Interfaces
Vol. 36, No. 6, November–December 2006, pp. 530–544
ISSN 0092-2102 | EISSN 1526-551X | 06 | 3606 | 0530
DOI 10.1287/inte.1060.0252

Defending Critical Infrastructure

Gerald Brown, Matthew Carlyle, Javier Salmerón, Kevin Wood
Operations Research Department, Naval Postgraduate School, Monterey, California 93943
{gbrown@nps.edu, mcarlyle@nps.edu, jsalmero@nps.edu, kwood@nps.edu}

We apply new bilevel and trilevel optimization models to make critical infrastructure more resilient against terrorist attacks. Each model features an intelligent attacker (terrorists) and a defender (us), information transparency, and sequential actions by attacker and defender. We illustrate with examples of the US Strategic Petroleum Reserve, the US Border Patrol at Yuma, Arizona, and an electrical transmission system. We conclude by reporting insights gained from the modeling experience and many "red-team" exercises. Each exercise gathers open-source data on a real-world infrastructure system, develops an appropriate bilevel or trilevel model, and uses these to identify vulnerabilities in the system or to plan an optimal defense.

Key words: critical infrastructure protection; bilevel program; trilevel program; mixed-integer program; homeland security; homeland defense.
History: This paper was refereed.

Our national strategy for homeland security deems these 13 infrastructure sectors critical to the United States: agriculture, banking and finance, chemical industry, defense industrial base, emergency services, energy, food, government, information and telecommunications, postal and shipping, public health, transportation, and water (Department of Homeland Security 2002, p. 30). In this paper, we introduce methods to identify vulnerabilities in these critical sectors and plan defensive measures. We also expand on conclusions found in a tutorial by Brown et al. (2005a).

Any critical-infrastructure system represents an enormous public investment. Even a minor disruption, randomly or deliberately caused, can degrade the system's performance and inflict substantial economic losses. How do we analyze the vulnerability of such a system to a set of coordinated terrorist attacks, and make informed proposals for reducing that vulnerability?

The techniques of system-reliability analysis have been proposed for gauging vulnerability (Garcia 2001, pp. 39–48). For example, real-time reliability assessment of an electric power grid may pronounce the system robust if there is no single point of failure (e.g., Wood and Wollenberg 1996, pp. 410–430). Fault-tree analysis, as used in transportation systems, power plants, and other critical systems (Roberts et al. 1981), typically identifies minimal sets of events, or "cut-sets," that are most likely to disrupt a system, and pronounces the system robust if the combined probability of occurrence is sufficiently low.

However, infrastructure that resists single points of random failure—these are single-element cutsets—or whose cutsets have low occurrence probabilities, may not survive an intelligently malicious attack. Random component failures offer a poor paradigm in a world with intelligent adversaries.

Vulnerability analysis must consider our adversary's ability to collect information about our infrastructure and use this information to identify weak points. A captured al Qaeda training manual advises: "Using [public sources] openly and without resorting to illegal means, it is possible to gather at least 80% of information about the enemy" (Federation of American Scientists 2006, p. UK/BM 80). In fact, we find that public sources often provide 100 percent of the information required to plan a devastating attack on an infrastructure system.

Al Qaeda also teaches the "overthrow of godless regimes [by] gathering information about the enemy, the land, the installations, and the neighbors . . . blasting and destroying the places of amusement, . . . embassies, . . . vital economic centers, . . . bridges leading

into and out of cities, . . . " (Federation of American Scientists 2006, p. UK/BM 12). Al Qaeda may not possess perfect models of our infrastructure, but its operatives are instructed to gather relevant information. That information can then be used to plan the most damaging attacks it can implement. Consequently, prudence dictates that we assume (1) that al Qaeda, or any other terrorist organization, will use its limited offensive resources to maximize damage to the infrastructure it decides to attack; and (2) that the terrorist organization has all the information necessary to accomplish its mission.

How would the military assess vulnerability when faced with an intelligent enemy? First, it would assume that our infrastructure will be attacked and would take steps to protect it, i.e., harden the infrastructure or improve its active defenses. The budget for this purpose will always be limited, but often not pre-specified. The military typically draws up a prioritized list of "defended assets" in need of protection, along with a list of potential protective measures, and presents these to policy makers. The latter parties make the final decisions after balancing costs, effectiveness, and intangibles, and after determining the budget. The United States Army (Department of the Army 2002a, b) applies four doctrinal components to evaluate and prioritize its defended assets (as well as those of its enemies): *criticality* (how essential is the asset?), *vulnerability* (how susceptible is the asset to surveillance or attack?), *reconstitutability* (how hard will it be to recover from inflicted damage?), and *threat* (how probable is an attack on this asset?).

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.

