, PSYOP messages. We assume the number of these resources is limited (in this example to 4). The third number is d_{ij}, the anticipated amount the relationship is diminished (the amount the distance between them increases) if it is disrupted.

Before any attacks are made against the relationships of this social network, the individual closeness centrality measures are given in Table 4.1. To illustrate how these are calculated, consider Node 1. The shortest distance from Node 1 to Node 2 is 1 unit, from Node 1 to Node 3 is 2 units, from Node 1 to node 4 is 1 unit, and from Node 1 to Node 5 is 2 units. Therefore, the sum across all shortest paths from Node 1 to all other nodes is 6 units. Closeness centrality is the reciprocal of this number or $\frac{1}{6}$.

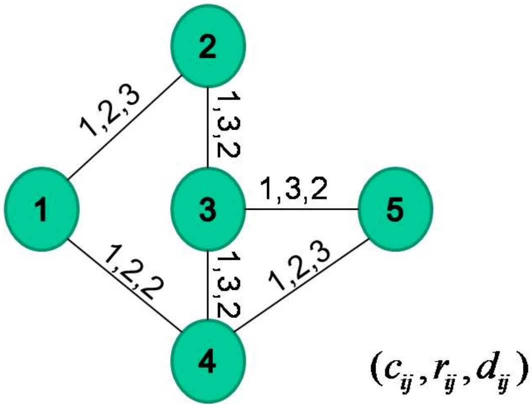


Figure 4.4: Notional Network

Table 4.1: Closeness Centrality

Node	Pre attack
Node 1	$\frac{1}{6}$
Node 2	$\frac{1}{6}$
Node 3	$\frac{1}{5}$
Node 4	$\frac{1}{5}$
Node 5	$\frac{1}{6}$

4.4.1 Individual Attacks

Suppose one wished to disrupt the social network in Figure 4.4 such that Node 1's closeness centrality is maximally disrupted. This is done by using the formulation in (4.6). This results in the following formulation:

$$
\begin{aligned}
\max \quad & \pi_2 + \pi_3 + \pi_4 + \pi_5 \\
\text{s.t.} \quad & -\pi_1 + p_2 - 3\delta_{1,2} \leq 1 \\
& -\pi_1 + \pi_4 - 2\delta_{1,4} \leq 1 \\
& \pi_1 - \pi_2 - 3\delta_{2,1} \leq 1 \\
& -\pi_2 + \pi_3 - 2\delta_{2,3} \leq 1 \\
& \pi_2 - \pi_3 - 2\delta_{3,2} \leq 1 \\
& -\pi_3 + \pi_4 - 2\delta_{3,4} \leq 1 \\
& -\pi_3 + \pi_5 - 2\delta_{3,5} \leq 1 \\
& \pi_1 - \pi_4 - 2\delta_{4,1} \leq 1 \\
& \pi_3 - \pi_4 - 2\delta_{4,3} \leq 1 \\
& \pi_4 + \pi_5 - 3\delta_{4,5} \leq 1 \\
& \pi_3 - \pi_5 - 2\delta_{5,3} \leq 1 \\
& \pi_4 - \pi_5 - 3\delta_{5,4} \leq 1 \\
& \delta_{1,2} - \delta_{2,1} = 0 \\
& \delta_{1,4} - \delta_{4,1} = 0 \\
& -\delta_{1,2} + \delta_{2,1} = 0 \\
& \delta_{2,3} - \delta_{3,2} = 0 \\
& -\delta_{2,3} + \delta_{3,2} = 0 \\
& \delta_{3,4} - \delta_{4,3} = 0 \\
& \delta_{3,5} - \delta_{5,3} = 0 \\
& -\delta_{1,4} + \delta_{4,1} = 0 \\
& -\delta_{3,4} + \delta_{4,3} = 0 \\
& \delta_{4,5} - \delta_{5,4} = 0 \\
& -\delta_{3,5} + \delta_{5,3} = 0 \\
& -\delta_{4,5} + \delta_{5,4} = 0 \\
& 2\delta_{1,2} + 2\delta_{1,4} + 2\delta_{2,1} + 3\delta_{2,3} + 3\delta_{3,2} + 3\delta_{3,4} + 3\delta_{3,5} \\
& \qquad + 2\delta_{4,1} + 3\delta_{4,3} + 2\delta_{4,5} + 3\delta_{5,3} + 2\delta_{5,4} \leq 8 \\
& \pi_1 = 0
\end{aligned}
\tag{4.16}
$$

The optimal solution to this program results in $\delta_{1,2} = \delta_{1,4} = 1$ (since relationships were symmetric, the reverse relationships are also selected for disruption, but for simplicity are not repeated here), which means relationships between node 1 and node 2 should be disrupted (influenced), as well as the relationships between nodes

1 and 4. With these disruptions, the sum of the shortest paths from Node 1 to all other nodes increases to 15. This reduces Node 1's closeness centrality to $\frac{1}{15}$.

Repeating this analysis for each node in the network depicted in Figure 4.4 has the results depicted in Table 4.2.

Table 4.2: Individual Closeness Centrality Disruption

Node	Pre attack	Relationships attacked	Post Attack
Node 1	$\frac{1}{6}$	$(1,2),(1,4)$	$\frac{1}{15}$
Node 2	$\frac{1}{6}$	$(1,2),(1,4)$	$\frac{1}{9}$
Node 3	$\frac{1}{5}$	$(2,3)$	$\frac{1}{7}$
Node 4	$\frac{1}{5}$	$(1,4),(4,5)$	$\frac{1}{8}$
Node 5	$\frac{1}{6}$	$(1,4),(4,5)$	$\frac{1}{8}$

These formulations were solved directly as a mixed integer program, and also with Benders' partitioning (both are described in Section 4.3.4). The solutions using both methods were the same, although solution times differed. For the small network, the mixed integer program solved in 0.013 seconds, while the Benders' partitioning version took 19.569 seconds. It should be noted that the Benders' version found the optimal solution almost immediately; however, it took several cuts to establish optimality. It is not uncommon for smaller problems to solve more quickly directly than the decomposed approach. [118, p. 810]

4.4.2 Group Attack

Suppose one wished to disrupt the social network in Figure 4.4 such that all closeness measures are maximally disrupted, subject to an attacker's resource constraints. This is done by using the formulation in (4.8). This results in the following formulation:

$$
\begin{array}{llllr}
\max & \pi_{1,2}+\pi_{1,3}+\pi_{1,4}+\pi_{1,5}+ & \pi_{2,1}+\pi_{2,3}+\pi_{2,4}+\pi_{2,5}+\pi_{3,1}+\pi_{3,2}+\pi_{3,4}+\pi_{3,5}+ & \pi_{4,1}+\pi_{4,2}+\pi_{4,3}+\pi_{4,5}+\pi_{5,1}+\pi_{5,2}+\pi_{5,3}+\pi_{5,4} & \\
\text{s.t.} & -\pi_{1,1}+p_{1,2} & & & -3\delta_{1,2}\leq 1\\
 & -\pi_{1,1}+\pi_{1,4} & & & -2\delta_{1,4}\leq 1\\
 & \pi_{1,1}-\pi_{1,2} & & & -3\delta_{2,1}\leq 1\\
 & -\pi_{1,2}+\pi_{1,3} & & & -2\delta_{2,3}\leq 1\\
 & \pi_{1,2}-\pi_{1,3} & & & -2\delta_{3,2}\leq 1\\
 & -\pi_{1,3}+\pi_{1,4} & & & -2\delta_{3,4}\leq 1\\
 & -\pi_{1,3}+\pi_{1,5} & & & -2\delta_{3,5}\leq 1\\
 & \pi_{1,1}-\pi_{1,4} & & & -2\delta_{4,1}\leq 1\\
 & \pi_{1,3}-\pi_{1,4} & & & -2\delta_{4,3}\leq 1\\
 & \pi_{1,4}+\pi_{1,5} & & & -3\delta_{4,5}\leq 1\\
 & \pi_{1,3}-\pi_{1,5} & & & -2\delta_{5,3}\leq 1\\
 & \pi_{1,4}-\pi_{1,5} & & & -3\delta_{5,4}\leq 1\\
 & & -\pi_{2,1}+\pi_{2,2} & & -3\delta_{1,2}\leq 1\\
 & & -\pi_{2,1}+\pi_{2,4} & & -2\delta_{1,4}\leq 1\\
 & & \pi_{2,1}-\pi_{2,2} & & -3\delta_{2,1}\leq 1\\
 & & -\pi_{2,2}+\pi_{2,3} & & -2\delta_{2,3}\leq 1\\
 & & \pi_{2,2}-\pi_{2,3} & & -2\delta_{3,2}\leq 1\\
 & & -\pi_{2,3}+\pi_{2,4} & & -2\delta_{3,4}\leq 1\\
 & & -\pi_{2,3}+\pi_{2,5} & & -2\delta_{3,5}\leq 1\\
 & & \pi_{2,1}-\pi_{2,4} & & -2\delta_{4,1}\leq 1\\
 & & \pi_{2,3}-\pi_{2,4} & & -2\delta_{4,3}\leq 1\\
 & & \pi_{2,4}+\pi_{2,5} & & -3\delta_{4,5}\leq 1\\
 & & \pi_{2,3}-\pi_{2,5} & & -2\delta_{5,3}\leq 1\\
 & & \pi_{2,4}-\pi_{2,5} & & -3\delta_{5,4}\leq 1
\end{array}
$$

$$
\begin{aligned}
&-\pi_{3,1}+\pi_{3,2} && -3\delta_{1,2}\leq 1\\
&-\pi_{3,1}+\pi_{3,4} && -2\delta_{1,4}\leq 1\\
&\pi_{3,1}-\pi_{3,2} && -3\delta_{2,1}\leq 1\\
&-\pi_{3,2}+\pi_{3,3} && -2\delta_{2,3}\leq 1\\
&\pi_{3,2}-\pi_{3,3} && -2\delta_{3,2}\leq 1\\
&-\pi_{3,3}+\pi_{3,4} && -2\delta_{3,4}\leq 1\\
&-\pi_{3,3}+\pi_{3,5} && -2\delta_{3,5}\leq 1\\
&\pi_{3,1}-\pi_{3,4} && -2\delta_{4,1}\leq 1\\
&\pi_{3,3}-\pi_{3,4} && -2\delta_{4,3}\leq 1\\
&\pi_{3,4}+\pi_{3,5} && -3\delta_{4,5}\leq 1\\
&\pi_{3,3}-\pi_{3,5} && -2\delta_{5,3}\leq 1\\
&\pi_{3,4}-\pi_{3,5} && -3\delta_{5,4}\leq 1\\
&-\pi_{4,1}+\pi_{4,2} && -3\delta_{1,2}\leq 1\\
&-\pi_{4,1}+\pi_{4,4} && -2\delta_{1,4}\leq 1\\
&\pi_{4,1}-\pi_{4,2} && -3\delta_{2,1}\leq 1\\
&-\pi_{4,2}+\pi_{4,3} && -2\delta_{2,3}\leq 1\\
&\pi_{4,2}-\pi_{4,3} && -2\delta_{3,2}\leq 1\\
&-\pi_{4,3}+\pi_{4,4} && -2\delta_{3,4}\leq 1\\
&-\pi_{4,3}+\pi_{4,5} && -2\delta_{3,5}\leq 1\\
&\pi_{4,1}-\pi_{4,4} && -2\delta_{4,1}\leq 1\\
&\pi_{4,3}-\pi_{4,4} && -2\delta_{4,3}\leq 1\\
&\pi_{4,4}+\pi_{4,5} && -3\delta_{4,5}\leq 1\\
&\pi_{4,3}-\pi_{4,5} && -2\delta_{5,3}\leq 1\\
&\pi_{4,4}-\pi_{4,5} && -3\delta_{5,4}\leq 1
\end{aligned}
$$

$$
\begin{aligned}
&- \pi_{4,1} + \pi_{4,2} && -3\delta_{1,2} \leq 1 \\
&- \pi_{4,1} + \pi_{4,4} && -2\delta_{1,4} \leq 1 \\
&\pi_{4,1} - \pi_{4,2} && -3\delta_{2,1} \leq 1 \\
&- \pi_{4,2} + \pi_{4,3} && -2\delta_{2,3} \leq 1 \\
&\pi_{4,2} - \pi_{4,3} && -2\delta_{3,2} \leq 1 \\
&- \pi_{4,3} + \pi_{4,4} && -2\delta_{3,4} \leq 1 \\
&- \pi_{4,3} + \pi_{4,5} && -2\delta_{3,5} \leq 1 \\
&\pi_{4,1} - \pi_{4,4} && -2\delta_{4,1} \leq 1 \\
&\pi_{4,3} - \pi_{4,4} && -2\delta_{4,3} \leq 1 \\
&\pi_{4,4} + \pi_{4,5} && -3\delta_{4,5} \leq 1 \\
&\pi_{4,3} - \pi_{4,5} && -2\delta_{5,3} \leq 1 \\
&\pi_{4,4} - \pi_{4,5} && -3\delta_{5,4} \leq 1 \\
& && \delta_{1,2} - \delta_{2,1} = 0 \\
& && \delta_{1,4} - \delta_{4,1} = 0 \\
& && -\delta_{1,2} + \delta_{2,1} = 0 \\
& && \delta_{2,3} - \delta_{3,2} = 0 \\
& && -\delta_{2,3} + \delta_{3,2} = 0 \\
& && \delta_{3,4} - \delta_{4,3} = 0 \\
& && \delta_{3,5} - \delta_{5,3} = 0 \\
& && -\delta_{1,4} + \delta_{4,1} = 0 \\
& && -\delta_{3,4} + \delta_{4,3} = 0 \\
& && \delta_{4,5} - \delta_{5,4} = 0 \\
& && -\delta_{3,5} + \delta_{5,3} = 0 \\
& && -\delta_{4,5} + \delta_{5,4} = 0
\end{aligned}
$$

$$\pi_{1,1} = 0$$

$$\pi_{2,2} = 0$$

$$\pi_{3,3} = 0$$

$$\pi_{4,4} = 0$$

$$\pi_{5,5} = 0$$

Table 4.3: Group Closeness Centrality Disruption

Node	Pre attack	Post Attack
Node 1	$\frac{1}{6}$	$\frac{1}{15}$
Node 2	$\frac{1}{6}$	$\frac{1}{9}$
Node 3	$\frac{1}{5}$	$\frac{1}{7}$
Node 4	$\frac{1}{5}$	$\frac{1}{7}$
Node 5	$\frac{1}{6}$	$\frac{1}{8}$
G	$\frac{9}{10}$	$\frac{1483}{2520}$

It should be noted that the constraints are arranged in blocks according to source node, and the coupling variables in the far right column. This is done to illustrate the the underlying structure which is exploited when using decomposition/partitioning techniques.

Using this formulation results in an optimal solution with $\delta_{1,2} = \delta_{1,4} = 1$ meaning that disrupting the relationships (and their reciprocals) between node 1 and nodes 2 and 4 maximally disrupts the closeness centrality of the group (as measured by their sum). This is displayed in Table 4.3.

As Table 4.3 shows, the sum of closeness centrality drops from 0.9 to less than 0.6. Stated another way, the disruption increased the shortest paths across the network by more than one-third.

This formulation was solved directly as a mixed integer program (see Section 4.3) and with Benders' partitioning (see Section 4.3.4). The solution using both methods were identical, although the solution times was radically different. The mixed integer program version solved in 0.016 seconds, while the Benders' partitioning version took 19.751 seconds. It should again be noted that the Benders' version found the optimal solution almost immediately; however, it took several cuts to establish optimality. This is partially a result of the particular implementation of Benders' in GAMS. For example, in each iteration, the subproblems were created from scratch instead of simply being modified based on the results of the master problem.

In addition, while in these examples the Benders' method required more computational effort than the mixed-integer versions, the Benders' algorithm provides information as it is being solved. In a very large network where solving the system as a single MIP is impractical, the Benders' algorithm could be used to provide bounds to the optimal solution as it proceeds. In addition, the shadow prices of the subproblem could be used as a proxy for reduced vulnerability obtained if additional resources could be obtained.

4.4.3 Borgatti Network

Borgatti considered the problem of disrupting social networks by identifying "key players" whose removal would maximally disrupt the network in [23]. To illustrate the problems with traditional centrality measures and promote his algorithm, he provided the example network depicted in Figure 4.5.

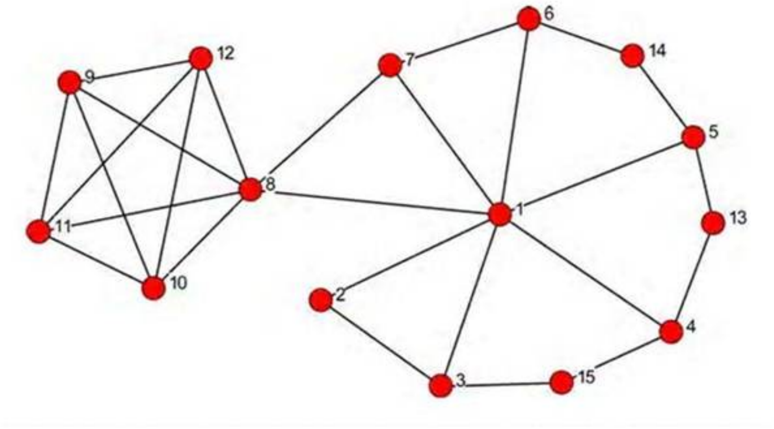


Figure 4.5: Borgatti Example [23, p. 23]

The method developed in this chapter compliments Borgatti's methods. Borgatti only considered the removal of nodes. Although the methods developed in this chapter could be modified to consider node removal (see Chapter III), the focus has been on the "interdiction" of relationships. Relationships/links are not directly explored in Borgatti's methods. To show how this method can complement Borgatti's work, consider the example in Figure 4.5. For simplicity, assume that each relationship can be degraded by one unit with one unit of interdiction resource. Assuming one unit of interdiction resource, the relationship between 1 and 8 is found to be key. Using the group attack formulation, the total closeness interdiction drops from $\frac{1}{486}$ to $\frac{1}{566}$. Using the individual attack formulation, the closeness centrality of Node 1 drops from $\frac{1}{21}$ to $\frac{1}{26}$.

Using the node only interdiction formulations (and excluding node 1 from susceptibility to influence), it is found that Node 8 is the optimal node/person to influence (again, assuming only one unit of influence is available). In the individual attack, this drops Node 1's centrality from $\frac{1}{21}$ to $\frac{1}{25}$. Using the group attack, the total closeness drops from $\frac{1}{486}$ to $\frac{1}{580}$.

This example illustrates a method to determine optimal disruptions using traditional social network measures such as centrality. This method addresses the limitations discussed by Borgatti in [23]. Note that in both nodal formulations, Node 8 was chosen for influence contrary to the examples provided by Borgatti.

4.4.4 Krebs Network

Krebs compiled the social network of the 9-11 hijackers based on open source data. [78] Krebs considered the resulting network and, in his expert opinion, additional edges needed to be added for the network to make sense. To decrease the average paths between the hijackers and increase collaboration among them, Krebs added six additional arcs. The resulting network is depicted in Figure 4.6.

Had this network been known before 9-11, the U.S. would have liked to maximally disrupt this network. With limited resources, the question becomes which relationships are most vital to the group as measured by closeness centrality. As a side note, finding the optimal relationships to interdict serves as a check to see how vital the edges added by Krebs really are.