There are other differences that distinguish military and civilian infrastructure vulnerability. Military infrastructure is usually "hard" and well protected, while most civilian infrastructure in the United States is "soft," i.e., open to surveillance and attack, from an enemy that could be anywhere. Military planners assess probabilities of winning and losing, while civilians assume that they will eventually recover from an attack, no matter how damaging. Military planners also have extensive experience in assessing the likelihood that an enemy will choose a particular plan of attack ("course of action"). As civilian security planners, we are new to such analysis; we must learn to plan for what is possible, rather than what subjective assessments indicate is likely. We need a better method to assess the vulnerability of civilian infrastructure. Worst-case analysis is critical.

We apply *attacker-defender models*, and other related bilevel and trilevel optimization models, to these problems. These models do not normally attempt to measure directly the importance or value of an individual system component, i.e., "asset." They model a complete infrastructure system and its value to society, including how losses of the system's assets reduce that value, or how improvements in the system mitigate lost value. The exact meaning of value will depend on the system being modeled. It may mean economic output, production of a commodity, or time to detection of a toxic substance. Furthermore, (operating) cost, the converse of value, will often be a more convenient measure of how well a system functions. (The attacker-defender model is often called an "interdiction model" in the literature, e.g., Golden 1978, Wood 1993.)

An attacker-defender model does address criticality, vulnerability, reconstitutability, and threat, but in a very different way than military planners might. We include reconstitutability, when appropriate, by representing the repair of damaged assets over time, and how repaired assets contribute to improved system value (Salmerón et al. 2004b). We assume that each system component is vulnerable to attack unless it is specifically hardened or defended. We address "threat" by positing different levels of offensive resources for the terrorists. At the end of an analysis, we can determine the criticality of a group of assets, i.e., the value of protecting or hardening a given set of assets. We can also determine the value of adding redundant assets to improve the system's robustness.

In essence, an *attacker-defender model* becomes a submodel in a formal model or informal procedure for identifying a near-optimal, budget-limited defense plan. The formal model is a *defender-attacker-defender model*. However, a simpler *defender-attacker model* may suffice for this purpose if the contribution of a single asset to system performance is easy to define. We cover each of these three model types in this paper.

We present the basic models in the next three sections. If the mathematics is not of interest, the reader may skim those sections and continue with the "Three Examples" and "Supply Chains and Other Systems" sections to learn what we have discovered and how we generalize our findings. Brown et al. (2005a) provide additional examples.

Attacker-Defender Models

The core of an attacker-defender model is an optimization model of an infrastructure system whose objective represents the system's value or cost to the defender, i.e., our society, while it operates. For instance, the maximum throughput of a pipeline network could measure that system's value, while power-generation costs, plus economic losses resulting from unmet demand, could measure the cost of operating an electric power grid. We use cost rather than value in the following model.

We assume that the defender operates a system to minimize cost, which is represented by a linear function. The defender's problem is

$$\text{(D)} \quad \min_{\mathbf{y}\in Y} \mathbf{cy}, \tag{1}$$

where $\mathbf{c}$ defines a vector of component operating costs (and/or penalties), $\mathbf{y}$ represents system operating decisions or activities, and $\mathbf{y} \in Y$ represents constraints on that operation and the requirements to be met. Of course, by appropriately defining variables and constraints, we can also represent or approximate certain nonlinear cost functions in this model.

We note that "defender" is actually a misnomer in these models because the models do not directly represent defensive actions. "System user" or "system operator" would be more accurate, but awkward.

The model posits that an attacker wishes to maximize the defender's optimal (minimum) operating cost, and will do so by restricting the defender's activities $\mathbf{y}$. Let $x_k = 1$ if the attacker attacks the defender's kth asset, let $x_k = 0$ otherwise, and let $\mathbf{x}$ denote the vector of attack decisions, i.e., an *attack plan*. For simplicity, we assume that if $x_k = 1$, asset k is disabled and $y_j = 0$ for any activity j that requires that asset. That is, attack of an asset stops the defender from carrying on activities that depend directly on that asset.

Binary restrictions on $\mathbf{x}$, and some reasonable set of constraints on the attacker's resources, are represented by $\mathbf{x} \in X$. Let $Y(\mathbf{x})$ represent the defender's set of feasible activities, restricted by the attack plan $\mathbf{x}$. Thus, the attacker solves this planning problem:

$$\text{(AD)} \quad \max_{\mathbf{x}\in X} \min_{\mathbf{y}\in Y(\mathbf{x})} \mathbf{cy}. \tag{2}$$

AD is a type of *bilevel program* (e.g., Moore and Bard 1990), and a bilevel program is a type of *Stackelberg game* (von Stackelberg 1952). The terms *leader* and *follower* in a Stackelberg game represent our attacker and defender, respectively. The key assumptions that make a Stackelberg game appropriate here are (1) the attacker's and defender's actions are sequential, (2) the attacker has a perfect model of how the defender will (or should) optimally operate the system, even after an attack, and (3) the attacker will manipulate that system to his best advantage. The latter two assumptions are strong but prudent for us: The defender can suffer no worse should the attacker possess a less-than-perfect model of the defender's system, or fail to implement a perfect attack plan. A defensive plan that hardens or protects the defender's activities will be prudently conservative if AD is used to evaluate the plan's effectiveness.

One can devise many generalizations of AD, including attacks that increase costs rather than limit activities, or attacks that reduce the capacity of an asset by less than 100 percent. We will cover some of these generalizations after establishing basic results.

Naturally, the defender may also lack perfect knowledge of the attacker's capabilities. That is, the defender may be guessing at the attack-resource constraints representing part of X. In this case, the defender will need to solve the model over a range of attack-resource levels, and use these results, along with some common sense, to determine system improvements.

For many situations, a linear program (LP) will provide an adequate model of the defender's system and its operations. For instance, the electric power industry commonly employs linearized "optimal power-flow models" for security analysis (Wood and Wollenberg 1996, p. 419). Therefore, we can express the optimal operation of the defender's system as

$$\text{(D0)} \quad \min_{\mathbf{y}\geq\mathbf{0}} \ \mathbf{cy} \tag{3}$$

$$\text{s.t.} \quad A\mathbf{y} = \mathbf{b}, \tag{4}$$

$$F\mathbf{y} \leq \mathbf{u}. \tag{5}$$

Constraints (4) correspond to general system-operations constraints (e.g., balance of current at junctions in an electric power network), and constraints (5) represent capacity limitations for asset $k \in K$ (e.g., maximum capacity, in megawatts, of the kth power line). Assets can include power lines, pipelines, roads, ports, communications hubs, and so forth.

We assume that an attack on asset k causes the loss of all its capacity u_k. Thus, the full AD model is

$$\text{(AD0)} \quad \max_{\mathbf{x}\in X} \min_{\mathbf{y}\geq\mathbf{0}} \ \mathbf{cy} \tag{6}$$

$$\text{s.t.} \quad A\mathbf{y} = \mathbf{b}, \tag{7}$$

$$F\mathbf{y} \leq U(\mathbf{1}-\mathbf{x}), \tag{8}$$

where $U = \text{diag}(\mathbf{u})$. The inner LP must be constructed to be feasible for any $\mathbf{x}$ because we expect the system to operate in some degraded fashion after any conceivable attack. This may require the use of invulnerable activities, i.e., extra variables y_j that do not appear in constraints. Also, if some amount of capacity $\mathbf{u}_0$ is invulnerable to attack, constraints (8) become $F\mathbf{y} \leq \mathbf{u}_0 + U(\mathbf{1}-\mathbf{x})$.

A natural approach to solving AD0 begins by reformulating it: Fix $\mathbf{x}$ temporarily; take the dual of the inner linear program; and then release $\mathbf{x}$ (Wood 1993). Unfortunately, this yields an unappealing, nonlinear, mixed-integer program. That model can be linearized, but there is a simpler method: Change the paradigm of "capacity attack" to "cost attack," and then take the dual of the inner problem (Cormican et al. 1998). Specifically, let $-\mathbf{p}$ strictly bound the dual variables associated with $F\mathbf{y} \leq U(\mathbf{1}-\mathbf{x})$ over all possible values of $\mathbf{x} \in X$. Thus, p_k bounds the value to the defender of a unit of asset k's capacity. Because AD0 is feasible even when asset k has been disabled and has no capacity, it must be possible to penalize use of that capacity to make any use "uneconomical": p_k is such a penalty. AD0 is thus equivalent to

$$\text{(AD1)} \quad \max_{\mathbf{x}\in X} \min_{\mathbf{y}\geq\mathbf{0}} \ (\mathbf{c}+\mathbf{x}^T PF)\mathbf{y} \quad \text{[dual variables]}$$

$$\text{s.t.} \quad A\mathbf{y} = \mathbf{b} \quad [\boldsymbol{\theta}],$$

$$F\mathbf{y} \leq \mathbf{u} \quad [\boldsymbol{\beta}],$$

where $P = \text{diag}(\mathbf{p})$, and "dual variables" denotes dual variables for the inner LP given fixed $\mathbf{x}$. (Note that nonstrict bounds $\mathbf{p}$ are actually valid for optimizing $\mathbf{x}$; see Cormican et al. 1998.)