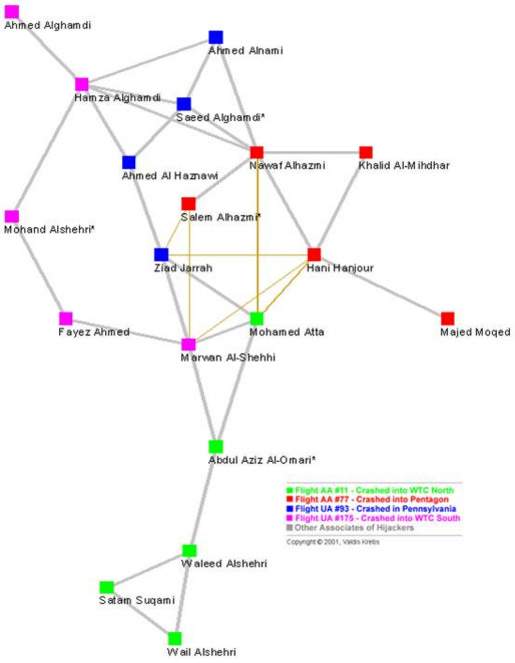


Figure 4.6: Krebs Example [78]

Since the goal would be to disrupt the entire network, not just the relationship of one individual with the rest of the network, the group interdiction formulation is used. For simplicity, it is assumed that all relationships are candidates for influence/disruption and consume one unit of resource. To facilitate a comparison with Krebs' work, six units of influence resources are made available. Finally, it is assumed that if an arc is interdicted (effected), the length of the relationship between two individuals increases by two units. With this, the optimal relationships to interdict (represented by maximizing the shortest path) is found to be the following: Abdul Aziz Al-Omari to Marwan Al-Shehhi, Abdul Aziz Al-Omari to Mohamed Atta, Abdul Aziz Al-Omari to Waleed Alsheri, Ahmed Alghamdi to Hamza Alghamdi, Hani Hanjour to Majed Moqed, Mohamed Atta to Nawaf Alhazmi. These interdictions reduce each individuals centrality as shown in Table 4.4.

As Table 4.4 shows, the individual closeness centrality dropped for every member of the network, and the average closeness dropped by over 23%. Interestingly, to maximally disrupt the network (as measured by closeness centrality), only one

Table 4.4: Krebs' 9-11 Network Centralities

Individual	Closeness Pre-Attack	Closeness Post-Attack
Abdul Aziz Al-Omari	$\frac{1}{42}$	$\frac{1}{67}$
Ahmed Al Haznawai	$\frac{1}{45}$	$\frac{1}{54}$
Ahmed Alghamdi	$\frac{1}{58}$	$\frac{1}{89}$
Ahmed Alnami	$\frac{1}{46}$	$\frac{1}{60}$
Fayez Ahmed	$\frac{1}{46}$	$\frac{1}{55}$
Hamza Alghamdi	$\frac{1}{41}$	$\frac{1}{55}$
Hani Hanjour	$\frac{1}{37}$	$\frac{1}{46}$
Khalid Al-M ihdhar	$\frac{1}{47}$	$\frac{1}{56}$
Majed Moqed	$\frac{1}{54}$	$\frac{1}{80}$
Marwan Al-Shehhi	$\frac{1}{37}$	$\frac{1}{46}$
Mohamed Atta	$\frac{1}{35}$	$\frac{1}{49}$
Mohand Alsheri	$\frac{1}{50}$	$\frac{1}{59}$
Nawaf Alhazmi	$\frac{1}{35}$	$\frac{1}{49}$
Saeed Alghamdi	$\frac{1}{44}$	$\frac{1}{58}$
Salem Alhazmi	$\frac{1}{41}$	$\frac{1}{50}$
Satam Suqami	$\frac{1}{71}$	$\frac{1}{109}$
Wail Alsheri	$\frac{1}{71}$	$\frac{1}{109}$
Waleed Alsheri	$\frac{1}{55}$	$\frac{1}{93}$
Ziad Jarrah	$\frac{1}{39}$	$\frac{1}{48}$
Average	0.022171	0.017066

of the relationships added by Krebs was selected for disruption: Mohamed Atta to Nawaf Alhazmi.

Using the node only formulations, the optimal individuals to influence can be determined. For example, using the group attack method of the node only model, it is found that influencing Abdul Aziz Al-Omari is the optimal, followed by Nawaf Alhazmi.

4.5 Computational Experiments

To demonstrate the potential of these algorithms on various sized networks, random networks were generated using the Erdos-Renyi method using UCINET. [22] Erdos and Renyi's method creates random graphs by joining n nodes by random chosen edges. One way to do this is to specify a desired average degree of nodes. [54,

p. 290] This method was chosen for several reasons. First, this method is prevalent in social network modeling; it is widely used and provided by UCINET. Second, this model has the feature that as the number of nodes increases, the model is likely to undergo a "phase transition" in which a "giant component" forms. [91, p. 2] This is important because the formulations developed in this research assume the social network is connected (i.e. no isolates). In fact, every network generated as a result of this study (including the smaller 10 node networks) was connected.

These formulations were solved using GAMS with the BARON/CPLEX solvers and on a Dell Precision M6300 with 2.50 gigahertz Intel Core2 Duo processor and 4094 megabytes of RAM.

To simplify the illustration, it is assumed in these examples that interdiction of each edge consumes one unit of resource, and if interdicted, the length of an edge increases by one unit. Of course, other numbers could be used, and this was done in experimentation. However, this makes the illustration more difficult to follow.

4.5.1 Individual Attacks

In the randomly generated networks described in Table 4.5, it was assumed that Node 1 was the person to be isolated. Therefore, the algorithms to minimize his individual centrality were used. The network sizes and solution times are reported in Table 4.5.

Table 4.5: Interdiction of an Individual's Centrality/Relationships

# Nodes	# Edges	Interdiction Resources	MIP Time	# Benders' Iterations & Time (H:MM:SS)
10	15	1	0:00:00.410	3 - 0:00:01.487
50	408	10	0:00:00.773	23 - 0:08:54.440
250	1500	20	0:00:01.215	11 - 0:06:19.005
500	3000	20	0:00:10.889	13 - 0:35:10.192

As Table 4.5 shows, moderately large social networks can potentially be solved using either the MIP or Benders' formulations. While it was somewhat surprising that the Benders' solution times were significantly higher than the MIP times, a closer examination of the GAMS output explains this. Almost all time reported for the Benders' solution times is a result of generating the large subproblems. Once the problem was generated, the time to solve each iterated sub/master problem was less than one second in all cases. Future research should be applied to this problem to reduce subproblem size and therefore reduce solution times. Alternatively, a more streamlined generation procedure might be considered.

4.5.2 Group Attack

In the randomly generated networks described in Table 4.6, it was assumed that the goal was to maximally increase the distance between all nodes. Therefore, the algorithm to maximally interdict all-shortest paths was used. The network sizes and solution times are reported in Table 4.5.

Table 4.6: Interdiction of Group's Centrality/Relationships

# Nodes	# Edges	Interdiction Resources	MIP Time	# Benders' Iterations & Time (H:MM:SS)
10	15	1	0:00:00.279	5 - 0:00:03.454
50	408	10	0:00:04.191	9 - 0:09:57.442
250	1500	20	0:04:19.844	14 - 0:39:57.442

As Table 4.6 shows,moderately large social networks can potentially be solved using either the MIP or Benders' formulations. Again, as with individual attacks discussed in Section 4.5.1, the Benders' solution times were significantly greater than respective MIP solution times.

Noticeably absent from Table 4.6 is a row for 500 nodes. GAMS was unable to process social networks of this size because of hardware (memory/RAM) limitations. The "break point" appears to be somewhere around 400 nodes with the given hardware and network specifications. However, this may not be a significant limitation. The terrorist social networks found/reported in open literature are less than 400 nodes; and even with larger networks, the subgroups of networks could be modeled. The largest network in the literature was developed by Dr. Sageman as a terrorist dataset. Therefore, to confirm this size constraint may not be an issue for Sageman's real-world dataset, that set was explored in the next section as an application.

Finally, if a network is found/developed for which this limitation is indeed significant, the formulation/algorithm could be implemented on a larger capacity platform; up to and including high performance computers. It is noted that many U.S. supercomputers already have GAMS installed. These larger formulations are also precisely where the benefits for Benders' partitioning should occur.

4.6 Application

4.6.1 Sageman

In response to the terrorist attacks on September 11, 2001, Sageman began collecting data on Al Qaeda using open source literature. [103, p. vii]. However, this database is based on Dr. Sagemans's 2004 publication and may be dated. Therefore, the analysis in this section is to demonstrate the potential of the methodology, and is limited to the data available. After all the isolates are removed, the remaining connected network from the Sageman database has 366 members and 2422 relationships. This network is depicted in Figure 4.7.

The Sageman database was selected because it is the largest terrorist social network found in the open social networks literature. In addition, this is the type of dataset this research's formulations are designed to be run against; terrorists networks are exactly the types of networks the U.S. would like to disrupt.

For this section, it was assumed that the length of each relationship was one unit, and disruption of a relationship increased this distance by one unit. For demonstration purposes, 5 units of interdiction resource were assumed. In effect, this determines the top 5 relationships to disrupt to maximally disrupt the network as

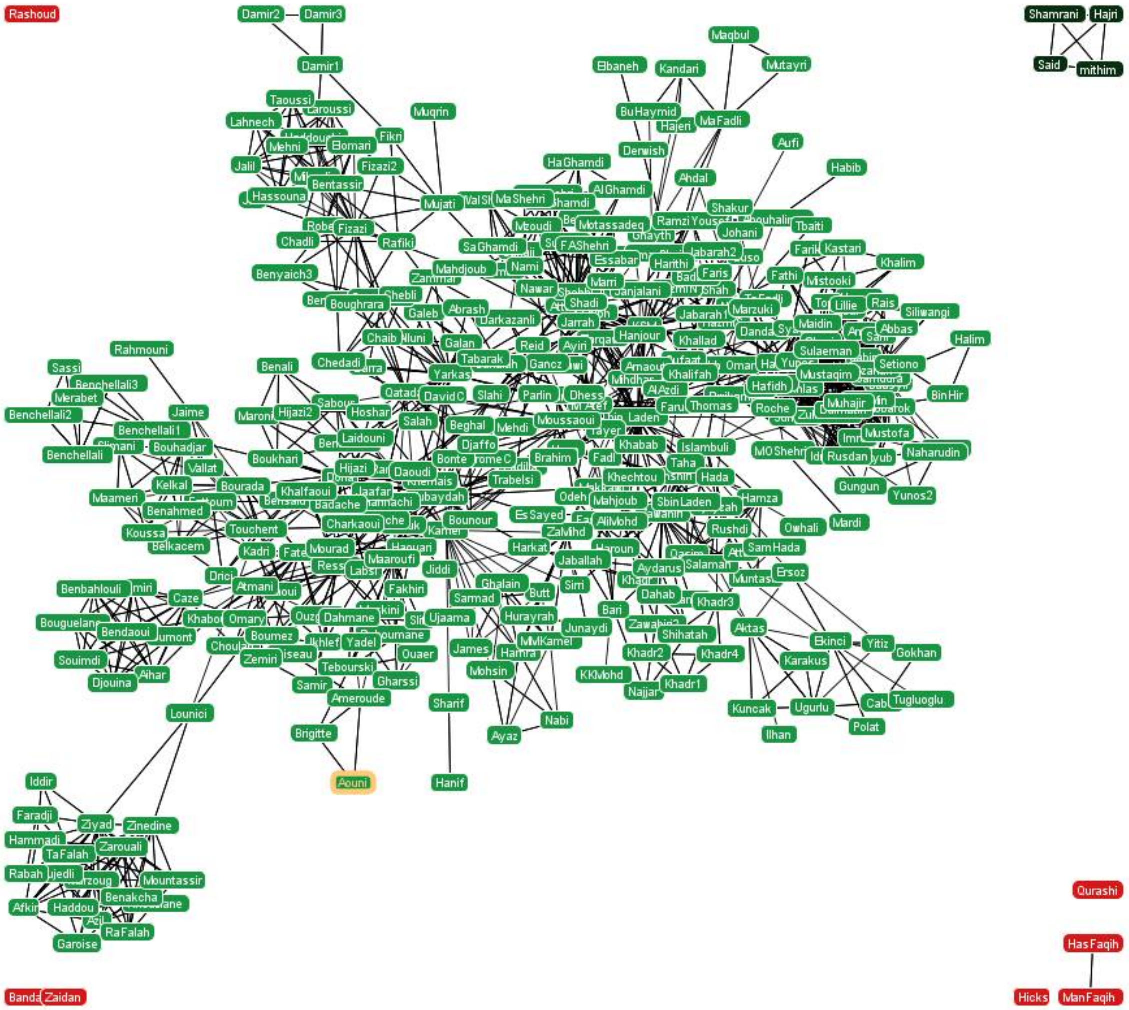


Figure 4.7: Al Qaeda Network from Sageman Database

measured by closeness centrality. Although this is a large network, GAMS solved the group interdiction of this network (with the MIP) in 12 minutes and 18 seconds (including problem generation which took approximately 10 minutes). The resulting solution indicates that the following relationships should be considered for disruption based on network structure: Enaam Arnaout to Muhammad Jamal Khalifah, Khader abu Hoshar to Saed Hijazi, Raed Hijazi to Saed Hijazi, Mohamed Mahjoub to Mahmoud Jaballah, and Mohammad Rais to Sardona Siliwangi.

Interestingly, none of these individuals is considered high profile, so interdiction of these relationships to disrupt the network, as measured by closeness centrality, is a unique result. If one wished to decrease the closeness of an individual high profile person, the individual closeness interdiction formulation should be used.

4.6.2 Communications Network

Krebs showed how lessons and measures from social network analysis can be applied to the infrastructure networks. [77] Specifically, he considered routers in a computer network and showed that "maximizing closeness between *all* routers improves updating and minimizes hop counts." [77, p. 16] Therefore, to disrupt a communications network, one approach would be to minimize the closeness between all components in the network. This can be done using the all-pairs/group interdiction approach. For example, this method can be applied against the communications network taken from Pinkstaff (and used in Section 3.7).

The graph in Figure 4.8 shows the effect of increasing interdiction resources on the sum of the closeness centrality of all communications equipment. As the graph shows, there is a significant increase at about 20 units of resource. Again, interdiction increases the shortest paths between the components, increasing centrality and disrupting centrality as discussed by Krebs. However, some areas of the graph are flat (for example, from 90 to 140) indicating that additional resources have no effect on the shortest paths. In these cases, a large increase in resources is necessary even to make minor changes in shortest paths.

4.7 Summary

The developed methodology extends network interdiction to social networks using traditional SNA measures to identify relationships whose influence would maximally disrupt the network. Along with the associated examples, this demonstrates

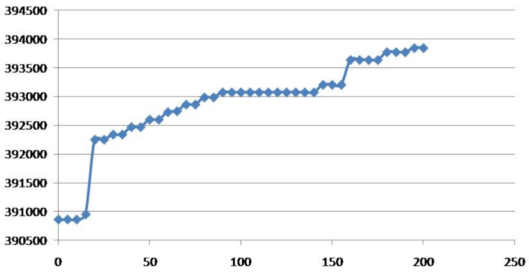


Figure 4.8: Interdiction Resources versus Closeness Centrality

that targeting specific relationships in a social network can reduce individual and overall social closeness as measured by centrality. In addition, it was also shown how these methods could be applied to infrastructure networks, such as a communications network. Depending on the specific mission objectives, and desired effects, this can be an effective approach.

In order to achieve this, shortest-path network interdiction was extended to both individual centrality interdiction and group centrality interdiction. Group centrality interdiction involves formulating an all-pairs shortest path interdiction model, which to date has not been done in the literature. For each of these models, both MIP and Benders partitioning formulations were developed, and computational experiments indicate the method is promising for large networks.

V. Multilayered Network Interdiction

5.1 Introduction

Traditional network models of single functionality do not typically account for the interdependent nature of layered networks. These networks are generally modeled individually, as an isolated network or with minimal recognition of interactions. This chapter develops a methodology to maximize disruptions over the individual networks while explicitly considering their interconnected effects.

As shown in Figure 5.1, multilayer interdiction builds on the concepts of multilayer models and network interdiction. In addition, if nodal attacks are desired, this formulation is further developed and combined with nodal interdiction. It is also shown that this formulation can be decomposed by variable type using Benders' Partitioning and solved to optimality using a Benders' partitioning algorithm. Finally, these new models and solution techniques are demonstrated using notional examples and computational experiments.

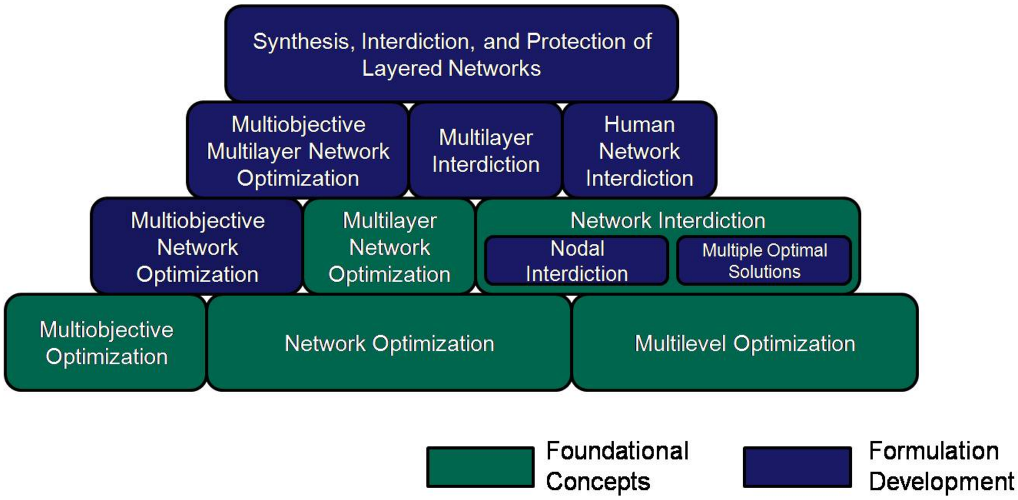

Synthesis, Interdiction, and Protection of Layered Networks
Multiobjective Multilayer Network Optimization
Multilayer Interdiction
Human Network Interdiction
Multiobjective Network Optimization
Multilayer Network Optimization
Network Interdiction
Nodal Interdiction
Multiple Optimal Solutions
Multiobjective Optimization
Network Optimization
Multilevel Optimization
Foundational Concepts
Formulation Development

Figure 5.1: Multilayer Interdiction Formulation

Wood developed techniques to model and solve maximum flow network interdiction problems. [137] This formulation (discussed in Section 2.5.2) is shown below:

$$\begin{aligned}
\min_{\alpha,\beta,\gamma} & \sum_{(i,j)\in A} u_{ij}\beta_{ij} \\
\text{s.t. } & \alpha_i - \alpha_j + \beta_{ij} + \gamma_{ij} \geq 0 \ \ \forall \ \ (i,j) \in A \\
& \alpha_t - \alpha_s \geq 1 \\
& \sum_{(i,j)\in A} r_{ij}\gamma_{ij} \leq R \\
& \alpha_i, \beta_{ij}, \gamma_{ij} \in \{0,1\}
\end{aligned} \tag{5.1}$$

where $\{N_s, N_t\}$ is a cutset partition, $\alpha_i = 1$ indicates node $i \in N_t$, $\alpha_i = 0$ indicates node $i \in N_s$, $\gamma_{ij} = 1$ if arc ij is a forward arc of the cutset and is interdicted (otherwise $\gamma_{ij} = 0$), $\beta_{ij} = 1$ if arc ij is a forward arc of the cutset but is not interdicted (otherwise $\beta_{ij} = 0$), r_{ij} is the resource required to interdict ij, and R is the total resource available for interdiction.

This model is robust and has been extensively studied; however, the variations currently available in the literature only consider single layers of networks. Real-world systems of networks, however, are often more complex. For example, infrastructure networks can be viewed as interdependent layers of networks. Therefore, in this effort, Wood's model is extended to account for layered effects similar to the network model developed by Kennedy (and discussed in Section 2.4).

5.2 Multilayered Network Model

Kennedy *et al.* developed a model to determine minimum cost cut-sets across multiple layers of networks. [73] This model finds the minimum cost (or maximum benefit) of a combined s,t-cut across all individual networks and shared elements by determining an overall cut set as driven by objective function. Before this formulation is introduced and developed, some additional variable definitions are explained.