After taking the dual of the inner minimization, a mixed-integer linear program (MILP) results:

$$\text{(AD1-MILP)} \quad \max_{\boldsymbol{\beta}\leq\mathbf{0},\,\boldsymbol{\theta},\,\mathbf{x}} \ \mathbf{b}^T\boldsymbol{\theta} + \mathbf{u}\boldsymbol{\beta}$$

$$\text{s.t.} \quad A^T\boldsymbol{\theta} + F^T\boldsymbol{\beta} - F^T P\mathbf{x} \leq \mathbf{c},$$

$$\mathbf{x} \in X.$$

See Errata, Note 1

We can solve this model directly or by using Benders' decomposition (Benders 1962). The standard Benders method for integer $\mathbf{x}$ begins by taking the dual of AD1-MILP with $\mathbf{x}$ fixed, which causes AD1 to reappear. Thus, the Benders decomposition applies naturally to these problems.

To illustrate, consider the following simplified model of a crude-oil pipeline network:

Data

- A node-arc incidence matrix for the network.
- $\mathbf{b}$ vector of supplies and demands: $b_i > 0$ defines a supply of b_i million barrels per day (mmbbl/day) at node i, $b_i < 0$ defines a demand of b_i mmbbl/day at i, and $b_i = 0$ implies that i is a transshipment node (pumping station), assumed invulnerable to attack.
- $\mathbf{c}_1$ vector of shipping costs by arc, i.e., pipeline segment ($/mmbbl/day).
- $\mathbf{c}_2$ vector of penalties for not taking available supply ("take-or-pay penalties") ($/mmbbl/day).
- $\mathbf{c}_3$ vector of penalties for unmet demand (e.g., spot-market cost) ($/mmbbl/day).
- $\hat{I}_2$ incomplete diagonal matrix with a one for each supply node, but with zeroes elsewhere.
- $\hat{I}_3$ incomplete diagonal matrix with a one for each demand node, but zeroes elsewhere.

Variables

$\mathbf{y}_1$ flows on pipelines (mmbbl/day).
$\mathbf{y}_2$ unused supply (mmbbl/day).
$\mathbf{y}_3$ unmet demand (mmbbl/day).

Formulation

$$(\mathrm{D0_P}) \quad \min_{\mathbf{y}\geq\mathbf{0}} \ \mathbf{c}_1\mathbf{y}_1 + \mathbf{c}_2\mathbf{y}_2 + \mathbf{c}_3\mathbf{y}_3 \tag{9}$$

$$\text{s.t.} \ \ A\mathbf{y}_1 - \hat{I}_2\mathbf{y}_2 + \hat{I}_3\mathbf{y}_3 = \mathbf{b}, \tag{10}$$

$$I\mathbf{y}_1 \leq \mathbf{u}. \tag{11}$$

Constraints (10) are "elastic flow-balance constraints" that allow unused supply and unmet demand; constraints (11) represent pipeline capacities. For simplicity, we will (1) ignore the oil's purchase price, (2) assume that $\mathbf{c}_2 = \mathbf{0}$, $\mathbf{c}_1 > \mathbf{0}$, and $\mathbf{c}_3 = (c_3, c_3, \ldots, c_3)$, and (3) assume that only pipeline segments can be attacked.

Now, we proceed directly to create a "cost-attack" variant of the attacker-defender model in the form of AD1. Let $\mathbf{x}$ be defined as in AD1, with "asset k" now meaning "pipeline segment k." We suppose that intelligence reports indicate that terrorists can form at most T squads to carry out a coordinated attack, so that

$$\mathbf{x} \in X \equiv \left\{ \mathbf{x} \in \{0,1\}^{|K|} \,\middle|\, \sum_{k\in K} x_k \leq T \right\}.$$

We further note that $p \equiv c_3$ exceeds the penalty incurred by not supplying one mmbbl/day because $\mathbf{c}_1 > \mathbf{0}$. Thus, letting $\mathbf{p} = (p, p, \ldots, p)$ and $P = \mathrm{diag}(\mathbf{p})$, the max-min attacker-defender model becomes

$$(\mathrm{AD1_P}) \quad \max_{\mathbf{x}\in X} \min_{\mathbf{y}\geq\mathbf{0}} \ (\mathbf{c}_1 + \mathbf{x}^T P)\mathbf{y}_1 + \mathbf{c}_2\mathbf{y}_2 + \mathbf{c}_3\mathbf{y}_3 \tag{12}$$

$$\text{s.t.} \ \ A\mathbf{y}_1 - \hat{I}_2\mathbf{y}_2 + \hat{I}_3\mathbf{y}_3 = \mathbf{b}, \tag{13}$$

$$I\mathbf{y}_1 \leq \mathbf{u}. \tag{14}$$

We leave it to the reader to take the dual of the inner minimization to create $\mathrm{AD1_P}$-MILP. However, there is a caveat: The quality of the LP relaxation of that MILP will depend directly on how small the penalties p_k are. Therefore, the modeler may need to work to identify small, valid values. For instance, for any $\varepsilon > 0$, each p_k in $\mathrm{AD1_P}$ can be validly reduced to $p_k - c_{1,\min} + \varepsilon$, where $c_{1,\min}$ is the smallest shipping cost a demand might incur while being satisfied.