First, let I be a node or arc(s) with common interdependencies across k networks. That is to say, I has common elements in all layers of the k networks of interest, or in some subset of the layers. In addition, let W_i be the set of all effects options, w, which can be applied to the elements in I. The option w may affect all the elements in I, or it may affect a subset of I.

Associated with each option against a particular interdependent element is the actual effect. For a given I and w_i, let δ_{ik} be the change (effect) on node i of network k given the selection of w_i. Define δ_{ijk} to be the change (effect) on arc (i, j) of network k given the selection of w_i. Of course, the effect may be zero in some or all networks. In addition, affecting a node could also affect a number of arcs.

$$\delta_{(ijk),w} = \begin{cases} 1, \text{ if arc}(i,j) \text{ of network } k \text{ is affected by option } w_i \in W \\ 0, \text{ otherwise} \end{cases} \tag{5.2}$$

For a given I, it is assumed $y_{w_i} = 1$ if option w_i is selected and zero otherwise.

$$y_{w_i} = \begin{cases} 1, \text{ if } w_i \in W_{\mathcal{I}} \text{ chosen} \\ 0, \text{ otherwise.} \end{cases} \tag{5.3}$$

In a targeting model, it can also be assumed that one would not wish to double strike a target, or w could be the level of strike(s) required (at least in initial planning). Therefore, at most, one of the common attack options w is selected, leading to the constraint:

$$\sum_{w \in W_I} y_w \leq 1 \quad \forall I \in C \tag{5.4}$$

where C is the set of all commonalities I. Finally, $\mathbb{C}_w$, represents the relative cost of cutting the interdependent arcs associated with using option w. This leads to the following model:

$$
\begin{aligned}
\min \sum_{k \in K} \sum_{(i,j) \in A_k} & c_{ijk} \nu_{ijk} + \sum_{I \in C} \sum_{w \in W_I} \mathbb{C}_w y_w \\
\text{s.t. } & \pi_{i_k} - \pi_{j_k} + \nu_{ijk} + \delta_{ijkw} y_w \geq 0 \\
& \pi_{t_k} - \pi_{s_k} \geq 1 \\
& \sum_{I \in C} \sum_{w \in W_I} y_w \leq 1 \\
& \pi, \nu, y \in \{0, 1\}
\end{aligned}
\tag{5.5}
$$

where c_{ijk} is the flow capacity along arc (i,j) of network k, ν_{ijk} is the dual variable associated with the capacity constraint of arc (i,j) of network k, π_{i_k} is the dual variable associated with the conservation of flow for node i of network k, s_k is the source node for network k, t_k is the sink node for network k.

In order to extend this minimum cost cut-set model to an interdiction model, it is first necessary to convert the notation to be consistent with interdiction literature. The variable names are therefore renamed as demonstrated in Table 5.1.

Table 5.1: Variable Naming Substitutions

	Layered Notation	Interdiction Notation
Flow capacity	c_{ijk}	u_{ijk}
Capacity constraint dual	ν_{ijk}	β_{ijk}
Node dual	π_{i_k}	α_{i_k}
Selection of interdependency	y_w	γ_w
Interdependent cost	$\mathbb{C}_w$	r_w

Making these variable naming substitutions, the following model results:

$$
\begin{aligned}
\min \sum_{k\in K} \sum_{(i,j)\in A_k} & u_{ijk}\beta_{ijk} + \sum_{I\in C} \sum_{w\in W_I} r_w\gamma_w \\
\text{s.t. } & \alpha_{i_k} - \alpha_{j_k} + \beta_{ijk} + \delta_{ijkw}\gamma_w \geq 0 \\
& \alpha_{t_k} - \alpha_{s_k} \geq 1 \\
& \sum_{I\in \mathbb{C}} \sum_{\gamma\in W_I} \gamma_w \leq 1 \\
& \alpha, \beta, \gamma \in \{0,1\}
\end{aligned}
\tag{5.6}
$$

where u_{ijk} is the flow capacity along arc (i,j) of network k, β_{ijk} is the dual variable associated with the capacity constraint of arc (i,j) of network k, α_{i_k} is the dual variable associated with the conservation of flow for node i of network k, s_k is the source node for network k, t_k is the sink node for network k.

5.3 Multilayered Network Interdiction

A quick comparison reveals that the traditional network interdiction model (formulation (5.1)) and the layered network cut-set model (formulation 5.6) are similar. In order to facilitate the extension of interdiction to multiple layers, it is helpful to have a conceptual understanding of how the maximum flow network interdiction model (formulation 5.1) works. This model sets nodes on either side of the cut, and it sets $\gamma_{i,j} = 1$ or $\beta_{i,j} = 1$ for forward arcs across the cut, to satisfy the main "dual" equation.

In other words, the cuts are identified by setting the α_i and α_j values. The model evaluates the capacity of the cut with the β_{ij} variables. However, the attacker/interdictor can avoid paying for some of the capacity (which would normally allow flow through) by interdicting arc (i,j) via $\gamma_{i,j} = 1$. That is, the $\gamma_{i,j}$ variables behave like the $\beta_{i,j}$ variables, except that the $\gamma_{i,j}$ do not have "costs" associated with them. Unfortunately for the attacker/interdictor, only a limited supply of γ's are available to be set to one (because of the attacker resource constraint).

The layered network cut-set model (formulation 5.6) works in a similar manner. As a first step in converting this formulation to an interdiction model, an interdiction variable is added to the dual constraints for each edge in the network. Of course, these interdictions do not count against the capacity or "costs," but are subject to the resource constraint of the attacker. In addition, the "cost" of selection of an effect γ_w is moved from the objective function to the resource constraint. In network interdiction, the primary objective is not to minimize (interdiction) cost. Instead, interdiction cost is converted into a constraint which is limited by the availability of the resource, R. In addition, in the layered network formulation (5.6), it was assumed that a target would not be attacked multiple times. Since minimizing cost is no longer a primary concern (so long as the resource constraint is not violated), this restriction can be dropped, if desired. Of course, in circumstances where it is important not to strike a target multiple times (for whatever reason), this restriction should be retained. Otherwise, other constraints make the constraint limiting attacks to single strikes redundant and unnecessary.

The commonality variables remain largely unchanged. Selection of a common effect works much like the selection of an interdiction variable. In each case, selection of the variable stops flow through the edge, and associated "costs" are limited by the resource constraint, not the objective function. The commonality variable type accounts for interdiction across multiple networks with a single cost. Therefore, the model determines which elements across the layers of networks should be interdicted to maximize disruption across the system of networks.

With the discussed modifications, the single level layered network interdiction model is as follows:

$$
\begin{aligned}
\min \sum_{k\in K} \sum_{(i,j)\in A_k} & u_{ijk}\beta_{ijk} \\
\text{s.t. } & \alpha_{i_k} - \alpha_{j_k} + \beta_{ijk} + \gamma_{ij} + \sum_{w} \delta_{ijkw}\gamma_w \geq 0 \\
& \alpha_{t_k} - \alpha_{s_k} \geq 1 \\
& \sum_{(i,j)\in A} r_{ij}\gamma_{ij} + \sum_{w\in W_I} r_w\gamma_w \leq R \\
& \alpha_{i_k} \in \{0,1\} \quad \gamma_w \in \{0,1\} \quad \beta_{ijk} \in \{0,1\}
\end{aligned}
\tag{5.7}
$$

When considering a system of layered networks as a holistic system, it is important to use commensurate units. Traditionally, networks are considered in isolation, partly because each network usually serves a specific purpose. For example, consider infrastructure layers as an example. Water, energy, and telecommunications all have different types of flows across their networks. Although these networks are connected, the material that is flowing does not cross networks (water never uses electrical lines for transport). Instead, interdependencies are created through other means as discussed in Section 2.2.8.1. For example, the water infrastructure requires electricity to power its pumps, SCADA systems, and so forth. In addition, water lines and electrical lines may cross the same bridge creating a geographic dependency.

Therefore, when considering the system of networks as a whole, the units must be scaled and/or normalized. For example, in considering a multi-modal system of transportation networks, a commensurate unit that could be used across all layers would be tonnage moved per unit of time. Another common unit used across multiple layers is cost/dollars. An additional option would be to use the approach by Wallace *et al.* and use a binary variable to represent connectivity of a critical network system without regard to units of physical flow. As discussed in (2.4.0.5), this allows the networks to retain their non-commensurate units, but still captures their interdependencies. Either approach (using commensurate units or binary variables) could be used with the developments in the section; but for consistency, commensurate units are assumed.

5.4 Benders' Partitioning

As with formulations 5.1 and 5.6, Benders' partitioning is applied to develop a master problem and subproblem (which is similar to the development in [46]).

For fixed interdictions, γ_{ijk} and y_w, the linear relaxation of (5.7) is a dual of a network flow problem which has an intrinsically integer solution. Therefore, the dual can be taken which results in the following program:

$$\begin{aligned} \min_{\gamma\in\Gamma, w\in W} \max & \sum_k x_{tsk} \\ \text{s.t.} & \sum_{j:(i,j)\in A_k} x_{ijk} - \sum_{j:(j,i)\in A_k} x_{jik} = 0 \\ & x_{ijk} \leq u_{ijk}(1-\gamma_{ijk})(1-y_w\delta_{ijkw}) \end{aligned} \tag{5.8}$$

where $\gamma_{ijk} : \gamma \in \{0,1\}$, $\sum\limits_{(i,j)\in A_k} r_{ijk}\gamma_{ijk} + \sum\limits_{w} r_w y_w \leq R$. If all extreme points of the inner maximization are enumerated, and the solution with the minimum value subject to $\gamma \in \Gamma$ and $w \in W$ is selected, then the model becomes

$$\min_{\gamma\in\Gamma, w\in W} \max_{x^l\in X} x^l_{tsk} - \sum_{(i,j)\in A_k} x^l_{ijk}\gamma_{ijk} - \sum_w \sum_{(i,j)\in A_k} x^l_{ijk} y_w \delta_{ijkw} \tag{5.9}$$

Alternatively, this can now be written as the following subproblem:

$$\begin{aligned} \max & \sum_k x_{tsk} \\ \text{s.t.} & \sum_{j:(i,j)\in A_k} x_{ijk} - \sum_{j:(j,i)\in A_k} x_{jik} = 0 \\ & x_{ijk} \leq u_{ijk}(1-\hat{\gamma_{ijk}})(1-\hat{y_w}\delta_{ijkw}) \end{aligned} \tag{5.10}$$

and the following master problem:

$$\begin{aligned}
&\min_{\gamma \in \Gamma} z \\
&\text{s.t. } z \geq x_{tsk}^{l} - \sum_{(i,j) \in A_k} x_{ijk}^{l} \gamma_{ijk} - \sum_{w} \sum_{(i,j) \in A_k} x_{ijk}^{l} y_w \delta_{ijkw} \qquad (5.11)\\
&\gamma_{ijk}, w \in \{0,1\}
\end{aligned}$$

The subproblem is just the summation across pure maximum flow problems, with fixed interdiction and commonality selection. Therefore, the subproblems can be solved as relaxed linear programs (but will still have integer solutions). The solution to the subproblem provides a lower bound to the optimal solution to formulation (5.7). The master problem is simply the attackers problem with "fixed" flows through the network. Therefore, it provides an upper bound to the optimal solution of formulation (5.7). As with the other cases discussed, a Benders' partitioning algorithm would iterate between these upper and lower bounds until they converged to this optimal solution.

5.5 Nodal Extension

The single level layered network interdiction formulation is developed to determine optimal edges (including interdependent edges) to interdict. As discussed in Chapter III, it may also be desired/necessary to model nodal interdiction, instead of / in addition to, edge interdiction.

Recall from Chapter III that a node-only max flow interdiction modifies a traditional edge interdiction formulation by forcing interdiction of a node to disrupt flow in (all/some) associated edges. In addition, the resource constraint must be modified to allow node interdiction as well. If only nodal interdiction is allowed, then the resource constraint is summed over all interdictable nodes. However, if nodes and edges are allowed, then modifications must be made to ensure edge disruptions associated with interdicted nodes are not double counted.

$$
\begin{aligned}
\min \sum_{k \in K} \sum_{(i,j) \in A_k} & u_{ijk}\beta_{ijk} \\
\text{s.t. } & \alpha_{i_k} - \alpha_{j_k} + \beta_{ijk} + \gamma_{ij} + \sum_{w} \delta_{ijkw}\gamma_w \geq 0 \\
& \alpha_{t_k} - \alpha_{s_k} \geq 1 \\
& \sum_{i} r_i\gamma_i + \sum_{w \in W_I} r_w\gamma_w \leq R \\
& \gamma_{ij} = \gamma_i \\
& \alpha_{i_k} \in \{0,1\} \quad \gamma_w \in \{0,1\} \quad \beta_{ijk} \in \{0,1\}
\end{aligned}
\tag{5.12}
$$

Formulation (5.12) is a formulation in which nodes and interdependencies can be targeted. This formulation includes the additional constraint $\gamma_{ij} = \gamma_i$ which forces all outgoing edges from an interdicted node to also be interdicted. The resource constraint is also appropriately modified.

Of course, additional variations are possible. As discussed in Chapter III, this could include formulations where nodes and edges are targetable, with the same or different resources. In addition, modifications could be made to make a subset of the nodes interdependent.

5.6 Notional Examples

5.6.1 Two Identical Layered Networks

Consider Network 1, depicted in Figure 5.2. If no arcs are interdicted, the maximum flow of this network is 26. However, suppose there is an attacker who wishes to minimize this maximum flow. For simplicity, assume each arc can be destroyed with 1 unit of resource, with a resource availability of 2. Using the model in (5.1), the optimal arcs to interdict are found to be $(4,6)$ and $(3,6)$ which reduces the maximum flow to 5. Since Network 2 is identical to Network 1, if they are solved independently, they have the same set of optimal solutions

In other words, if there is no interdependence considered, each network could be solved separately. Of course, this would lead to a combined maximum flow (after interdiction) of 10 units (with 4 units of interdiction resource).

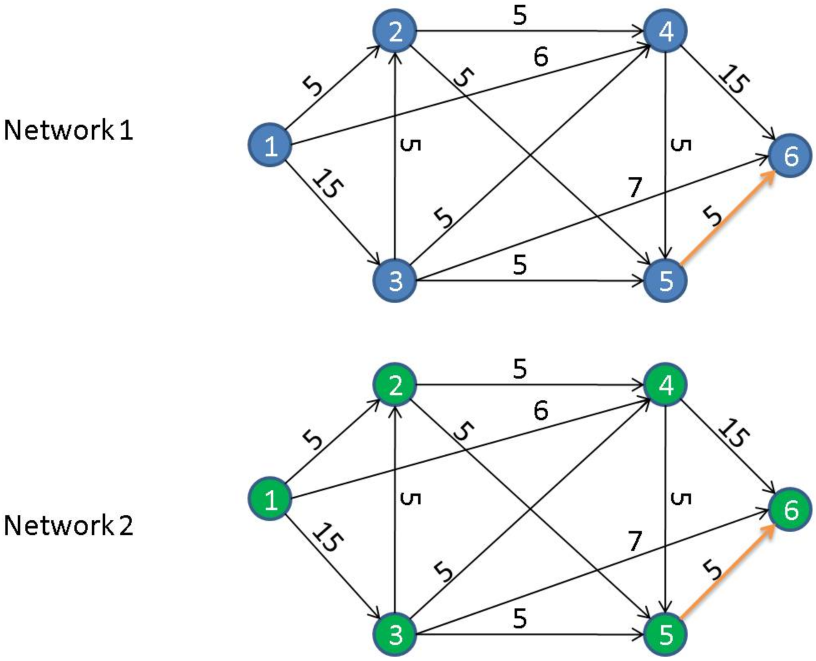


Figure 5.2: Two Identical Layered Networks

Now suppose that the two networks are interdependent. Specifically, assume edges $(5, 6)$ of both networks share a common corridor. That is, although the arcs could be interdicted separately, they could also be interdicted together with a single resource cost of 1 unit. Specifically, this means $\delta_{ijkw} = \delta_{5611} = \delta_{5621} = 1$.

To account for this interdependence, the formulation in (5.7) is used. The optimal arcs to be interdicted are $(4, 6)$ from Network 1; $(3, 6)$ and $(4, 6)$ from Network 2; and interdependent arcs $(5, 6)$. Again, this consumes all 4 units of interdiction resources, but drops the combined maximum flow to 7 (all from Network 1, as no flow is possible in Network 2). Thus, for the same amount of resources, the flow can be further reduced accounting for interdependencies.

5.6.2 Multiple Different Layered Networks

Consider the networks depicted in Figure 5.3. If no arcs are interdicted, the maximum flow of Network 1 is 16, Network 2 is 26, Network 3 is 33, and Network 4 is 24. Therefore, the combined maximum flow across all networks is 99 (assuming commensurate units). From the max-flow min-cut theorem, it can be shown that the cost to cut all of these networks is also 99.

If interdependencies are included, then the cost may be reduced. Assume the colored edges in Figure 5.3 represent common corridors. Therefore, there are three potential interdependencies. The first (represented by the blue edges) are edges $(3, 2)$ in Network 2 and $(1, 2)$ in Network 3. The second (represented by orange edges) are edges $(2, 3)$ in Network 1, $(3, 7)$ in Network 3, $(4, 3)$ in Network 4. The third (represented by green edges) is $(3, 4)$ in Network 2, and $(1, 3)$ in Network 4. Further, assume the costs associated with cutting these interdependencies is 5, 3, and 5; respectively.

Incorporating these interdependencies, formulation (5.5) can be used to determine the minimum cost cut, which is 91. However, suppose there is an attacker who wishes to minimize this maximum flow, but did not have enough resources (91) to completely stop the flow in all networks. This leads to the network interdiction problem. To facilitate a comparison, assume the cost to interdict each arc is the same as the upper capacity of each arc.

As a network interdiction problem with multiple interdependent layers, Formulation (5.7) can be used. If a resource constraint of 91 is used, this formulation confirms that all flow through the networks can be cut. However, when a resource level less than 91 is used, it is possible to determine which arcs should be interdicted to maximally disrupt the network flows. The graph in Figure 5.4 shows the residual flow decreases as an interdictor's resources increase from 0 to 91.

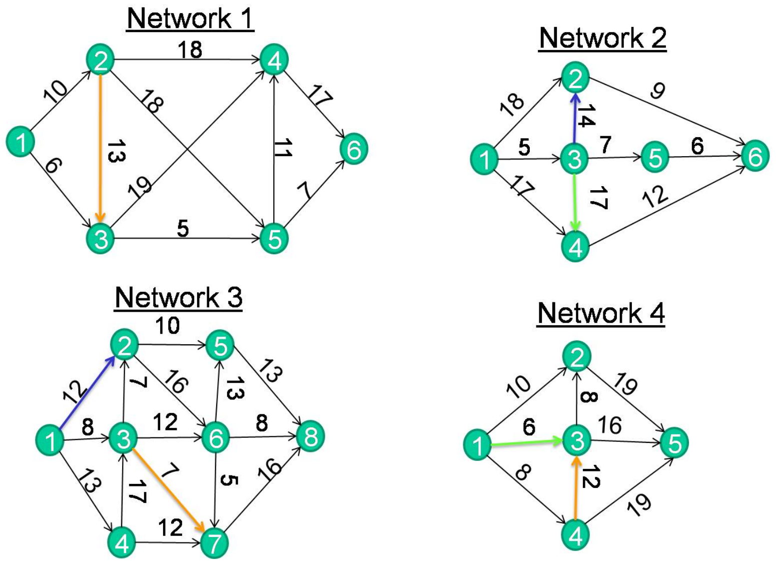


Figure 5.3: Different Layered Networks

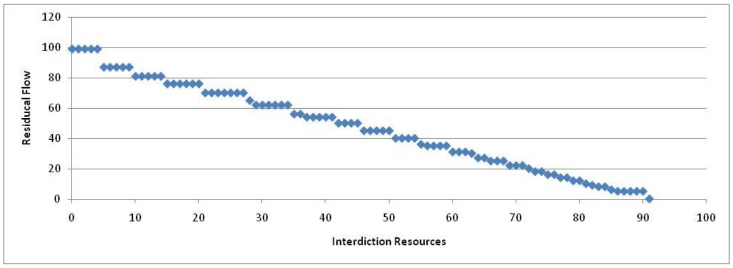


Figure 5.4: Residual Flow versus Interdiction Resources

For example, at a interdiction resource level of 12 units, the optimal solution is to select the first and third interdependent edges (represented by the blue and green edges). This reduces the total flow through all 4 networks from 99 to 81. However, if the interdiction resource is increased by one additional unit, then the optimal solution changes to the first interdependent set and edge $(1, 4)$ in the fourth network. This reduces the maximum flow (across all 4 networks) further to 79 units.