Actually, a cost-attack model like AD1 will sometimes apply directly to infrastructure analysis. For instance, suppose that D0, with constraints (5) eliminated, corresponds to a minimum-traversal-time (shortest-path) problem in a road network. Rather than having an attack on link k reduce that link's capacity, a more natural model may simply add a delay d_k to the nominal traversal time c_k. Thus, this model becomes

$$(\mathrm{AD1_R}) \quad \max_{\mathbf{x}\in X} \min_{\mathbf{y}\geq\mathbf{0}} \ (\mathbf{c} + \mathbf{x}^T D)\mathbf{y}$$
$$\text{s.t.} \ \ A\mathbf{y} = \mathbf{b}$$

(Israeli and Wood 2002), where $D = \mathrm{diag}(\mathbf{d})$, with $\mathbf{d}$ being the vector of delays d_k. (Hereafter, we will not announce the bold, vector versions of variables and data, except when used to define matrices such as D.)

Defender-Attacker Models

The solution of an attacker-defender model identifies a set of most-critical assets (components) for a system. The ability to identify such assets leads to some obvious heuristics for approximating the solution to the "optimal defense problem," i.e., for identifying a near-optimal defense plan, given a limited defense budget. But, how do we identify truly optimal solutions?

In theory, one merely embeds the bilevel attacker-defender model in a trilevel defender-attacker-defender model (DAD) such as

$$(\mathrm{DAD}) \quad \min_{\mathbf{w}\in W} \max_{\mathbf{x}\in X(\mathbf{w})} \min_{\mathbf{y}\in Y(\mathbf{x})} \mathbf{c}\mathbf{y}. \tag{15}$$

Here, $\mathbf{w}$ denotes a binary vector of defensive decisions (e.g., $w_k = 1$ if asset k is hardened and made invulnerable, and $w_k = 0$, otherwise), $\mathbf{w} \in W$ denotes the binary restrictions on $\mathbf{w}$ together with budgetary and other possible constraints, and the inner max-min problem simply represents an attacker-defender model with a restricted set of attack strategies, $X(\mathbf{w})$. Thus, the defender wishes to identify a defense plan $\mathbf{w}^*$ so that when the attacker solves

$$\max_{\mathbf{x}\in X(\mathbf{w}^*)} \min_{\mathbf{y}\in Y(\mathbf{x})} \tag{16}$$

the "benefit" the attacker perceives, i.e., the worst damage the attacker can inflict, is minimized.

In general, we believe that DADs will solve only with difficulty because conversion to a monolithic MILP will usually be impossible, necessitating more complicated decomposition techniques. We discuss this topic later in the paper. Fortunately, certain optimal-defense problems lend themselves to easier bilevel, defender-attacker models of the following form:

Indices

k asset the defender may want to defend, and the attacker may want to attack (we use a one-to-one relationship here for simplicity).

Data

c_k value to the attacker of attacking undefended asset k.

p_k reduction in value of attacking asset k if that asset is defended, i.e., the attacker receives benefit $c_k + p_k$, $p_k \le 0$, by attacking defended asset k.

Variables

$$x_k = \begin{cases} 1 & \text{if the defender defends his } k\text{th asset,} \\ 0 & \text{otherwise.} \end{cases}$$

$$y_k = \begin{cases} 1 & \text{if the defender's } k\text{th asset is attacked,} \\ 0 & \text{otherwise.} \end{cases}$$

Constraints

$\mathbf{x} \in X$ resource constraints and binary restrictions on the defender's defense plan, e.g., $X = \{\mathbf{x} \in \{0,1\}^n \mid G\mathbf{x} \le \mathbf{f}\}$.

$\mathbf{y} \in Y$ resource constraints and binary restrictions on the attacker's attack plan, e.g., $Y = \{\mathbf{y} \in \{0,1\}^n \mid A\mathbf{y} = \mathbf{b}\}$.