These examples demonstrate how network interdiction against layered networks provides alternatives and more information than traditional cut-sets and are most beneficial when an interdictor's/attacker's resources are limited.

This problem was also solved via the Benders' partitioning method. However, due to the small nature of the example, there was no discernable difference in computation times between the different solution methods. (Both methods result in a GAMS reported 0.0 solution time.)

5.6.3 Nodal Interdiction Example

In the build-up to Operation Iraqi Freedom, CENTAF was concerned calculating the maximum flow from its pre-positioning warehouses to locations within Kuwait. There are several storage locations shown in Figure 5.5 and several methods of transportation available to move material from there storage locations to Kuwait.

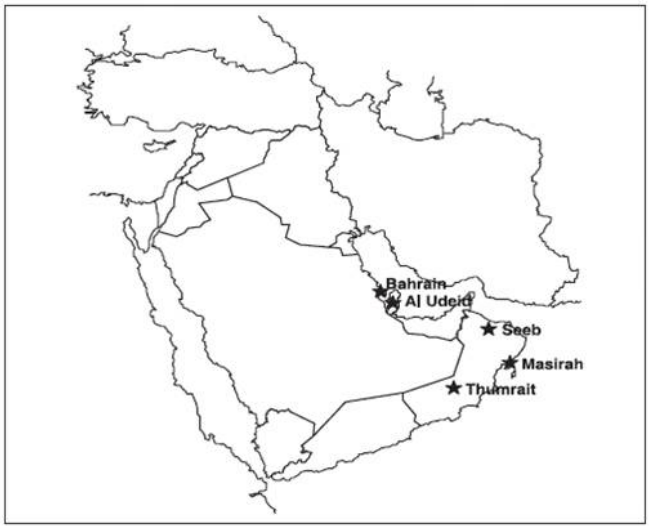


Figure 5.5: War Reserve Material Prepositioned Locations

As shown in Figure 5.6, this scenario could be modeled and solved as a single network. However, for demonstration purposes, these networks were modeled separately (as an airlift network, a sealift network, and a ground/road based network). To determine vulnerabilities associated with these networks, it is assumed that a terrorist organization has the capabilities to stop flow from any one location, or alternatively to stop all sealift through the Strait of Hormuz (i.e. no sealift to Al

Udeid). Disruption of flow in one location could disrupt flow in all three transportation networks as the same nodes appear in multiple networks.

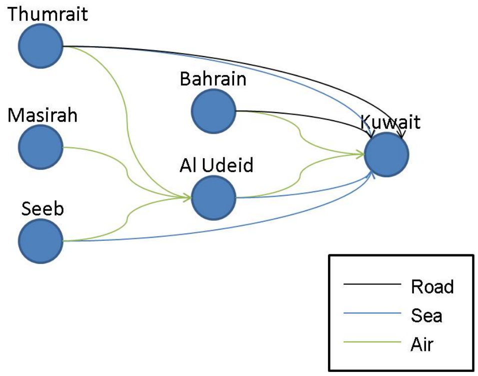


Figure 5.6: Network Representation

The three network flow models were formulated as discusses in Section 5.5. Of course, the nodes appearing in multiple networks create an independency, and an additional effect was modeled which reduced sealift to Al Udeid to zero if the Straight of Hormuz was disrupted. The optimal solution to disrupt the flow of material from storage locations to Kuwait is stop flow from Al Udeid. The second best solution is to stop sealift to Al Udeid; however, disruption of flow from Al Udeid also stops flow from local warehouses in Al Udeid (which would not be affected in a Straights of Hormuz disruption).

5.7 Computational Experiments

Random networks were generated using NETGEN. [76] In order to ensure a feasible network is created, NETGEN first constructs a skeleton network which allows the specified level of flow. NETGEN then adds additional random arcs until the total number of arcs is equal to the number requested.

NETGEN has several input parameters: random number seed, number of nodes, number of arcs, number of sources, number of sinks, amount of flow from source to sinks, range on arc costs, and range on arc capacities. Klingman *et al.*

included 40 example inputs which could be used to generate random networks. [76, p. 818] The Center for Discrete Mathematics and Theoretical Computer Science (DIMACS) website (ftp://dimacs.rutgers.edu/pub/netflow/generators/) make the source code for NETGEN (written in Fortran and C), and all 40 sample inputs provided by Klingman *et al.* available on their website. In addition, eight representative examples from this sample were selected and provided for algorithm testing.

Seven of these eight inputs were used in this research to develop random networks for computational testing (the eighth was a pure assignment problem). However, some modifications were necessary. NETGEN was originally designed to create random minimum cost network flow problems. Since this chapter is concerned with maximum flow problems, the flow demand constraints were dropped. In addition, the costs assigned to each arc are reinterpreted to mean cost to interdict an arc (instead of cost per unit of flow on the arc).

To create layers (interconnections) between the networks, arcs were selected at random to be interdependent across the networks under consideration. This random selection was done via Excel VBA, with roughly 1% of the arcs randomly selected. These layered networks were then analyzed in GAMS with the BARON/CPLEX solvers; and on a Dell Precision M6300 with 2.50 gigahertz Intel Core2 Duo processor and 4094 megabytes of RAM. The results are listed in Table 5.2.

For example, in the first example, the two networks (which were created from inputs 20 and 27 from Klingman's original list) were created by NETGEN and converted to a maximum flow problem with 1% of the arcs between the two networks interdependent. These networks were then analyzed using both the MIP formulation and the Benders' partitioning formulation with solution times provided. This process was repeated for various combinations of the seven networks presented in the table.

As the table demonstrates, both solution approaches can solve moderately sized networks (1000's of nodes and 10,000's of edges) relatively quickly. Therefore, there

Table 5.2: Computational Results of Layered Maximum Flow Interdiction Formulations

Test	Network	Nodes	Arcs	MIP Time	Benders' Iterations & Total Clock # Iterations / (H:MM:SS)
1	Network 1 (20)	400	1416	0:00:01.702	11 / 0:00:10.750
	Network 2 (27)	400	2676		
2	Network 1 (28)	1000	2900	0:00:02.635	6 / 0:00:10.252
	Network 2 (32)	1500	4342		
3	Network 1 (36)	8000	15000	0:01:32.981	2 / 0:00:48.128
	Network 2 (38)	3000	35000		
4	Network 1 (28)	1000	2900	0:00:33.518	15 / 0:02:46.546
	Network 2 (38)	3000	15000		
	Network 3 (40)	3000	23000		

are indications that these modelling and solution approach can be applied to social networks, infrastructure networks, and other potentially large networks.

5.8 Application

In order to demonstrate the potential of the model, it has been applied to a realistic communications network. One such notional communications network is depicted in Figure 5.7.

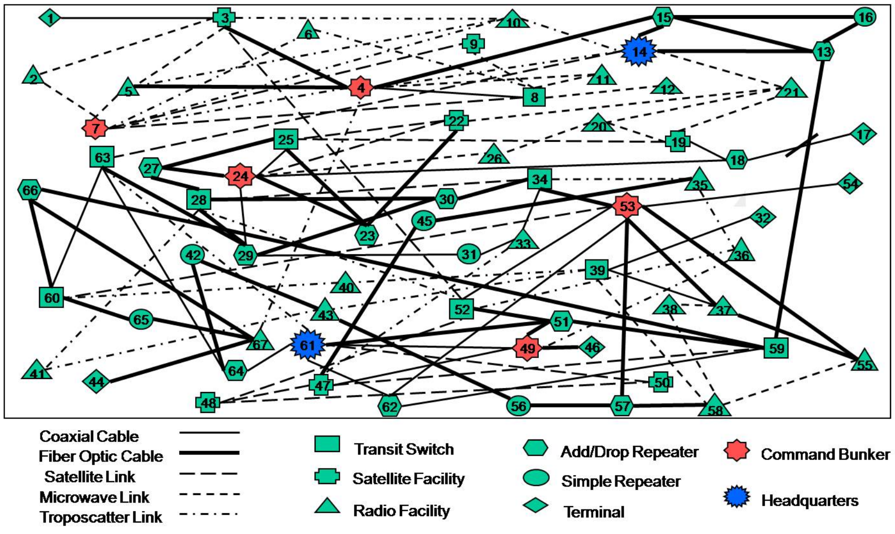


Figure 5.7: Notional Communications Network [95, p. 55]

To use this notional network, it is assumed that one goal of this network is to maximize the amount of information flow from the ”headquarters” to the “Command Bunkers.” Specifically, for this example, the goal was to maximize flow from the headquarters at 61 to the command bunker at 7. In addition, because of their unique capabilities, policies, ownership, and so forth; the ground based portions and satellite based portions are modeled separately.

Even though the two networks are modeled separately, there is, of course, some overlap. Specifically, ground based satellite facilities contribute to flow through both networks. Therefore, disruption of a ground based facility would disrupt flow through the satellite network and associated ground based network systems.

To make this scenario as realistic as possible, it is assumed that satellites and intra-satellite communication can not be disrupted (not necessarily for technical, but for policy/legal reasons). Therefore, disruption of the satellite network requires disruption of the ground based transmitters/recievers. The resulting network model, with two layered maximum flow networks, to determine optimal disruptions to the layered system are modeled accordingly.

The maximum flow from the headquarters to the command bunker with no interdictions/disruptions is 61.92. Given this model and scenario, as long as there are sufficient resources to disrupt the ground based satellite facilities, the model always does so. There are technical/geographic difficulties associated with jamming/disrupting satellite down links, but those difficulties were not included here. If resources are such that this is the only option selected, then the resulting maximum flow drops from 61.92 to 54.816. Of course, as interdiction resources increase, the resulting maximum flow decreases further.

5.9 Conclusion

This chapter extended network interdiction techniques to apply to layered networks. It was shown how traditional network interdiction models can be developed to account for the interdependent nature of layered networks. This allows the maximization of the disruption across all the individual networks and shared elements in the overall cut set. This formulation is further developed and combined with nodal interdiction.

Both a MIP formulation and a Benders' formulation were developed and described. These formulations were implemented in GAMS and computational experiments indicate the potential for use against very large networks, such as infrastructure networks. This would allow a decision maker to consider the effects of an (interdiction) attack across multiple layers of networks, vice the single network effects traditionally considered.

VI. Synthesis of Robust Networks

As discussed in the introduction, the formulation development discussed in this section is the development and synthesis of robust networks. As shown in Figure 6.1, this formulation combines two blocks: multiobjective network programming and multiobjective multilayer programming. First, a multiobjective optimization technique is combined with network optimization to form a multiobjective network design. This extended formulation is then combined with multilayer network programming to form multiobjective multilayer network design.

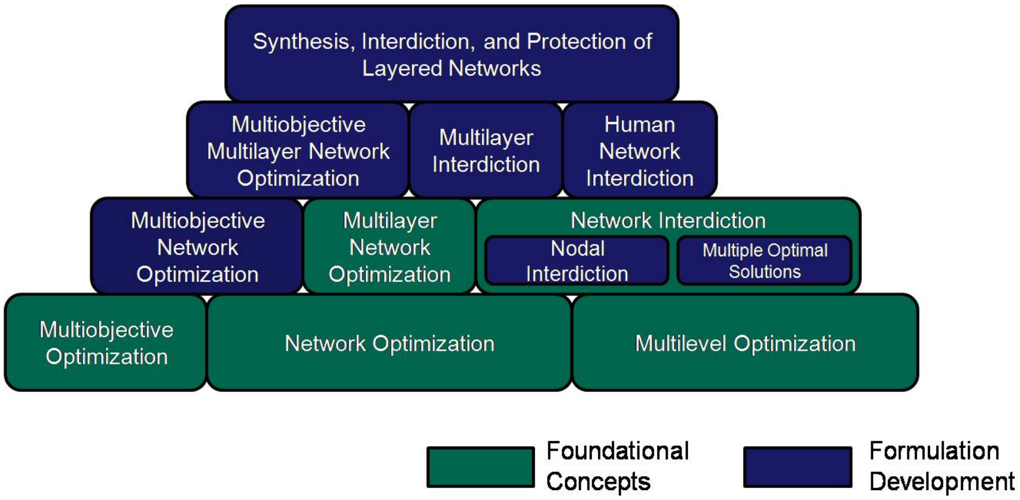

Synthesis, Interdiction, and Protection of Layered Networks
Multiobjective Multilayer Network Optimization
Multilayer Interdiction
Human Network Interdiction
Multiobjective Network Optimization
Multilayer Network Optimization
Network Interdiction
Nodal Interdiction
Multiple Optimal Solutions
Multiobjective Optimization
Network Optimization
Multilevel Optimization
Foundational Concepts
Formulation Development

Figure 6.1: Robust Network Design

In Section 2.2.9.1, the vulnerability and survivability of networks was discussed. As these subjects center on the concept of connectivity, an importation distinction must be made. “Performance of a network, viewed in terms of either vulnerability or survivability, ultimately centers on connectivity and whether flow can move between origins and destinations.” [89] “Interdiction does not necessarily depend upon level of connectivity, but rather flow between origins and destination.” [89, p. 112] Of course, the two are closely related, as connectivity can be determined by sending one unit of flow from a source-sink pair.

6.1 Maximize Robustness in Network Design

This section begins with the network design formulation developed and discussed in Section 2.2.9:

$$\begin{aligned}
\min \quad & \sum_{ij\in E} c_{ij} z_{ij} & & \\
\text{s.t.} \quad & \sum_{i\in W}\sum_{j\in V\setminus W} z_{ij} \geq r_{st} & & \forall (s,t)\in V, s\neq t, \forall W\subseteq V, s\in W, t\notin W \\
& \sum_{i\in W}\sum_{j\in V\setminus (Z\cup W)} z_{ij} \geq d_{st} & & {\scriptstyle \forall (s,t)\in V, s\neq t, \forall z\subseteq V\setminus\{s,t\}, |Z|=k_{st}, \forall W\subseteq V\setminus Z, s\in W, t\notin W} \\
& z_{ij}\in\{0,1\} & & \forall ij\in E
\end{aligned} \tag{6.1}$$

Initially, this formulation is simplified to only consider edge disjoint networks. It is noted that this is not a restrictive constraint, but it simplifies the notation and bookkeeping. In this formulation, the objective was to minimize cost subject to a minimum level of robustness. Robustness is defined in terms of survivability requirements. Specifically, that there be at least r_{st} edge disjoint paths so that the removal of at least r_{st} edges is required to disconnect the graph. Often, however, budgets are limited, and an organization seeks to maximize its robustness subject to a budget constraint. Therefore, the objective function is modified to find the maximum level of robustness subject to a budget constraint, β.

With this modification, the following formulation results:

$$\begin{aligned}
& \text{Max } \alpha \\
& \quad \text{s.t.} \sum_{i\in W}\sum_{j\in V\setminus W} z_{ij} \geq \alpha \quad \forall S\subseteq N, \emptyset\neq S\neq N \\
& \quad\quad \sum_{\forall (i,j)\in E} c_{ij} z_{ij} \leq \beta \\
& \quad\quad z_{ij} = \{0,1\} \quad \forall (i,j)\in E
\end{aligned} \tag{6.2}$$

where the new variable α is the redundancy value to be maximized and β is the budget for the (construction of) edges. With this, there now exists a formulation to maximize robustness of a single layer network subject to budget constraints.

This formulation is now modified to consider flows. As mentioned in the introduction, formulations for survivable design usually consider connectivity, not the amount of flow between nodes. However, in the design of some survivable networks, such as infrastructure networks, it may be important to give at least minimal consideration to flows to ensure demands are met. The following formulation includes the flow balance constraints (including the demand node), and only allows flows across edges that are built.

$$
\begin{aligned}
&\text{Max } \alpha \\
&\text{s.t. } \sum_{i \in W} \sum_{j \in V \setminus W} z_{ij} \geq \alpha \quad \forall S \subseteq N, \emptyset \neq S \neq N \\
&\qquad \sum_{\forall (i,j) \in E} c_{ij} z_{ij} \leq \beta \\
&\qquad \sum x_{ik} - \sum x_{kj} = \begin{cases} -d, & \text{for } i = s; \\ 0, & 0 \; \forall \;\; i \in N \setminus \{s,t\}; \\ d, & \text{for } i = t. \end{cases} \\
&\qquad x_{ij} \leq u_{ij} \times z_{ij} \\
&\qquad z_{ij} = \{0,1\} \quad \forall (i,j) \in E
\end{aligned}
\tag{6.3}
$$

6.2 Maximize Robustness across Layered Networks

Given the interdependent nature of networks, this formulation is now combined with the layered networks formulation discussed in Section 2.4. When moving from a single layered network model in 6.2 to a multilayer model, it is recognized that two different types of edges can be added: an edge that remains within a single layer and edges that effect multiple layers (interdependent). In addition, the A_k matrices for the individual layers take on a slightly different meaning. Instead of describing the network edges of a fixed network, they now represent fully connected layers. Any of these links actually used will incur a cost which is subject to the budget limitation. In addition, the cost values now assume a value of the cost of constructing an edge.

$$\begin{aligned}
&\text{Max } \alpha \\
&\text{s.t. } \sum_{i \in W} \sum_{j \in V \backslash W} z_{ij} \geq \alpha \quad \forall S \subseteq N, \emptyset \neq S \neq N \\
&\qquad \sum_{\forall (i,j) \in E} c_{lij} z_{lij} - \sum_{w} (c_{lij} y_w \delta_{(ijl),w}) + \mathbb{C}_w y_w \leq \beta \\
&\qquad \sum x_{lik} - \sum x_{lkj} = \begin{cases} -d, & \text{for } i_l = s; \\ 0, & 0 \; \forall \;\; i \in N \backslash \{s,t\}; \\ d, & \text{for } i_l = t. \end{cases} \qquad (6.4) \\
&\qquad x_{lij} \leq u_{lij} z_{lij} + \sum_{w \in W} y_w d_w \\
&\qquad z_{lij} \geq \sum_{w \in W} y_w \delta_{(ijl),w} \\
&\qquad z_{lij}, y_w = \{0,1\} \quad \forall (i,j) \in E
\end{aligned}$$

where d_w is the change in capacity associated with included edges of effect option w, and all other variables are as previously defined.

The objective function and the first constraint together maximize the robustness of the multilayered network by maximizing the number of edge-disjoint paths (as discussed in Section 2.2.9). The second constraint ensures that the cost of edges (both inter- and intra-layer) does not exceed the budget. Note: the selection of interdependent edges automatically creates the associated individual network edges. However, as this cost is incurred as part of the interdependent arc selection, the cost of the automatically created edges is subtracted in this constraint. Finally, the last constraint ensures that if an interdependent set of edges is selected for inclusion, then the corresponding individual network components are also selected (but as noted above, with no additional increase in cost).

The model as formulated finds the most robust layered network possible given the budget limitation while ensuring that the demands are met. This budget parameter, β, can be varied to determine the effect of changing budgets on the robustness of the system. As discussed in section 2.3.1, by varying this parameter, all efficient solutions (Pareto optimal) can be found. However, eventually, if the budget parameter is dropped too low, the problem becomes infeasible, as demand can not be met.

In addition, as discussed in Section 2.3.1, an alternative formulation may speed computation: the elastic constraint formulation. In this formulation, instead of varying the budget directly, deviations from (above and below) the specified budget are allowed (although positive deviations are penalized). Although likely to be computationally faster, this formulation is still $\mathcal{NP}$-complete.

This formulation can also be applied to existing networks to determine optimal additions to the network to maximize robustness. All components in the existing network are assigned a cost of zero in formulation (6.4). The optimal solution to this model includes new components to be constructed and includes any used portions of the already existing network.

6.3 Notional Examples

6.3.1 Maximum k-Connectedness with Feasible Flow

Consider the nodes pictured in Figure 6.2. Suppose one wishes to build a survivable network with these nodes. A directed arc can be built between any two nodes at a cost of one unit, and any built edge has an upper capacity of 5 units. In addition, Node 1 can supply 25 units, and Node 6 demands 25 units. Finally, assume there are 10 units of resource available to build edges.

With this information, formulation (6.3) can be used. This results in a 2-edge connected network ($\alpha = 2$) as depicted in Figure 6.3. Note, all edges (except edge $(2, 3)$) carry 5 units of flow.

As mentioned previously, the amount of resources available can be varied to determine the set of efficient solutions. To show this, the amount of resources is

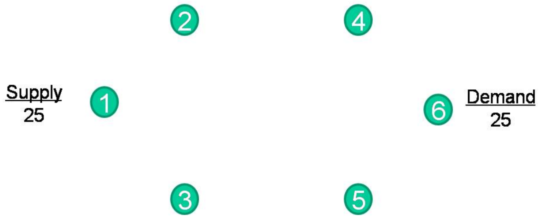


Figure 6.2: Six Nodes

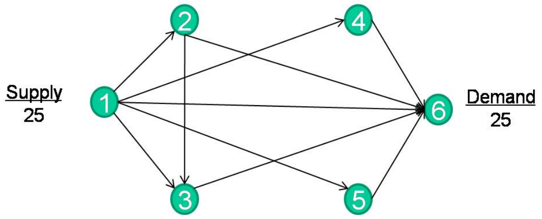


Figure 6.3: Survivable Network with Feasible Flow

varied from 0 to 30 (which would result in a maximally connected network since the cost of all edges is 1).

As shown in Figure 6.4, the problem is infeasible until $\beta = 9$. As resources increase, the potential maximum robustness rises until the maximum possible level of robustness is reached for this network (which is 10).

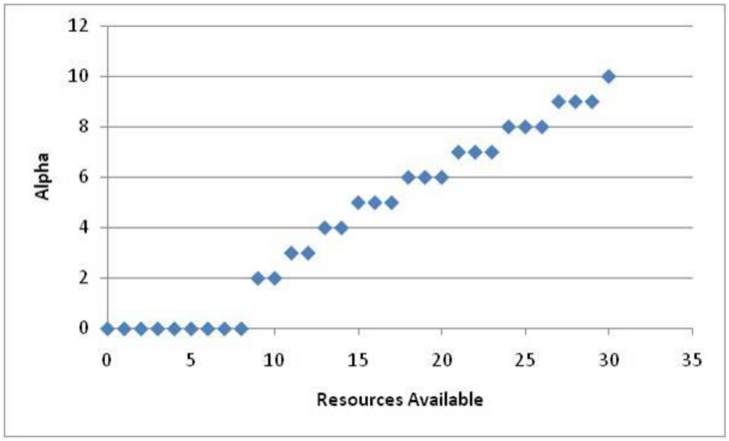


Figure 6.4: Efficient Solutions for Six Node Network

6.3.2 Layered Survivable Network Synthesis

This example illustrates how the methodology has been extended to consider multiple layers of networks when designing survivable networks with feasible flows. To do this, consider the nodes in Figure 6.5. As in the previous example, the cost to create an edge from one node to any other node in the network is one. However, to facilitate the demonstration, only those edges in grey in the figure allow flow. Other edges may be incorporated to increase connectivity (and maximize robustness), but may not have flow in the optimal solution. For the edges in grey, if the arc is selected for inclusion, the first number is the cost (per unit of flow) of flow across the arc, and the second number is capacity of the edge. Note: the formulation does not actually consider the cost of flows, but this information is still included as the formulation could be easily modified to consider it. Finally, edges $(4, 5)$ in both networks are interdependent. Either/both edges could be chosen separately (through the individual networks) or combined with a cost of 1 unit. In addition, this interdependency is modeled to include a change in capacity of both arcs to a maximum of 10 units (from an initial 1 unit).

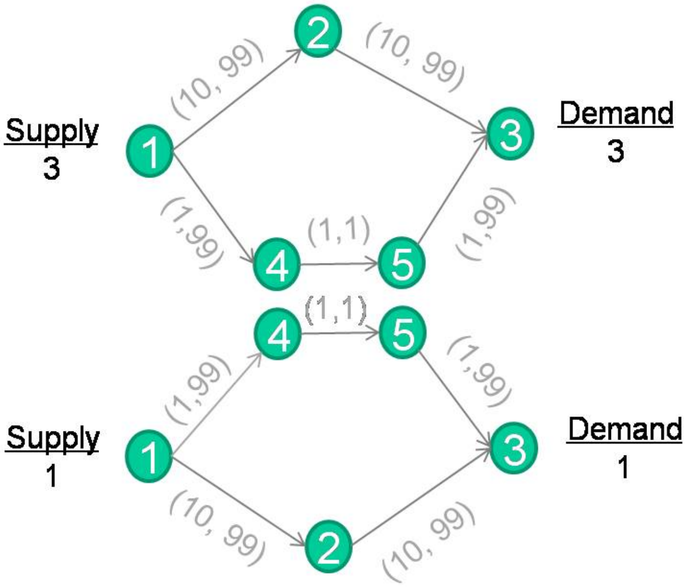


Figure 6.5: Layered Synthesis Example

With this information, formulation (6.4) can be used to determine which edges should be included to maximize robustness across both networks with the given constraints. This results in the networks depicted in Figure 6.6.

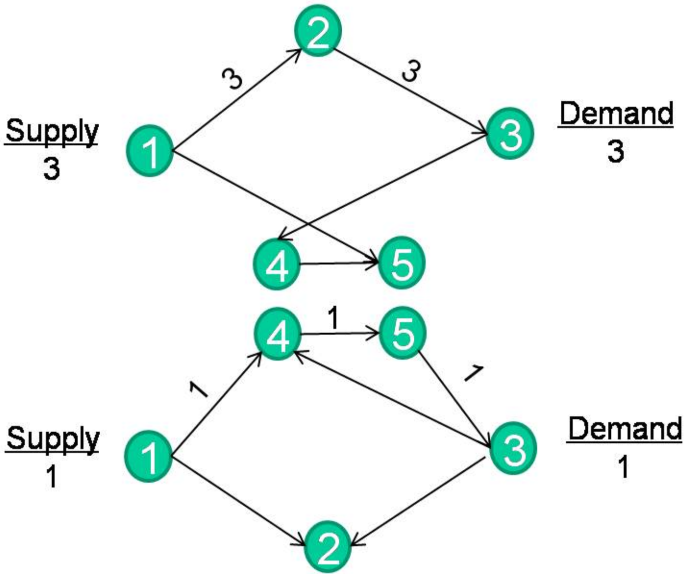


Figure 6.6: Layered Synthesis Example Results

Both networks have $\alpha = 2$. Although the interdependent effect was chosen (which brought in both edges $(4, 5)$) which modified the upper bound of both arcs, the algorithm did not elect to send any/all flow through these arcs. This is because the algorithm only considers achieving feasible flow, then maximizing robustness. If flow costs were also included in the objective function, less expensive flow could have been achieved by using these (interdependent) edges.

As done with the previous example, the amount of resources available can be varied to determine the set of efficient solutions. To show this, the amount of resources is varied from 0 to 39 (which would result in a maximally connected network since the cost of all edges is 1). (Each node can connect to at most 4 other nodes. As edges are directed, one edge can be connected in each direction to each node. This leads to a maximum of 8 edges per node. Since there are 5 nodes, this leads to a maximum of 40 edges (per network) which could be selected. However, since 2 edges can be selected with a single cost of 1, this maximum is reached at 39.)

As shown in Figure 6.7, the problem is infeasible until $\beta = 7$. As resources increase, the potential maximum robustness rises until the maximum possible level of robustness is reached for this network (which is 8).

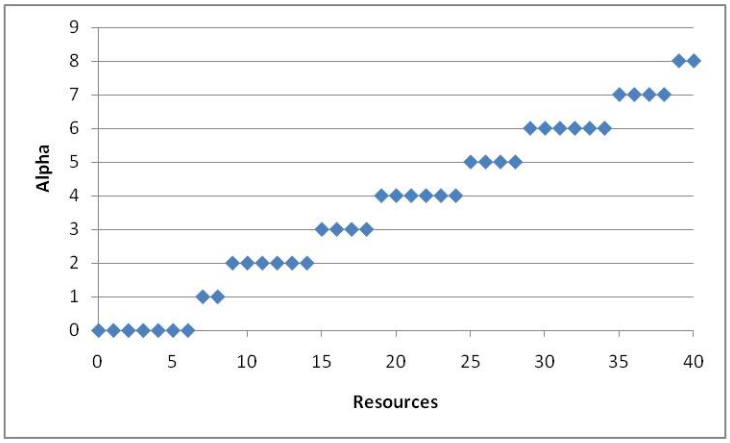


Figure 6.7: Efficient Solutions for Layered Five Node Network

6.4 Computational Issues

Unfortunately, the “connectivity” constraints make direct application of these formulations to large networks impractical. To ensure connectivity, these constraints must be formed for every possible combination of subsets (known as a powerset) of nodes. Therefore, if there are n nodes, then there are 2^n possible separations of these nodes into 2 groups. In the formulations discussed in this chapter, the 2 cases where one of the subsets is empty can be ignored because there can be no connectivity to an empty subset. Therefore, these formulations require $2^n - 2$ connectivity constraints. In addition, numerous other constraints are required, such as capacity constraints for every possible edge between these subsets.

However, even if only the connectivity constraints are considered, then a network of 30 nodes would require over one billion constraints ($2^{30} = 1,073,741,824$).

Therefore, the formulations as stated should only be used for designing small networks (or aggregated large ones). Otherwise, as discussed previously, small changes to an existing network could be modeled by assigning a zero cost to existing network components. The model would then find the set of additional edges to add to the network to maximally increase robustness/connectivity.

Fortunately, the formulations can also be useful for large network designs under some circumstances. For example, instead of considering all possible subsets of nodes, a decision maker could pre-define the subsets between nodes that are of concern in the given network. In this case, the connectivity constraints would only be necessary for these predefined subsets. Alternatively, these formulations could be useful when only small changes to an existing network are being contemplated. The models as currently formulated, examine every possible edge between nodes. However, if only a small number are feasible candidates for consideration (for whatever reason), then only these connections would need to be considered.

Finally, other possible techniques could be considered. There are many heuristic techniques that have been developed to deal with formulations that have an explosion in the number of constraints (such as this and the traveling salesman problem). In addition, high performance computing has been successfully used to solve (to optimality) large instances of problems of this type (such as the traveling salesman problem). Therefore, with sufficient computational resources, even large instances of this problem can be solved.

6.5 Summary

This chapter developed extended formulations for the synthesis of robust networks. This development combined many aspects of operations research such as multiobjective programming, network design, and multilayer programming. These new formulations allow the a decision maker to maximize robustness in network design across single and layered networks. These models can be directly applied to develop small networks or to specialized large networks. Otherwise, heuristics would likely provide robust solutions in a reasonable amount of time.

VII. Multiple Optimal Solutions

7.1 Overview

This chapter considers aspects of the impact of multiple optimal solutions on network interdiction and related models. First, multiple optimal solutions are discussed in the context of "pure" interdiction problems (where the objective functions are identical but diametrically opposed). This is followed by a discussion of the benefits of multiple optimal solutions in an interdiction problem. Specifically, this allows decision makers to consider non-quantifiable objectives in selecting from among the optimal solutions. Finally, a discussion of the problems associated with multilevel models is discussed, along with methods to explore these issues and determine a range of options and expected reactions.

7.2 Network Interdiction

Traditionally, as discussed in Section 2.3.3, multiple optimal solutions are not normally a consideration in "pure" network interdiction problems because the objective functions in network interdiction models are diametrically opposed. Therefore, the rational reaction set consists of one possible solution *for the follower.* [70, p. 113] In other words, there are not multiple solutions for the follower which would change the leader's/attacker's objective function, if one assumes all functions have been captured in the objectives.

7.2.1 Follower Solutions

However, even in pure interdiction, there are circumstances where multiple optimal solutions of the follower impact the decision or effectiveness of the leader/attacker. Consider the simple social network depicted in Figure 4.4 and reproduced (with minor changes) in Figure 7.1.

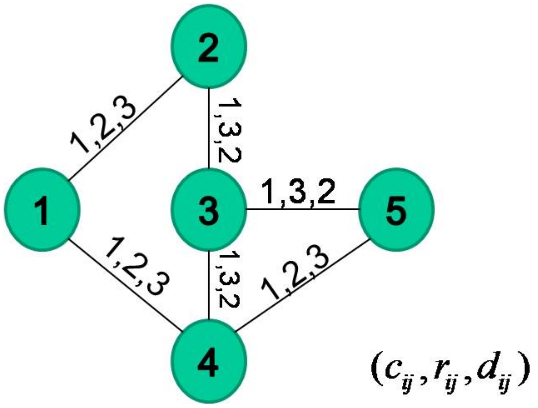


Figure 7.1: Notional Network

From Node 1 to Node 3, there are two (edge and node disjoint) independent shortest-paths: 1-2-3 and 1-4-3. In addition, unless the interdictor has resources of at least 4 units, there is insufficient resources to disrupt both paths. This is a case where the multiple optimal solutions of the follower (the shortest path model) may prevent the attacker (shortest path interdiction) from obtaining any impact until a specific amount of interdiction resources is obtained.

With $r = 3$, there are six optimal solutions to the shortest-path interdiction from Node 1 to Node 3. Specifically, there is an optimal solution to interdict each of the six arcs in the network. While any simple program will generate one of these solutions, each of which is indeed (mathematically) optimal for the stated objective function, the solution(s) miss the insight that none of these solutions actually impact the shortest-path from 1 to 3. Therefore, while the multiple optimal solutions of the follower do not change the value of the leader's objective function, it is important to determine their existence and impact on the leader's solutions.

7.2.2 Attacker Solutions

In addition, the "attacker" in network interdiction problems often faces multiple optimal solutions. Unless these multiple optimal solutions are specifically requested (i.e. from math programming software such as GAMS), their existence and number will likely remain unknown. For example, in the previous chapters, an optimal solution to example/notional problems was presented. While the solutions provided are optimal, the existence and consideration of alternative optimal solutions that may exist was ignored.

Therefore, to explore these problems for multiple optimal solutions, the procedure outlined by Danna *et al.* for mixed integer programs was used. [48] Specifically, the "one tree" algorithm developed and implemented in GAMS was used to find all multiple optimal solutions to MIP's developed in the previous chapters. The "one tree" algorithm is a modification of the standard branch-and-bound algorithm. This algorithm proceeds in two phases. In the first phase, an optimal solution is found with a traditional branch-and-bound algorithm, but all nodes are kept for a second phase where the tree is reused to explore for multiple optimal solutions. These solutions are found through modifications to the way the branch-and-bound algorithm stores integer solutions, fathoms nodes, branches, and (dual) tightens. [48, p. 283]

A sample of each respective model was modified with this procedure to determine the existence/number of multiple optimal solutions. Specifically, the models in each chapter's application to the notional communications network(s) was modified. For example, in the node-only interdiction of the communications network (discussed in Section 3.7) with interdiction resources of 75 units, there are six multiple optimal solutions. In the all-pairs shortest path network interdiction (used for calculating closeness centrality in Section 4.6.2), this same network has three multiple optimal solutions. Finally, the same network was modeled as layered networks in Section 5.8, where 20 multiple optimal solutions were found.

While mathematically all multiple optimal solutions have the same value for there objective function, there may be subjective reasons why some of these solutions may be preferable over others. Given a set of multiple optimal solutions, a decision maker can consider aspects not explicitly modeled to determine the best course of action.

> Because of the differences between the mathematical model and the real business problem, it is interesting to generate multiple optimal or near-optimal solutions for the mathematical model so that the decision maker can examine them, and in the end, choose the best solution overall, i.e., the one that also performs best for the criteria that could not be expressed in the MIP model. [48, p. 281]

For example, in traditional network interdiction modeling, the model's solution determines the optimal edges to disrupt to maximally disrupt the network. If there are multiple optimal solutions, some of these solutions may require fewer munitions, put pilots at less risk, relate to information that is not available to the analyst, or risk fewer civilian causalities; none of which are directly discussed in the network interdiction model. These concerns may be considered during an equity review. Fortunately, if the decision makers are given a set of solutions to choose from during this review, then the decision makers can consider a variety of factors and choose the "best" optimal solution.

7.2.2.1 Diversion Example

For a more specific example, consider network diversion. The traditional network diversion problem is to

> identify a minimum cost set of directed edges to cut, so that any directed path from a specified source node s to a specified sink node t must include at least one directed edge from a specified subset of edges. [47, p. 35]

Unlike network interdiction, the goal of network diversion is not necessarily to reduce the amount of flow from the source to sink, but to redirect it through a specified set of edges through the removal of a minimum cost set of edges. However, a decision maker may wish to combine these ideas (network interdiction and diversion).

For example, suppose the primary goal of an attacker is the traditional interdiction objective to minimize the maximum flows through the network subject to resource constraints. In addition, the attacker would like to maximize the amount of (post-attack) remaining/residual flow that flow through a predefined "diversion" set of edges. For example, suppose an attacker wishes to disrupt a communications network. However, the attacker does not have sufficient resources to completely cut the network. Therefore, the attacker would like the remaining communications after an attack to be forced through channels/links that are less secure.

One approach to solve this problem would be first solve the network interdiction problem for all optimal solutions. The second step would be to search these multiple optimal solutions for those which have any (the most) flow across the diversion set of edges. To illustrate this potential, consider the network in Figure 7.2. This example (slightly modified from [47, p. 36]) has a flow capacity of one unit across all edges and requires one unit of interdiction resource to interdict each edge. The maximum flow across this network is 3 units.

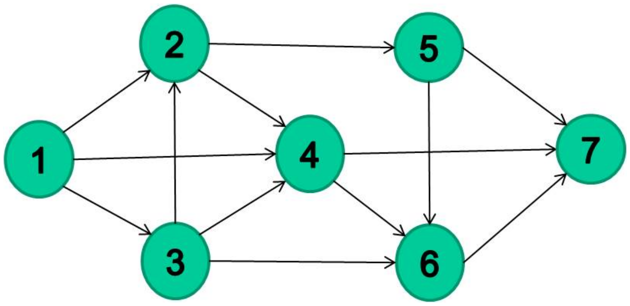


Figure 7.2: MOS Example

Suppose an interdictor wishes to disrupt this network and has two units of interdiction resources. Using maximum flow network interdiction, this network can be reduced to one unit of flow. When solved using formulation (2.66) in GAMS, the optimal solution of interdicting edges $(1, 2)$ and $(1, 3)$ is returned. However, using Danna's one-tree algorithm, seven additional solutions are found: $(1, 2)$ & $(1, 4)$; $(1, 3)$ & $(1, 4)$; $(2, 5)$ & $(4, 7)$; $(2, 5)$ & $(6, 7)$; $(4, 7)$ & $(5, 7)$; $(4, 7)$ & $(6, 7)$; and $(5, 7)$ & $(6, 7)$. All these solutions have an objective function value of one. However, suppose a secondary objective was to maximize the remaining flow in the network through edges $(3, 6)$ and/or $(4, 6)$. With this secondary objective, then the best optimal solution is to target edges $(2, 5)$ and $(4, 7)$. With this solution, all remaining flow must flow through one of the diversion arcs. In [47, p. 36], Curet was only interested in the diversion set (and found the same solution); however, this methodology ensured remaining flow was minimal and diverted. The variety of solutions allows flexibility for the decision maker which would not have been found without specifically looking for them (i.e. via the "one tree" algorithm).