Formulation

$$\text{(DA1)} \quad \min_{\mathbf{x}\in X} \max_{\mathbf{y}\in Y} (\mathbf{c} + \mathbf{x}^T P)\mathbf{y}.$$

A simplified example illustrates this model. Suppose that intelligence reports indicate that a terrorist organization, "the attacker," intends to send out b teams to attack b different subway stations in a city having $M > b$ total stations. Municipal authorities, "the defender," have $m < M$ teams with which to defend stations. The value to the defender of station k is $c_k > 0$, and we assume that the attacker assigns the same values. Let $p_k = -c_k$. Thus, a defended station becomes invulnerable, and the attacker gains no benefit by attacking it. We formulate this "subway-defense problem" as

$$(\text{DA1}_{\text{SUB}}) \quad \min_{\mathbf{x}\in X} \max_{\mathbf{y}\in\{0,1\}^M} \sum_{k=1}^{M} (c_k + x_k p_k) y_k \tag{17}$$

$$\text{s.t.} \sum_{k=1}^{M} y_k = b, \tag{18}$$

where $X = \{\mathbf{x} \in \{0,1\}^M \mid \sum_{k=1}^{M} x_k = m\}$.

In general, DA1 and instances like DA1_{SUB} are difficult to solve because the inner maximization is not an LP. Thus, no general transformation exists to convert DA1 into an MILP as we converted AD1 into AD1-MILP. This can be resolved in one of three ways:

Case 1. We decide that continuous attack effort represents a reasonable approximation of reality; therefore, we convert Y to $Y_{\text{CONT}} = \{\mathbf{y} \in R_+^n \mid A\mathbf{y} = \mathbf{b},\ \mathbf{y} \le \mathbf{1}\}$ (Golden 1978).

Case 2. The LP relaxation of Y, $Y_{\text{LP}} = \{\mathbf{y} \in R_+^n \mid A\mathbf{y} = \mathbf{b},\ \mathbf{y} \le \mathbf{1}\}$, yields intrinsically binary solutions, making a conversion from DA1 into DA1-MILP possible. Such is the situation with DA1_{SUB}, and we invite the reader to work out the details. Typically, Case 2 will arise when Y_{LP} corresponds to a network-flow problem which, having a totally unimodular constraint matrix, possesses integer extreme points (e.g., Ahuja et al. 1993, pp. 447–449). Indeed, Y_{LP} for DA1_{SUB} describes a simple network flow problem. Brown et al. (2005b) present a more complex instance involving theater ballistic missile defense.

Case 3. Neither of the cases above pertains, and we must include restriction $\mathbf{y} \in \{0,1\}^n$ in the definition of Y.

Case 3 requires special techniques to solve, but solution methods better than brute-force enumeration do exist (e.g., Israeli and Wood 2002, Skroch 2005). This paper focuses on Cases 2 and 3 because Case 1 seems unrealistic for our applications.

We offer one final observation on the DA model. DA cannot incorporate a detailed operational model of the defender's system. However, by manipulating $\mathbf{x} \in X$, we can describe limited operational detail. For instance, suppose that the defender's system loses value $c > 0$ if either asset k or k' is attacked (when undefended), but loses no additional value if both

k and k' are attacked. The constraint $x_k + x_{k'} \leq 1$ handles this situation perfectly when added to the constraints defining X.

Defender-Attacker-Defender Models

Although difficult, we can sometimes solve a trilevel DAD model exactly, to prescribe an optimal defensive plan for an infrastructure system. The DAD must assume a fixed level of offensive resources, but results will be believable if we make appropriately conservative assumptions. For instance, can we really believe that a group of terrorists will be able to strike more than 10 electric power substations simultaneously in a particular region? Limiting the number of attacks to 10 may be deemed appropriately conservative.

For simplicity, we assume that if asset k is defended, i.e., $w_k = 1$, then that asset becomes invulnerable. We let $h^+ \equiv \max\{0, h\}$ apply componentwise in a vector, so that $(\mathbf{x} - \mathbf{w})^+$ denotes the "net attack plan" that results from attack plan $\mathbf{x}$ implemented against defense plan $\mathbf{w}$. Using AD0 as the inner, bilevel model, the trilevel model becomes

$$\begin{aligned}(\text{DAD0}) \quad z_D^* = \min_{\mathbf{w}\in W} \max_{\mathbf{x}\in X} \min_{\mathbf{y}\in Y} \mathbf{cy} \\ \text{s.t.} \quad A\mathbf{y} = \mathbf{b}, \\ \mathbf{0} \leq \mathbf{y} \leq U(\mathbf{1} - (\mathbf{x} - \mathbf{w})^+).\end{aligned}$$