7.2.2.2 Resources Example

Another discriminator among multiple optimal solutions is the amount of interdiction resources used. Consider the example in Figure 7.3, and assume an interdictor has 11 units of interdiction resources available to attack this network.

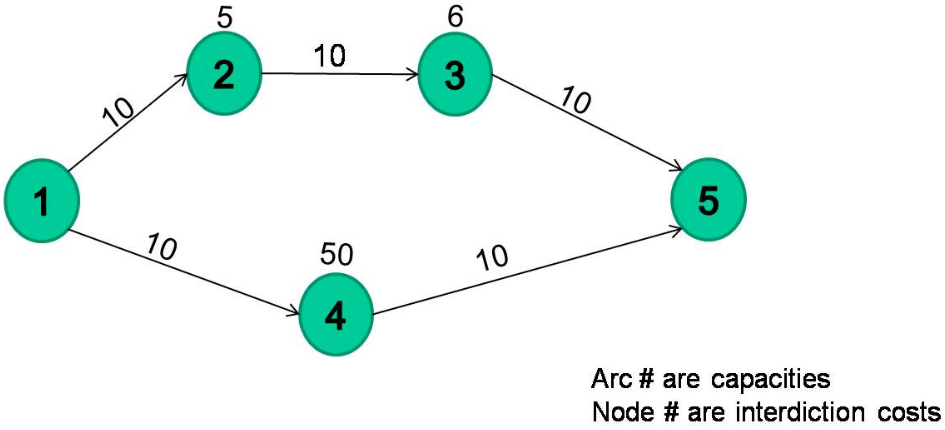


Figure 7.3: Notional Network for Resources Example

When this example is modeled and solved via GAMS with the CPLEX solver, the optimal solution returned is to disrupt node 3 (using 6 units of resource); which reduces potential maximum flow to 10. However, on further examination, there are two additional optimal solutions: disrupt node 2, and disrupt both nodes 2 and 3. All three of these solutions reduce maximum flow to 10 units. However, all three of them use different levels of resources to achieve this. The least resource intensive solution is to disrupt node 2 (which uses 5 units of interdiction resource), while the most expensive (using all 11 units of interdiction resources) is to attack both nodes 2 and 3.

While all three solutions are feasible (i.e. do not violate the cost constraint), it may be an ineffective use of resources to use more than necessary to achieve the optimal solution. It is important to reiterate that the first solution found by GAMS uses more resources than another optimal solution. Therefore, if costs/resources are indeed a concern, then either this cost should be included in the objective function, or an examination is needed among all multiple optimal solutions.

For example, consider the notional communications network in Figure 7.4.

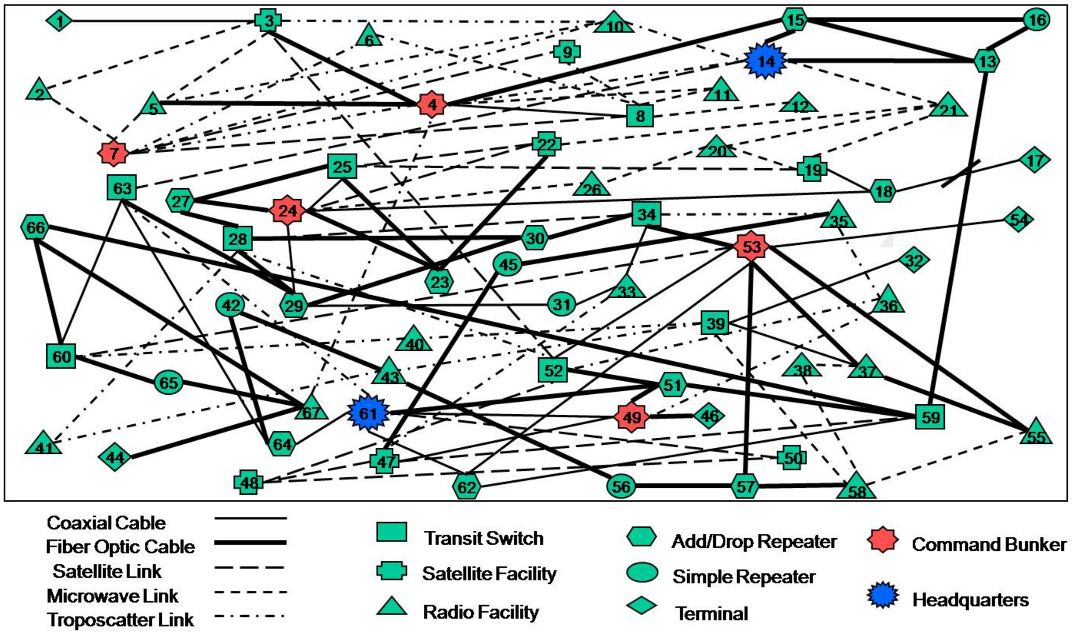


Figure 7.4: Notional Communications Network

As discussed previously, with interdiction resources of 75 units, there are six multiple optimal solutions to the node only interdiction of this network (as presented in Section 3.7). An examination of these solutions indicates that the option that uses the least resources (60) is interdiction of nodes 13, 15, 21, 51, 62, and 64. The other solutions either disrupt nodes unnecessarily, or swap out one node for a more expensive one, with the same results. Therefore, all things being equal, the cheapest solution should be chosen.

7.3 Multilevel Programs

The previous section focused on the role of multiple optimal solutions in network interdiction types of problems. This issue of multiple optimal solutions becomes more complicated when differing objective functions are involved.

Bullock *et al.* discussed the necessity of using different objective functions when dealing with terrorist organizations, and they created a methodology (using Value Focused Thinking (VFT)) to develop likely strategies and courses of action for all players. [30] This methodology goes well beyond the contribution of each network asset to be potentially attacked, and considers the values, fundamental objectives, and means objectives. [30, p. 1866]

Unfortunately, the introduction of multiple objective functions leads to potentially unsolvable problems. This was discussed in Section 2.3.3 along with potential mitigation strategies. This breakdown is a result of multiple optimal solutions of the follower's problem. Since the objective functions of the leader and the follower are different, the value of the leader's objective function could potentially change depending on which of the solutions the follower chooses. As the follower is indifferent to these solutions and the leader can not dictate which solution the follower should chose, there is no general way to converge to a solution.

Mitigation strategies either make additional assumptions such as the follower will always chose the solution that is best for the leader (the optimistic strategy) or will always chose the solution that is worst for the leader (the pessimistic strategy). However, neither of these strategies are appropriate when dealing with terrorist adversaries.

> But both the optimistic and pessimistic solutions can in general not be assumed to be good approximations of the realized solutions in practice. [49, p. 217]

In addition, as Cao and Leung point out:

> But there is little justification to assume that the follower will only behave in the two extreme ways and that the leader should only choose between optimistic and pessimistic approaches. [31, p. 135]

This is especially true of terrorist organizations. Often the decision making process of terrorist organizations is not understood, and terrorist actions may seem unpredictable and counterintuitive. Therefore, the range of possible solutions bounded by the possible choices of the follower should be examined. To determine the lower bound, the pessimistic strategy is followed. This bound is determined by replacing the follower's objective function $f_2(y)$, and it is replaced by subtracting a small portion of the leader's objective function from the follower's: $f_2'(y) = f_2(y) - \varepsilon(f_1(x, y))$. Conversely, to find the upper bound, a small portion of the leader's objective function is added to the follower's objective function: $f_2'(y) = f_2(y) + \varepsilon(f_1(x, y))$. [17]

The optimistic and pessimistic strategies bound the leader's objective function based on potential optimal solution selection by the follower. Fortunately, if the leader has some understanding of the follower's selection process, he can mitigate non-cooperative selections by choosing alternate solutions. Specifically, this can be done using the model developed by Cao and Leung as follows: [31, p. 138]

$$\begin{aligned}
\max_{x} \quad & c_1x + \beta c_2 y + (1-\beta)c_2 z \\
\text{s.t.} \quad & A_1 x \leq x \leq b_1 \\
\max_{y} \quad & c_3 y + c_3 z - \varepsilon c_2 z \\
\text{s.t.} \quad & A_3 y \leq b_2 - A_2 x \\
& A_3 z \leq b_2 - A_2 x \\
& x, y, z \geq 0
\end{aligned} \tag{7.1}$$

where β is the expected "cooperation" level of the follower. The cooperation level can refer to an actual level of expected cooperation, a level the follower can be "influenced" to choose, or simply the expected choice based on past experience and/or subjective knowledge of the follower's decision making process.

In addition, Cao and Leung proved that the optimistic solution to this problem will find the optimal solution for the given cooperation level. [31, p. 138] Moore demonstrated how to solve problems of the form (7.1) by replacing the follower's model with its equivalent KKT conditions (as discussed in Section 2.3.6). Solving the reformulated model for various cooperation levels gives the decision maker the set of solutions and reactions based on the predicted level of cooperation/non-cooperation.

7.3.1 Example

Consider the following example (from [31, p. 135])

$$\begin{aligned} \max_{x} \quad & 8x_1 + 10x_2 + 2y_1 - y_2 \\ \text{s.t.} \quad & x_1 + x_2 \leq 10 \\ \max_{y} \quad & y_1 + y_2 \\ \text{s.t.} \quad & y_1 + y_2 - x_1 + x_2 \leq 20 \\ & x_1, x_2, y_1, y_2 \geq 0 \end{aligned} \tag{7.2}$$

By solving for the pessimistic and optimistic solutions, it is found that the leader's objective function could vary between 50 and 140 depending on the choice of the follower among his optimal solutions. However, by using the process described in Section 7.3, it is shown that (based on the predicted cooperation level), the decision maker (leader) can actually bound the problem between 90 and 140. To do this, (7.2) is transformed to the form of (7.1).

$$\begin{aligned}
\max_{x} \quad & 8x_1 + 10x_2 + \beta(2y_1 - y_2) + (1-\beta)(2z_1 - z_2) \\
\text{s.t.} \quad & x_1 + x_2 \leq 10 \\
\max_{y} \quad & y_1 + y_2 + z_1 + z_2 - \varepsilon(2z_1 - z_2) \\
\text{s.t.} \quad & y_1 + y_2 - x_1 + x_2 \leq 20 \\
& z_1 + z_2 - x_1 + x_2 \leq 20 \\
& x_1, x_2, y_1, y_2, z_1, z_2 \geq 0
\end{aligned} \tag{7.3}$$

Now, using the reformulation technique of Bard and Moore, the following model results: [11]

$$\begin{aligned}
\max_{x} \quad & 8x_1 + 10x_2 + \beta(2y_1 - y_2) + (1-\beta)(2z_1 - z_2) \\
\text{s.t.} \quad & x_1 + x_2 \leq 10 \\
& (1-u_1)y_1 = 0 \\
& (1-u_1)y_2 = 0 \\
& ((1-2\varepsilon) - u_2)z_1 = 0 \\
& ((1+\varepsilon) - u_2)z_2 = 0 \\
& 1 - u_1 + \nu_1 = 0 \\
& 1 - u_1 + \nu_2 = 0 \\
& (1-2\varepsilon) - u_2 + \nu_3 = 0 \\
& (1+\varepsilon) - u_2 + \nu_4 = 0 \\
& u_1(20 + x_1 - x_2 - y_1 - y_2) = 0 \\
& u_2(20 - z_1 - z_2 + x_1 - x_2) = 0 \\
& y_1 + y_2 - x_1 + x_2 \leq 20 \\
& z_1 + z_2 - x_1 + x_2 \leq 20 \\
& x_1, x_2, y_1, y_2, z_1, z_2 \geq 0 \\
& u_1, u_2, \nu_1, \nu_2, \nu_3, \nu_4 \geq 0
\end{aligned} \tag{7.4}$$

Solving (7.4) for a variety of $\beta \in [0, 1]$ results in the graph in Figure 7.5.

As Figure 7.5 demonstrates, an analysis of potential follower selections among multiple optimal solutions can lead to better results for the leader. As discussed previously, if the leader followed a strict pessimistic strategy (and the follower indeed

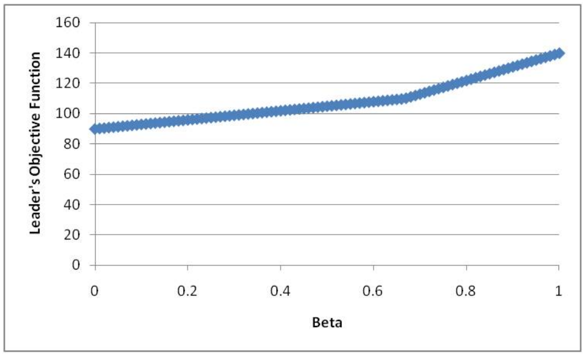


Figure 7.5: Multiple Follower Solutions Example

choose the worst solution for the leader, then the leader's objective function value would be 50. If the decision maker believes the follower will be uncooperative, he can choose the conservative strategy and do no worse than 90 (not the 50 under the pessimistic strategy). However, he can do no better than 110 using this strategy. Therefore, if he believes the follower is likely to be more cooperative (or can be influenced or decieved to chose a higher level), then he should switch to the cooperative strategy.

To achieve these results, the leader should choose $x_1 = 0$ and $x_2 = 10$ if the decision maker believes the follower will be cooperative at the 0.67 level or less. However, if the decision maker believes the follower will be more cooperative than that, then $x_1 = 10$ and $x_2 = 0$ should be chosen. This change in decision variables is reflected as a change in the slope in the graph in Figure 7.5.

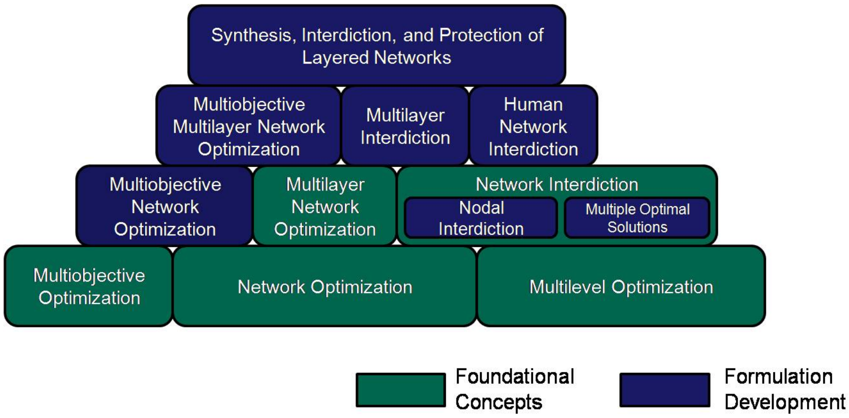
This example demonstrates how a decision maker can take subjective information into account when making decisions. Because the follower has multiple optimal solutions to choose from, the decision maker can evaluate likely decisions by the follower.

7.4 Summary

This chapter illustrated the importance of considering multiple optimal solutions when solving network interdiction and related problems. Both advantages and disadvantages of multiple optimal solutions were illustrated. In traditional bilevel/multilevel programming, it was shown how a technique developed to predict coalitions can be modified to give decision makers a suite of solutions of potential actions/reactions to determine a proper course of action.

VIII. Summary

Figure 8.1 once again presents an overview of how the chapters build on each other and fit together to provide a foundation for the synthesis, interdiction, and protection of infrastructures. Although the models/theory developed are general in nature, specific attention is reserved for infrastructure networks because our success as a nation may depend on our ability to protect our infrastructures. Infrastructures are the basis of our economy, wealth, and power. Each chapter developed the foundation and theory for a set of analysis techniques to assist decision makers in each of these areas.



Synthesis, Interdiction, and Protection of Layered Networks
Multiobjective Multilayer Network Optimization
Multilayer Interdiction
Human Network Interdiction
Multiobjective Network Optimization
Multilayer Network Optimization
Network Interdiction
Nodal Interdiction
Multiple Optimal Solutions
Multiobjective Optimization
Network Optimization
Multilevel Optimization
Foundational Concepts
Formulation Development

Figure 8.1: Dissertation Blueprint

Specifically, new models and solution techniques were developed for the nodal interdiction, synthesis of robust networks, multilayer interdiction, and human network interdiction. In addition, the importance and effects of multiple optimal solutions for these (and similar) models is discussed.

8.1 Theoretical Developments

This research contributed to the modeling and application of Operations Research in nodal interdiction, social network interdiction, multilayered network interdiction, and cost and robustness modeling in layered networks.

In the nodal interdiction chapter (Chapter III), network interdiction models were extended to directly include node interdiction. Previous literature only directly considered edge interdiction, with nodes needing to be substituted out through node-splitting. The extensions in Chapter III were done through the introduction of variables and constraints to model the impact of an attack on a node to associated edges. In addition, modifications to the resource constraints were developed to allow node only interdiction, node and edge interdiction with shared resources, and node and edge interdiction with different resources providing the opportunity for the analyst to directly model a wider array of operational settings.

In the social network interdiction chapter (Chapter IV), a new technique to target individual relationships based on shortest-path network interdiction was presented. It was shown that this idea can be extended by summing shortest path interdiction across all an individual's relationships. This new model combined the shortest path network interdiction model and the social network measure of closeness centrality. In addition, a model to disrupt the group as a whole was developed by extending shortest path network interdiction to an all-pairs shortest path network interdiction. It was shown how this proposed all-pairs formulation can also be used to target individual members of a network to maximize social closeness (as defined by closeness centrality). Both of these models (individual relationship interdiction and all-pairs interdiction) were also extended and combined into nodal interdiction by the addition of variables/constraints from the nodal interdiction chapter. These formulations were modified via Benders' partitioning and solution algorithms were developed, providing a methodology to investigate larger networks.

In the multilayered network interdiction chapter (Chapter V), network interdiction modeling was extended to layered network formulations. These extensions identify the maximum protection/disruption possible across layered networks with limited resources. This is accomplished primarily through the introduction of a "commonalities" variable that impacts/effects multiple networks via a single disruption. Resource constraints are also modified to account for both inter-network and intra-network attacks and effects. This formulation is also combined with nodal interdiction to present a new model which allows both edges and nodes to be interdicted across these multiple layers. A Benders' partitioning version of the formulation was developed and demonstrated which provides the opportunity to consider/analyze larger layered networks.

The synthesis chapter (Chapter VI), extends the formulations for robust network design to maximize robustness of a single layer network subject to budget constraints. Using ε-constraint methods, it was shown that the tradeoff between cost and robustness could be determined. This formulation was further modified to consider flow requirements (via the addition of flow constraints) as well as connectivity requirements. Finally, the model was modified to allow for considerations across multiple layered networks. The primary addition is a constraint that ensures that if an interdependent set of edges is selected for inclusion, then the corresponding individual network components are also selected (but with no additional increase in cost). The model as formulated, finds the most robust layered network possible given the budget limitation while ensuring that the demands are met. This will allow the decision maker to allocate resources to build the most robust network possible that meets flow demands with the budged constraint. It is also shown that using the ε-constraint, again allows an analysis of the effect of changing budgets on the robustness of the system (and finds all Pareto optimal solutions).

8.2 Application Developments

All models developed in this research were implemented in research level code following the guidelines of the Committee on Algorithms (COAL). With these tools, computational testing along with moderately realistic applications were considered to demonstrate the potential of the theory/models.

8.2.1 Multiple Optimal Solutions

This research also studied the potential impact of multiple optimal solutions in network-type interdictions. First, the dangers of ignoring alternate optimal solutions in pure network interdiction problems were discussed; this was followed by a discussion of the opportunities presented by multiple optimal solutions. The issue of multiple optimal solutions for followers in multilevel programming was also presented and discussed. It was observed that instead of only examining optimistic and pessimistic strategies, techniques that determine the range of potential solutions resulting from multiple optimal follower solutions needs to be followed.

8.2.2 Infrastructure Protection

This goal of the research was to develop theory with applied capabilities for the synthesis, interdiction, and protection of layered networks. The previous chapters have developed tools which aid in this analysis. While the theory and tools have general applications to single networks and system of systems of networks, they have been developed and built to be especially suited for infrastructure modeling. This section summarizes current infrastructure guidance and illustrates where this research effort can potentially be applied.

Critical infrastructures and their vital role in the nation's health was discussed in Section 2.2.8. Asymmetric adversaries, both foreign and domestic, pose a risk to the availability and efficiency of critical infrastructures. Following the 9/11 attacks, considerable investments have been made in protecting infrastructure networks. However, there are insufficient resources and funding to fully protect all the nation's critical infrastructures. In addition, much of this infrastructure is in private hands. Even if resources were greatly increased, it would be fiscally and physically impossible to protect everything, particularly while an open society is maintained. As Frederick the Great is reported to have said, "he who tries to protect everything protects nothing."

Therefore, continued analysis is needed to investigate how to optimally invest in infrastructure protection, i.e., how to obtain the most protection with limited resources. This is especially important in light of both potential terrorist attacks and state-sponsored asymmetric warfare being suggested by rising powers such as China (see *Unrestricted Warfare* [98]).

Most critical infrastructures are in the hands of the private sector whose first responsibility is to shareholders, and not necessarily to homeland security. Therefore, it may be difficult to motivate some private firms to invest in target hardening. Lewis suggests approaching this problem by coupling investments in security with productivity and efficiency enhancements. [81, p. 7] This would achieve greater security through redundancy, providing a cushion against both heavy loading and failure. [81, p. 20] The question remains how to best allocate a budget to protect an infrastructure against damage.

8.2.2.1 Allocation Strategies

Lewis suggests four allocation strategies are available: ranked vulnerability, apportioned vulnerability, optimal vulnerability, and manual vulnerability reduction. [81, p. 146]

Ranked Vulnerability/Allocation

Ranked allocation funds the highest-ranking components first, the second-highest next, and so forth; where ranking is defined in terms of vulnerability or risk. This is the most commonly used strategy by practitioners. [81, p. 145] For example, this is how DHS and infrastructure protection plans indicate allocation should be done. [130, p. 23]

The *National Infrastructure Protection Plan* recognizes that “Resources must be directed to areas of greatest priority to enable effective management of risk.” [92, p. 91] Therefore, the NIPP serves as a “unifying framework to ensure that critical infrastructures and key resource investments are coordinated and address the highest priorities ...” [92, p. 91] To do this, the NIPP directs the DHS to combine information from sector specific reports and state reports to assess the protection status and requirements. Based on this the “DHS will develop funding recommendations for programs and initiatives designed to reduce national-level risk in the critical infrastructure / key resources protection mission area.” [92, p. 92-93]

> However, a prioritized list of defended assets has a serious flaw for our applications. Such a list creates a preferred set of n+1 assets by adding one asset to the preferred set of size n. But, we know that an optimal set of size n and an optimal set of size n+1 may have nothing in common. For instance, a community with funds to build a new facility for one bomb disposal truck would select the most central location. However, if the community has money available for two facilities and two trucks, it would select two completely different facility locations, based on their ability to provide better average response time. [28, p. 531]

While a single prioritized list may not be the best solution, a range of solutions to present to decision makers is logical. For example, Chapter VI developed a technique to maximize robustness across layered networks subject to a cost constraint. It was shown that using the elastic-constraint method, the cost can be varied to find the efficient front (Pareto-optimal solutions). Using the proposed approach, decision makers can be presented with the trade-offs between budget and increased robustness in making the selection of assets to build.

Similarly, the vulnerabilities identified through interdiction techniques (whether traditional, nodal, social network, or layered network) are all dependent on the attacker’s resource constraint(s). However, as was demonstrated in each respective chapter, these techniques can be parameterized by cost to find the tradeoffs between attacker’s cost and the identification of vulnerable systems.

Optimal Allocation

Optimal allocation minimizes a combined vulnerability across all networks. Developing tools to assist this analysis has been the focus of this research. Of course, there are various ways to define this vulnerability function: minimize the probability of any event, minimize the probability of the worst event, minimize the probability of expected events, and so forth. This strategy seeks to find the absolute minimum vulnerability given cost estimates. While it seems ideal, “it is brutal in its exactness, leaving some components unfunded.” [81, p. 187]

Chapter VI developed techniques to consider cost versus robustness in network design. These concepts were further developed to allow the maximization of robustness across multiple layers of networks. This has direct application in infrastructure networks, which themselves are layered networks. The tools developed in Chapter VI will help a decision maker decide which facilities to construct (or fund construction) to maximize robustness across the layered infrastructure networks.

Instead of designing a network to a desired (or maximized) robustness level, another approach to robust network design is to design a network subject to predicted interdictions. For example, Garg and Smith start with the multicommodity network design formulation and modify it to account for potential failure scenarios. These scenarios, which are inputs to the model, are the set of arcs that could all fail at any one time. [60, p. 2] However, this method requires prediction of “likely” attacks, which requires subjective beliefs from subject matter experts.

Instead of trying to predict likely attacks, trilevel models can build the network synthesis problem given worse-case (optimal) network interdictions. Smith *et al.* took a step in this direction with [110]. Their trilevel model can be visualized as three stages. First, a network designer seeks to construct and/or expand a network subject to a budget constraint (note: this budget constraint could also be a second objective and multiobjective techniques could be used, but this was not done by these authors) on arc construction costs. In the second stage, an enemy interdicts this network to minimize the maximum flow. Finally, in the third stage, the network users solve a minimum cost network flow problem.

In their formulation, the designer’s objective function is “a weighted combination of flow profits before and after enemy interdiction minus arc construction costs.”

[110, p. 4] Of course, part of this objective function involves solving the network interdiction model (which is itself a bilevel program reformulated as a single level program). Smith *et al.* also considered heuristic interdiction strategies such as greedily destroying the largest-capacity arcs. However, as they point out in their analysis, these heuristic strategies only achieve 50% of the interdiction possible compared to an optimal interdiction. [110, p. 25]

Regardless of the methodology used, some research indicates large improvements in reliability can often be obtained with small increases in design cost. This is because many of these types of problems have many near-optimal solutions, "some of which may have desirable properties like reliability." [112, p. 4] Therefore, Snyder *et al.* suggest developing trade-off curves between formulations that account for failures and those that do not. To do this, they suggest using a weighted sum of the two objectives, where various weights will generate all non-dominated solutions. [112, p. 12]

In addition, the issues discussed in Chapter VII should be considered in any modeling effort involving networks where multiple optimal solutions may be a concern.

Apportioned & Manual Allocation

Apportioned allocation is a "middle of the road strategy that meets two objectives: (1) reduce risk and (2) fund as many counter-threat target hardening projects as possible." [81, p. 172] In essence, this is the technique used to "satisfy politicians" by reducing risk while funding projects across multiple congressional districts to satisfy political concerns.

Similarly, a policy maker may take subjective and/or intangible considerations into account and allocate resources manually. For example, while not essential to the operation of any other critical infrastructures and key assets, the destruction of a national icon such as the Statue of Liberty would create unknown psychological damage to the nation. [81, p. 145] This would seem to advocate using a "soft" operations research (OR) approach such as value-focused thinking to determine a protection strategy. For example, in [96] Pruitt developed such a hierarchy that could be used.

8.3 Future Research

8.3.1 Nodal Interdiction

The nodal interdiction formulations developed in Chapter III do not require the node splitting techniques required in traditional nodal interdiction. Therefore, without the artificial nodes and edges, it is believed that a robust sensitivity analysis may be possible. This sensitivity analysis would provide the decision maker with additional information about the nodes and their removal which is not possible with traditional interdiction methods.

8.3.2 Social Network Interdiction

Sociologists use measures adapted from graph theory in developing centrality measures. However, many of these measures are computationally inefficient and not easily extended. Therefore, research needs to be conducted to incorporate more operations research techniques into social network analysis. Chapter IV provided the foundation for one potential proxy measure by using shortest path network interdiction. By analyzing the network before and after potential interdictions, a measure of difference could be developed to more accurately determine a person's closeness and/or potential for targeting.

This method could be combined with the techniques developed to measure Network Centric Warfare by Wong-Jiru *et al.* (as discussed in Section 2.2.6.4). In this way, a comparison of the social networks prior to and after an interdiction could be quantified in several dimensions. The measure of vulnerability on these bases may capture more information than traditionally conveyed in traditional social network analysis measures.

More research needs to be conducted to determine the impact of missing information on these SNA measures. Most open source literature focuses on the impact of missing information on random networks (see [25]). However, social networks are not random. The impact of missing information on the models developed in Chapter IV needs to be determined. This would include missing information such as a cellular structure, unknown cells, unknown relationships, and so forth. Using a realistic social network generator (see [87]), one could test the effects of selected missing data. Techniques to create realistic social networks have been recently developed and should be used to test these models/measures.

Benders' partitioning version of the models developed in Chapter IV were presented, but they need to be further developed. In each case, the implementation of these models did not incorporate any techniques to increase efficiency. For example, in each iteration, the subproblems were created from scratch. It would be more efficient to store information from previous iterations and update them as necessary, instead of recreating them each time. In addition, techniques such as flow dispersion developed by Cormican could likely be used to reduce computational times of partitioning methods. [46]

The developments in Chapter IV focused on the disruption of enemy networks. However, there are other potential applications, particularly in a counter-insurgency environment. For example, with small modifications, it is believed that the models could be used to maximize closeness within a specified group while minimizing closeness between one or more groups. The goal in a counter-insurgency is to disrupt the terrorist group while winning the hearts and minds of the people. This involves bringing together (maximizing closeness of) coalition forces, government forces, and neutral forces; while minimizing the closeness of insurgent forces. There may also be potential to combine this model with strategy selection using coalition formulations discussed in Section 2.3.4. Again, this would allow the selection of a strategy that considers all aspects of coalition counter-insurgency action, not just what may be best for a single decision maker.

8.3.3 Multilayered Network Interdiction

As with social network interdiction, a Benders' partitioning version of the model developed in Chapter V was developed. Again, this implementation did not incorporate any techniques to increase the efficiency of the partitioning model. Techniques to increase computational efficiency need to be researched and applied to these models, including those discussed in the previous subsection: storing iteration information for reuse, flow dispersion, and so forth. These extensions will facilitate the investigation of increasingly larger systems of systems.

8.3.4 Synthesis of Robust Networks

In Chapter VI, a method to consider cost versus robustness in layered networks was developed. However, this chapter focused on the edge disjoint version of the synthesis problem. Although less common, a model which also considers robustness in terms of node disjoint paths also needs to be developed and incorporated with the edge formulation.

In addition, although the theoretical developments of Chapter VI are sound, additional developments are necessary to broaden the models' applications to large networks. The application of heuristics developed for similar models (such as the traveling salesman) may provide some necessary time savings. Parametric programming on the models presented here will provide additional post-optimality analysis to aid the decision maker.

One of the developments of Chapter VI is the formulation in which the network designer has an explicit objective function to maximize robustness subject to a cost constraint. A potential application of this tool is to combine this objective function with the trilevel model developed by Smith *et al.* For example, the lowest level would be the network user who seeks network flow across multiple layers. The second level (above the network user) would be modeled as an attacker seeking optimal network interdiction against this layered network. Finally, the top level decision maker, would be the decision maker (such as the government or a combatant commander) who allocates resources to maximize robustness subject to this potential network interdiction.

In such a proposed formulation, the inner (two) levels will still be network interdiction problems. In addition, the interdiction model across layered networks was also developed in Chapter V which could be used for this inner formulation. Of course, if nodes are a consideration, along with edges, then the tools on nodal interdiction (Chapter III) could be used as well. If the networks involved contain social networks, then techniques and models developed in Chapter IV should be considered as well.

8.3.5 Multiple Optimal Solutions

Section 7.3 introduced the need to examine cases of non-symmetric interdiction where attackers and defenders do not have diametrically opposed objective functions. For example, it is the goal of infrastructure operators to maximize efficiency or minimize cost. However, the terrorist goal is to maximize loss of life (CNN effect), not maximize cost of operating infrastructure systems. The impact of these differing objective functions needs to be explored in more depth.

The impact of secrecy and deception of one or more players needs to be explored. Multilevel programs assume rational players with perfect information available to all players. An examination of the impact of relaxing these assumptions needs to be done. Both traditional game theory and stochastic interdiction may allow some insight. As part of this research, a model needs to be built to allow for partial information, beliefs/probabilities, and so forth. Part of this modeling effort should incorporate behavioral modeling.

8.4 Concluding Remarks

This research developed the foundation, theory, and framework for a set of analysis techniques to assist decision makers in analyzing questions regarding the synthesis, interdiction, and protection of infrastructure networks. While there is still ample work to do, it is hoped that the suite of analysis techniques developed will assist decision makers and aid in national defense. Specifically, it is hoped that this research will help identify critical people, relationships, and/or assets to attack/exploit or protect; maximally protect or disrupt layered networks with limited resources; balance robustness, cost, and risk in designing or expanding networks; and provide decision makers with potential ramifications of multiple optimal solutions.

Bibliography

1. *AFDD 2-5 Information Operations.* 2005.

2. Ravindra K. Ahuja, Thomas L. Magnanti, and James B. Orlin. *Network Flows: Theory, Algorithms, and Applications.* Prentice Hall, 1993.

3. Ibrahim Akgun. The K-Group Maximum-Flow Network-Interdiction Problem. Master's thesis, Naval Postgraduate School, 2000.

4. D.L. Alderson. Catching the Network Science Bug: Insight and Opportunity for the Operations Researcher. *Operations research*, page 1047, 2008.

5. Massoud Amin. Toward Secure and Resilient Interdependent Infrastructures. *Journal of Infrastructure Systems*, 8:67, 2002.

6. G. Anandalingam. A Mathematical Programming Model of Decentralized Multi-level Systems. *Journal of the Operational Research Society*, 39(11):1021–1033, 1988.

7. G. Anandalingam and V. Apprey. Multi-level Programming and Conflict Resolution. *European Journal of Operational Research*, 51(2):233–247, 1991.

8. G. Anandalingam and D.J. White. A Solution Method for the Linear Static Stackelberg Problem Using Penalty Functions. *Automatic Control, IEEE Transactions on*, 35(10):1170–1173, 1990.

9. Jonathan F. Bard. An Investigation of the Linear Three Level Programming Problem. *IEEE Transactions on Systems, Man, and Cybernetics*, 14(5):711–717, 1984.

10. Jonathan F. Bard. *Practical Bilevel Optimization: Algorithms and Applications.* Kluwer Academic Publishers, 1998.

11. Jonathan F. Bard and James T. Moore. A Branch and Bound Algorithm for the Bilevel Programming Problem. *SIAM Journal on Scientific and Statistical Computing*, 11:281, 1990.

12. M. Bellmore, G. Bennington, and S. Lubore. A Network Isolation Algorithm. *Naval Research Logistics Quarterly*, 17:461–469, 1969.

13. Omar Ben-Ayed and Charles E. Blair. Computational Difficulties of Bilevel Linear Programming. *Operations Research*, 38(3):556–560, 1990.

14. Jacques F. Benders. Partitioning Procedures for Solving Mixed-Variables Programming Problems. *Numerische Mathematik*, 4(3), 1962.

15. D.P. Bertsekas. *Linear network optimization: algorithms and codes*. MIT Press, 1991.

16. Wayne Bialas and Mark Karwan. On Two-level Optimization. *Automatic Control, IEEE Transactions on*, 27(1):211–214, 1982.

17. Wayne Bialas and Mark Karwan. Two-Level Linear Programming. *Management Science*, 30(8):1004–1021, 1984.

18. Wayne F. Bialas. Cooperative n-Person Stackelberg Games. *Decision and Control, 1989., Proceedings of the 28th IEEE Conference on*, pages 2439–2444, 1989.

19. Wayne F. Bialas and Mark N. Chew. Coalition Formation in n-Person Stackelberg Games. *Proceedings of the 21st IEEE Conference on Decision and Control*, pages 761–765, 1980.

20. Levent Bingol. A Lagrangian Heuristic for Solving a Network Interdiction Problem. Master's thesis, Naval Postgraduate School, 2001.

21. Paul T. Boggs, Stephen G. Nash, and Susan Powell. Guidelines for Reporting Results of Computational Experiments. Report of the Ad Hoc Committee. *Mathematical Programming: Series A and B*, 49(3):413–425, 1991.

22. Everett M.G. Borgatti, S.P. and L.C. Freeman. Ucinet for windows: Software for social network analysis., 2002.

23. Steve P. Borgatti. Identifying Sets of Key Players in a Social Network. *Computational & Mathematical Organization Theory*, 12(1):21–34, 2006.

24. Steve P. Borgatti, Kathleen M. Carley, and David Krackhardt. On the Robustness of Centrality Measures Under Conditions of Imperfect Data. *Social Networks*, 2005.

25. Steve P. Borgatti, Kathleen M. Carley, and David Krackhardt. On the Robustness of Centrality Measures Under Conditions of Imperfect Data. *Social Networks*, 28(2):124–136, 2006.

26. Gerald Brown, Matt Carlyle, Terry Harrison, Javier Salmerón, and Kevin Wood. How to Attack a Linear Program. *71st Military Operations Research Society Symposium, Quantico, VA, June*, pages 10–12, 2003.

27. Gerald Brown, Matt Carlyle, Javier Salmerón, and Kevin Wood. Analyzing the Vulnerability of Critical Infrastructure to Attack and Planning Defenses. *INFORMS Tutorials in Operations Research, Institute for Operations Research and the Management Sciences, Hanover, MD*, pages 102–123, 2005.

28. Gerald Brown, Matt Carlyle, Javier Salmerón, and Kevin Wood. Defending Critical Infrastructure. *Interfaces*, 36(6):530–544, 2006.

29. Patrick S. Brown. Optimizing the Long-Term Capacity Expansion and Protection of Iraqi Oil Infrastructure. Master's thesis, Naval Postgraduate School, 2005.

30. Richard K. Bullock, Richard F. Deckro, and Jeffery D. Weir. Methodology for competitive strategy development. *Computers and Operations Research*, 35(6): 1865–1873, 2008.

31. D. Cao and L.C. Leung. A partial cooperation model for non-unique linear two-level decision problems. *European Journal of Operational Research*, 140 (1):134–141, 2002.

32. Kathleen M. Carley. Dynamic Network Analysis. *Dynamic Social Network Modeling and Analysis: Workshop Summary and Papers*, pages 133–145, 2003.

33. Kathleen M. Carley. Estimating Vulnerabilities in Large Covert Networks. *Proceedings of the 9 thInternational Command and Control Research and Technology Symposium*, 2004.

34. Arthur K. Cebrowski. The Implementation of Network-Centric Warfare. *Office of Force Transformation, Department of Defense*, 2005.

35. Joint Warfighting Center. *Commanders Handbook for an Effects-Based Approach to Joint Operations*. Suffolk, Va.: US Joint Forces Command, 2006.

36. V. Chankong Chankong and Yacov Y. Haimes. *Multiobjective Decision Making: Theory and Methodology*. North-Holland NY, 1983.

37. Harold Chestnut. *Systems Engineering Methods*. New York: Wiley, 1967.

38. Mark N. Chew. A Game Theoretic Approach to Coalition Formation in Multilevel Decision Making Organizations. Master's thesis, State University of New York at Buffalo, 1981.

39. Richard L. Church and Maria P. Scaparra. Protecting Critical Assets: the r-Interdiction Median Problem with Fortificatin. *Geographical Analysis*, 39: 129–146, 2007.

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

41. Clinton R. Clark. Modeling and Analysis of Clandestine Networks. Master's thesis, Air Force Institute of Technology, 2005.

42. Lloyd W. Clarke and G. Anandalingam. A Bootstrap Heuristic for Designing Minimum Cost Survivable Networks. *Computers and Operations Research*, 22 (9):921–934, 1995.

43. William J. Clinton. Executive Order 13010-Critical Infrastructure Protection. *Government Printing Office, Washington, DC*, 1996.

44. William J. Clinton. Presidential Decision Directive 63. *Government Printing Office, Washington, DC*, 1998.

45. US Congress. Homeland Security Act of 2002. *Public Law No*, 107, 2002.

46. Kelly J. Cormican. Computational Methods for Deterministic and Stochastic Network Interdiction Problems. Master's thesis, Naval Postgraduate School, 1995.