We warned about taking the dual of the inner minimization before, but now we have

$$z_D^* = \min_{\mathbf{w}\in W} \max_{\mathbf{x}\in X} \max_{\boldsymbol{\alpha},\boldsymbol{\beta}} \boldsymbol{\alpha}\mathbf{b}^T + \boldsymbol{\beta}U(\mathbf{1} - (\mathbf{x} - \mathbf{w})^+) \quad (19)$$

$$\text{s.t.} \quad \boldsymbol{\alpha}A + \boldsymbol{\beta}I \leq \mathbf{c}, \quad (20)$$

$$\boldsymbol{\beta} \leq \mathbf{0}, \quad (21)$$

$$= \min_{\mathbf{w}\in W,\, z} \quad z \quad (22)$$

$$\text{s.t.} \quad z \geq \hat{\boldsymbol{\alpha}}_l \mathbf{b}^T + \hat{\boldsymbol{\beta}}_l U(\mathbf{1} - (\hat{\mathbf{x}}_l - \mathbf{w})^+), \quad l \in L, \quad (23)$$

where L enumerates all combinations of maximal attack plans $\hat{\mathbf{x}} \in X$ and extreme points $(\hat{\boldsymbol{\alpha}}, \hat{\boldsymbol{\beta}})$ from (20) and (21).

The final formulation indicates that DAD0 can be solved just as we might solve a DA with a Benders decomposition, except: (1) the subproblems will be instances of AD solved via AD1, and (2) the master problem will require constructs to handle the "+" operator. The fact that the subproblems can be solved by decomposition leads to interesting possibilities for a "nested decomposition" (O'Neill 1976).

We have only just begun to explore DADs, and a host of alternative or complementary solution techniques must be tested. One technique has already proven useful—the addition of "super-valid inequalities" (Israeli and Wood 2002) to the relaxed master problem, i.e., the version of the master problem (22)–(23), that is solved during the Benders decomposition algorithm. In particular, as the algorithm generates each Benders cut (23) based on a new solution $\widehat{\mathbf{w}}_l$, we also add a constraint that represents $\mathbf{w} \neq \widehat{\mathbf{w}}_l$. The upper bound from the relaxed master problem remains valid if we have not identified an optimal solution; if we have identified such a solution, the value of the bound is irrelevant. Because $\mathbf{w}$ is binary, simple linear constraints will implement the super-valid inequalities.

Actually, one can implement a version of Benders' decomposition with a master problem whose constraints consist only of super-valid inequalities, and with an objective function that represents any of the lower-bounding functions in (23). Brown (2005) applies this technique to a model for planning the reconstruction of the Iraqi oil pipeline system and defending it from insurgents.

Three Examples

This section describes AD, DA, and DAD models applied to problems of protecting specific instances of critical infrastructure. We have created and tested many of these models by (1) defining a hypothetical but realistic scenario, (2) assembling a "red team" of well-trained, military officer-students to gather data from strictly public sources, (3) advising the team on creating and solving an appropriate model, and (4) helping analyze results.

The results have led to valuable insights. We have found cases in which a given set of attackers can do more—or less—damage than we would have predicted, and sometimes the attacks do not target the "obvious" components revealed in single-point-of-failure analysis.

The Strategic Petroleum Reserve: Attacker-Defender

We first consider the US Strategic Petroleum Reserve (SPR), which stores 700 million barrels of crude oil in underground caverns, and which can deliver this oil (about two months' supply for the United States) via its pumps and pipelines to refiners, ports, and export pipelines. Terrorists have certainly planned attacks on infrastructure like the SPR elsewhere in the world (Luft and Korin 2003).

We seek a defensive plan for a section of the SPR that lies in Louisiana. (We base Figure 1 on a series of telephone and e-mail discussions during April and May 2005 with J. Holbrook and P. Withers, analysts for the Space Countermeasures Hands On Program at Kirtland Air Force Base, Albuquerque, NM.) Figure 1 depicts that section as a network, showing (1) the Bayou Choctaw and West Hackberry storage sites as source nodes, (2) four refineries, four ports, and 14 export pipelines as sink nodes, (3) a number of pumping stations and junctions as transshipment nodes, and (4) a number of pipeline sections connecting the nodes as network arcs.

We suppose that the United States is in a state of emergency and that the defender requires maximum output from the SPR and consequently measures the "cost" of operating the system in terms of any reduction below that maximum. We could create a formal DAD as the basis for analysis—the network is small and the corresponding DAD would solve easily. However, we imagine that analysts have just begun their work, and prefer to explore a set of discrete options to "get a feel for the problem." So, for this limited scenario, the analysts' toolkit consists of the attacker-defender model $AD1_P$, (Equations (12)–(14)).