47. Norman Curet. The Network Diversion Model. *Military Operations Research*, pages 35–45, 2001.

48. E. Danna, M. Fenelon, Z. Gu, and R. Wunderling. Generating multiple solutions for mixed integer programming problems. *Lecture Notes in Computer Science*, 4513:280, 2007.

49. S. Dempe. *Foundations of bilevel programming.* Kluwer Academic Publishers, 2002.

50. H.D. Derbes. Efficiently Interdicting a Time-Expanded Transshipment Network. Master's thesis, Naval Postgraduate School, 1997.

51. Donald D. Dudenhoeffer, May R. Permann, and M. Manic. CIMS: A Framework for Infrastructure Interdependency Modeling and Analysis. *Proceedings of the 37th conference on Winter simulation*, pages 478–485, 2006.

52. Matthias Ehrgott. *Multicriteria Optimization.* Springer, 2005.

53. Matthias Ehrgott and David M. Ryan. Constructing Robust Crew Schedules with Bicriteria Optimization. *Journal of Multi-Criteria Decision Analysis*, 11 (3):139–150, 2002.

54. P. Erdos and A. Renyi. On random graphs. *Publ. Math. Debrecen*, 6:290–297, 1959.

55. Howard Frank and Ivan T. Frisch. *Communication, Transmission, and Transportation Networks.* Addison-Wesley Educational Publishers Inc., US, 1971.

56. Linton C. Freeman. A Set of Measures of Centrality Based on Betweenness. *Sociometry*, 40(1):35–41, 1977.

57. Linton C. Freeman. Centrality in Social Networks: Conceptual Clarification. *Social Networks*, 1(3):215–239, 1979.

58. D. Ray Fulkerson and Gary C. Harding. *Maximizing the Minimum Source-Sink Path Subject to a Budget Constraint: Another View of the Minimum Cost Flow Routine*, 1975. Research Report ADA014422.

59. D. Ray Fulkerson and Gary C. Harding. Maximizing the Minimum Source-Sink Path Subject to a Budget Constraint. *Mathematical Programming*, 13(1): 116–118, 1977.

60. Manish Garg and J.Cole Smith. Models and Algorithms for the Design of Survivable Multicommodity Flow Networks with General Failure Scenarios. *Omega*, 36(6):1057–1071, 2008.

61. John Garstka. *Network Centric Operations Conceptual Framework Version 1.0.* DTIC, 2003. Research Report ADA457620.

62. Jennifer L. Geffre. A layered social and operational network analysis. Master's thesis, Air Force Institute of Technology, 2007.

63. R.E. Gomory and T.C. Hu. Synthesis of a Communication Network. *J. Soc. Indust. Appl. Math*, 12:348–369, 1964.

64. Martin Grötschel, Clyde L. Monma, and Mechthild Stoer. Design of Survivable Networks. *Handbooks in OR and MS*, 7:617–672.

65. Yacov Y. Haimes and Barry M. Horowitz. Modeling Interdependent Infrastructures for Sustainable Counterterrorism. *Journal of Infrastructure Systems*, 10: 33, 2004.

66. Jonathan T. Hamill. *Analysis of Layered Social Networks*. PhD thesis, Air force Institute of Technology, 2006.

67. Pierre Hansen, Brigitte Jaumard, and Gilles Savard. New Branch-and-Bound Rules for Linear Bilevel Programming. *SIAM Journal on Scientific and Statistical Computing*, 13:1194, 1992.

68. Travis J. Herbranson. Isolating key players in clandestine networks. Master's thesis, Air Force Institute of Technology, 2007.

69. White House. Homeland Security Presidential Directive 7 (HSPD-7):Critical Infrastructure Identification. *Prioritization, and Protection[http://www. whitehouse. gov/news/releases/2003/12/20031217-5. html]*.

70. Eitan Israeli. *System Interdiction and Defense*. PhD thesis, Naval Postgraduate School, 1999.

71. Eitan Israeli and R. Kevin Wood. Shortest-path Network Interdiction. *Networks*, 40(2):97–111, 2002.

72. Kevin T. Kennedy. An analysis of multiple layered networks. Master's thesis, Air Force Institute of Technology, 2003.

73. Kevin T. Kennedy, Richard F. Deckro, James W. Chrissis, and Victor D. Wiley. An analysis of multiple layered networks. *Military Operations Research*, (forthcoming).

74. L. Khachiyan, E. Boros, K. Borys, K. Elbassioni, V. Gurvich, G. Rudolf, and J. Zhao. On short path interdiction problems: total and node-wise limited interdiction. RUTCOR Research Report RRR 25-2006, RUTCOR, Rutgers Center of Operations Research, 2006.

75. Jon A. Kimminau. The Psychology of Coercion: Merging Airpower and Prospect Theory. Master's thesis, School of Advanced Airpower Studies, 1998.

76. D. Klingman, A. Napier, and J. Stutz. NETGEN: A program for generating large scale capacitated assignment, transportation, and minimum cost flow network problems. *Management Science*, pages 814–821, 1974.

77. Valdis E. Krebs. The social life of routers. *The Internet Protocol Journal*, 3(4): 14–25, 2000.

78. Valdis E. Krebs. Uncloaking terrorist networks. *First Monday*, 7(4-1), 2002.

79. Larry J. LeBlanc and David E. Boyce. A Bilevel Programming Algorithm for Exact Solution of the Network Design Problem with User-Optimal Flows. *Transportation Research*, 20B(3):259–265, 1986.

80. Earl E. Lee, John E. Mitchell, and William A. Wallace. Assessing Vulnerability of Proposed Designs for Interdependent Infrastructure Systems. *Proceedings of the 37th Annual Hawaii International Conference on System Sciences*, pages 54–61, 2004.

81. Ted G. Lewis. *Critical Infrastructure Protection in Homeland Security: Defending a Networked Nation*. Wiley-Interscience, 2006.

82. Churlzu Lim and J. Cole Smith. Algorithms for Discrete and Continuous Multicommodity Flow Network Interdiction Problems. *IIE Transactions*, 39(1): 15–26, 2007.

83. Yibing Lv, Tiesong Hu, Guangmin Wang, and Zhongping Wan. A Penalty Function Method Based on Kuhn–Tucker Condition for Solving Linear Bilevel Programming. *Applied Mathematics and Computation*, 188(1):808–813, 2007.

84. Patrice Marcotte and Gilles Savard. A Note on the Pareto Optimality of Solutions to the Linear Bilevel Programming Problem. *Computers and Operations Research*, 18(4):355–359, 1991.

85. R. A. Miller and I. Lachow. Strategic Fragility: Infrastructure Protection and National Security in the Information Age. *Defense Horizons*, (59):1–6, 2008.

86. James Moore. *Extensions to the Multilevel Linear Programming Problem*. PhD thesis, School of Mechanical Engineering, University of Texas, 1988.

87. James F. Morris, Jerome W. O'Neal, and Richard F. Deckro. A random graph generation algorithm designed for social network analysis. FOIL Working Paper, June 2009.

88. John Moteff and Paul Parfomak. Critical Infrastructure and Key Assets: Definition and Identification. 2004.

89. Alan T. Murray, Timothy C. Matisziw, and Tony H. Grubesic. Critical Network Infrastructure Analysis: Interdiction and System Flow. *Journal of Geographical Systems*, 9(2):103–117, 2007.

90. William H. Nesbitt. Robust Interdiction of Covert Social Networks. Master's thesis, Naval Postgraduate School, 2006.

91. M.E.J. Newman. Random graphs as models of networks. *Arxiv preprint cond-mat/0202208*, 2002.

92. Department of Homeland Security. National Infrastructure Protection Plan. 2006.

93. C. Patvardhan, V.C. Prasad, and V. Prem Pyara. Vertex cutsets of undirected graphs. *IEEE transactions on reliability*, 44(2):347–353, 1995.

94. P. Pederson, Donald Dudenhoeffer, S. Hartley, and M. Permann. Critical Infrastructure Interdependency Modeling: A Survey of US and International Research. 2006.

95. Michael S. Pinkstaff. An approach to disrupting communication networks. Master's thesis, Air Force Institute of Technology, 2001.

96. Kristopher A. Pruitt, Richard F. Deckro, and Stephen P. Chambal. Modeling Homeland Security. *Journal of Defense Modeling and Simulation*, 1.

97. J. Pub. Jp 3-03. *Doctrine for Joint Interdiction*, 03 May 2007.

98. L. Qiao and W. Xiangsui. Unrestricted warfare. *Beijing, China: Peoples Liberation Army Literature and Arts Publishing House*, 1999.

99. Robert S. Renfro. *Modeling and Analysis of Social Networks*. PhD thesis, Air Force Institute of Technology, 2001.

100. Robert S. Renfro and Richard F. Deckro. A Flow Model Social Network Analysis of the Iranian Government. *Military Operations Research*, 8(1):5–16, 2003.

101. Steven M. Rinaldi, James P. Peerenboom, and Terrence K. Kelly. Identifying, Understanding, and Analyzing Critical Infrastructure Interdependencies. *Control Systems Magazine, IEEE*, 21(6):11–25, 2001.

102. Johannes O. Royset and R. Kevin Wood. Solving the Bi-Objective Maximum-Flow Network-Interdiction Problem. *INFORMS Journal on Computing*, 19(2): 175, 2007.

103. M. Sageman. *Understanding terror networks*. University of Pennsylvania Press, 2004.

104. Gilles Savard. *Contributions à la Programmation Mathématique à Dex Niveaux*. PhD thesis, Université de Montréal, École Polytechnique, 1989.

105. Maria P. Scaparra and Paola Cappanera. Optimizing Security Investments in Transportation and Telecommunication Networks. *INFORMS Annual Meeting, San Francisco, CA*, 2005.

106. Maria P. Scaparra and Richard L. Church. A Bilevel Mixed-integer Program for Critical Infrastructure Protection Planning. *Computers and Operations Research*, 35(6):1905–1923, 2008.

107. Chenggen Shi, Jie Lu, Guangquan Zhang, and Hong Zhou. An Extended Branch and Bound Algorithm for Linear Bilevel Programming. *Applied Mathematics and Computation*, 180(2):529–537, 2006.

108. S. Sinha. A Comment on Anandalingam (1988). A Mathematical Programming Model of Decentralized Multi-Level Systems. *Journal of the Operational Research Society*, 52(5):594–596, 2001.

109. J. Cole Smith. *Enhanced Cutting Plane Techniques for Bilevel Optimization Algorithms*. DTIC, 2008.

110. J. Cole Smith, Churlzu Lim, and Fransisca Sudargho. Survivable Network Design Under Optimal and Heuristic Interdiction Scenarios. *Journal of Global Optimization*, 38(2):181–199, 2007.

111. Jeffrey J. Smith. Communicating for Effect: Operationalizing and Analyzing Weapons of Mass Influence. Master's thesis, Air University: CADRE/AR – Air Force Fellows, 2008.

112. Lawrence V. Snyder and Mark S. Daskin. Models for Reliable Supply Chain Network Design. *Reliability and Vulnerability in Critical Infrastructure: A Quantitative Geographic Perspective*, 2006.

113. Lawrence V. Snyder, Maria P. Scaparra, Mark S. Daskin, and Richard L. Church. Planning for Disruptions in Supply Chain Networks. *TutORials in Operations Research, edited by H. Greenberg. Baltimore, Md.: INFORMS*, 2006.

114. Margareta Soismaa. A Note on Efficient Solutions for the Linear Bilevel Programming Problem. *European Journal of Operational Research*, 112(2):427–431, 1999.

115. Malcolm K. Sparrow. *The Application of Network Analysis to Criminal Intelligence: An Assessment of the Prospects*. Elsevier, 1991.

116. Robert G. Spulak and Jessica G. Turnley. *Theoretical Perspectives of Terrorist Enemies as Networks*. JSOU Press, 2005.

117. Sara Sterling. Aggregation techniques for social network analysis. Master's thesis, Air Force Institute of Technology, 2004.

118. Vaidy Sunderam, Geert van Albada, Peter Sloot, and Jack Dongarra. *Computational Science-ICCS 2005: 5th International Conference, Atlanta, Georgia, Proceedings, Part III.* Springer, 2005.

119. Maksim Tsvetovat and Kathleen M. Carley. Structural Knowledge and Success of Anti-Terrorist Activity: The Downside of Structural Equivalence. *Journal of Social Structure*, 6, 2005.

120. Adnan Uygun. Network Interdiction by Lagrandian Relaxation and Branch-and-Bound. Master's thesis, Naval Postgraduate School, 2002.

121. Luis N. Vicente and Paul H. Calamai. Bilevel and Multilevel Programming: A Bibliography Review. *Journal of Global Optimization*, 5(3):291–306, 1994.

122. William A. Wallace, David Mendona, Earl Lee, John Mitchell, and Joe Chow. Managing disruptions to critical interdependent infrastructures in the context of the 2001 world trade center attack. In M. F. Myers, editor, *Impacts of and Human Response to the September 11, 2001 Disasters: What Research Tells Us.* Natural Hazards Research and Applications Information Center, University of Colorado, 2007.

123. Qian Wang, Fengmei M. Yang, Shouyang Y. Wang, and Yi H. Liu. Bilevel Programs with Multiple Followers. *Systems Science and Mathematical Sciences*, 13(3):265–276, 2000.

124. Stanley Wasserman and Katherine Faust. *Social Network Analysis: Methods and Applications.* Cambridge University Press, 1994.

125. Ue-Pyng Wen and Wayne F. Bialas. The Hybrid Algorithm for Solving the Three-level Linear Programming Problem. *Computers and Operations Research*, 13(4):367–377, 1986.

126. Ue-Pyng Wen and Shuh-Tzy Hsu. Efficient Solutions for the Linear Bilevel Programming Problem. *European Journal of Operational Research*, 62(3):354–362, 1992.

127. D. J. White. Penalty Function Approach to Linear Trilevel Programming. *Journal of Optimization Theory and Applications*, 93(1):183–197, 1997.

128. D. J. White and G. Anandalingam. A Penalty Function Approach for Solving Bi-level Linear Programs. *Journal of Global Optimization*, 3(4):397–419, 1993.

129. House White. National Strategy for Homeland Security. *Washington, DC: The White House*, 2002.

130. House White. National Strategy for the Physical Protection of Critical Infrastructures and Key Assets. *Washington, DC: The White House*, 2003.

131. House White. National Strategy to Secure Cyberspace. *Washington, DC: The White House*, 2003.

132. Philip S. Whiteman. Improving Single Strike Effectiveness for Network Interdiction. *Military Operations Research*, 4:15–30, 1999.

133. Philip S. Whiteman. A Target Selection Tool for Network Interdiction. In *64th MORS Symposium Working Group 1*, 1996.

134. William J. Willick. Solution Techniques for Cooperative N-Person Stackelberg Games. Master's thesis, State University of New York, 1990.

135. William J. Willick. *A Power Index for Cooperative Games with Applications to Hierarchical Organization*. PhD thesis, State University of New York, 1995.

136. Ann Wong-Jiru, John Colombi, Laura Suzuki, and Robert Mills. Graph theoretical analysis of network centric operations using multi-layer models. In *CSER 2007*, pages 1–10.

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

138. Yiming Yao, Thomas Edmunds, Dimitri Papageorgiou, and Rogelio Alvarez. Trilevel Optimization in Power Network Defense. *Applications and Reviews, IEEE Transactions on Systems, Man, and Cybernetics*, 37(4):712–718, 2007.

139. Pengcheng Zhang, Srinivas Peeta, and Terry Friesz. Dynamic Game Theoretic Model of Multi-Layer Infrastructure Networks. *Networks and Spatial Economics*, 5(2):147–178, 2005.

140. Wang Zhi-Wei, Nagasawa Hiroyuki, and Nishiyama Noriyuki. An Algorithm for a Multiobjective, Multilevel Linear Programming. *Journal of the Operations Research Society of Japan*, 39(2):176–187, 1996.

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704–0188

1. REPORT DATE *(DD–MM–YYYY)*
10–09–2009

2. REPORT TYPE
Doctoral Dissertation

3. DATES COVERED *(From — To)*
Sept 2006 — Sep 2009

4. TITLE AND SUBTITLE
Synthesis, Interdiction, and Protection of Layered Networks

5a. CONTRACT NUMBER

5b. GRANT NUMBER

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)
Kevin T. Kennedy, Major, USAF

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Air Force Institute of Technology
Graduate School of Engineering and Management (AFIT/EN)
2950 Hobson Way
WPAFB OH 45433-7765

8. PERFORMING ORGANIZATION REPORT NUMBER
AFIT/DS/ENS/09-01

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)
Joint DoD Organization

10. SPONSOR/MONITOR'S ACRONYM(S)

11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION / AVAILABILITY STATEMENT
APPROVED FOR PUBLIC RELEASE; DISTRIBUTION UNLIMITED

13. SUPPLEMENTARY NOTES
This material is declared a work of the U.S. Government and is not subject to copyright protection in the United States.

14. ABSTRACT This research developed the foundation, theory, and framework for a set of analysis techniques to assist decision makers in analyzing questions regarding the synthesis, interdiction, and protection of infrastructure networks. This includes extension of traditional network interdiction to directly model nodal interdiction; new techniques to identify potential targets in social networks based on extensions of shortest path network interdiction; extension of traditional network interdiction to include layered network formulations; and develops models/techniques to design robust layered networks while considering trade-offs with cost. These approaches identify the maximum protection/disruption possible across layered networks with limited resources, find the most robust layered network design possible given the budget limitations while ensuring that the demands are met, include traditional social network analysis, and incorporate new techniques to model the interdiction of nodes and edges throughout the formulations. In addition, the importance and effects of multiple optimal solutions for these (and similar) models is investigated. All the models developed are demonstrated on notional examples and were tested on a range of sample problem sets.

15. SUBJECT TERMS
network interdiction, social networks, infrastructure protection

16. SECURITY CLASSIFICATION OF:			**17. LIMITATION OF ABSTRACT**	**18. NUMBER OF PAGES**	**19a. NAME OF RESPONSIBLE PERSON**
a. REPORT	**b. ABSTRACT**	**c. THIS PAGE**			Richard F. Deckro, AFIT/ENS
U	U	U	UU	243	**19b. TELEPHONE NUMBER** *(include area code)* (937) 255–3636, ext 4325

AFIT Air Force Institute of Technology

BLPP bilevel programming problem

BLPPs bilevel programming problems

C4I Command, Control, Communications, Computers, and Intelligence

CI counter-intelligence

CNAT Critical Network Analysis Tool

COAL Committee on Algorithms

COG center of gravity

COIN-OR COmputational INfrastructure for Operations Research

CONUS continental United States

DHS Department of Homeland Security

DIMACS Center for Discrete Mathematics and Theoretical Computer Science

DoD Department of Defense

DTIC Defense Technical Information Center

FOUO For Official Use Only

GAMS General Algebraic Modeling System

GIS Geographic Information System

GUI Graphical User Interface

HIIM Holistic Interpersonal Influence Measure

HSIP Homeland Security Infrastructure Program

ISR Intelligence, Surveillance, and Reconnaissance

JFC Joint Forces Commander

JUNG Java Universal Network/Graph

LNM Layered Network Model

KKT Karush-Kuhn-Tucker

LP linear program

MANA Map Aware Non-Uniform Automata

MILDEC military deception

MIP Mixed Integer Program

MORS Military Operations Research Society

NART Network Analysis Research Tool

NCW Network Centric Warfare

NGA National Geospatial-Intelligence Agency

NIPP National Infrastructure Protection Plan

OA operational assessment

OPSEC operations security

OR operations research

PA public affairs

PDD Presidential Decision Directive

PSYOP psychological operations

SLLNIM single level layered network interdiction model

SLNIM single level network interdiction model

SCADA Supervisory Control and Data Acquisition

SLNIM-MP master problem of the single level network interdiction

SLNIM-SP sub problem of the single level network interdiction

SNA social network analysis

SoSA system of systems analysis

SVI supervalid inequality

US United States

VBA Visual Basic for Applications

VFT Value Focused Thinking