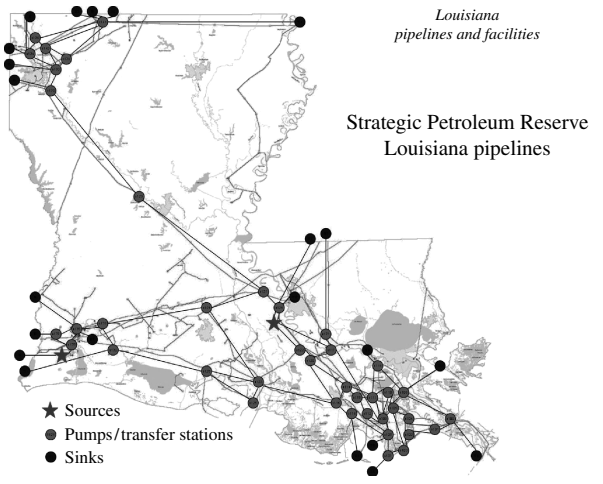

Figure 1: The US Strategic Petroleum Reserve has two storage locations in Louisiana connected by a system of pumps and pipelines to refiners, ports, and export pipelines. We model defense of the maximum system output. Several simple defense plans make the system highly robust against multiple attacks (Benedetto et al. 2005).

Analysts working for the SPR would have precise data for pipeline capacities and pumping rates, but we believe that our estimates, derived from public sources, should suffice for purposes of demonstration. They should also suffice for purposes of a terrorist organization. Now, for each of three defensive options, we evaluate optimal attack plans assuming that the attacker can destroy no network components (nodes or arcs), one component, two components, and so forth. The options and results follow.

Defense Option A: *Baseline, no defense*. The destruction of only two system components, the sources, reduces optimum system output to zero (i.e., leads to the most costly system operation possible). Thus, a sensible defensive plan must include the sources.

Defense Option B: *Protect critical core components*. We discover a "critical backbone" of components, which, if protected, ensures connection of the two sources to many, normally redundant, parts of the distribution network. With the backbone defended, at least seven (undefended) components must be destroyed before maximum system output drops below half.

Defense Option C: *Protect a* 10*-mile-radius security zone around each source*. This protects three-quarters of the system capacity for any conceivable number of attacks.

Border Patrol: Defender-Attacker

The porosity of the United States' borders has received much attention in recent years, with emphasis placed on the lack of border-patrol resources (e.g., General Accounting Office 2004). We believe that operations research can help make better use of limited budgetary and human resources here. In particular, we want to improve the probability that border defenses will detect an alien, who may be a terrorist, trying to infiltrate the country from Mexico. For simplicity, we assume a single "infiltrator" will choose

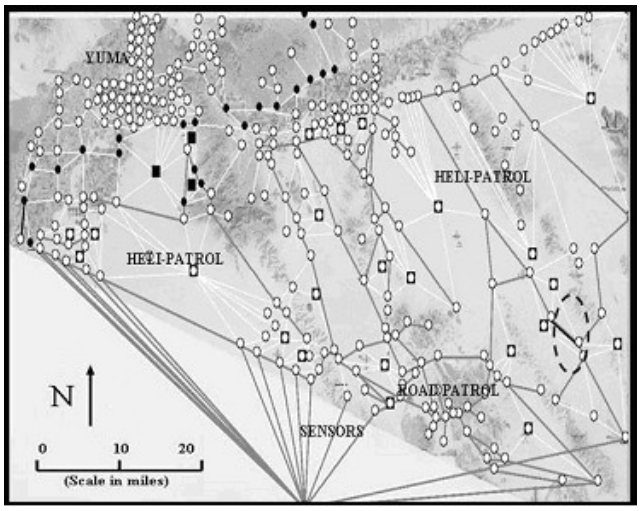

Figure 2: Limited patrol assets can be allocated optimally to detect illegal incursions into the United States through the Yuma, Arizona border area. A map of the area is overlaid with a skeleton of a network that represents potential infiltration routes from Mexico. Arcs not shown represent movements from border entry points through sensor fields, and along roads and footpaths. The dark nodes identify a route that minimizes the maximum probability of detection for an infiltrator (Pulat 2005).

from among a set of well-known routes to attempt to enter the United States.

Figure 2 shows a map of the Yuma border area, along with a skeleton of the "infiltration network" that describes the paths an infiltrator could take from Mexico, into the United States through conventional portals, or via incursions over roads or footpaths. (The full network contains too many arcs to depict.) The intent is to spend a limited security budget on procedural changes, sensors, road patrols, and helicopter patrols to increase detection probabilities on individual arcs and thereby maximize overall detection probability. The options for procedural changes include closing off certain legal portals, and using sensors or helicopter patrols for detection and cueing ground units. Ground units are vehicles and crews that we position independently, or position to follow up on cues from helicopter patrols.

Probability of nondetection proves to be a useful concept for modeling this problem. For simplicity, we assume that every arc k in the network possesses a nominal probability $1 \geq q_k > 0$: This is the current probability of nondetection if the infiltrator traverses arc k. If we spend c_k dollars at arc k, a new sensor will be installed, or a new procedure implemented, and the nondetection probability becomes $\bar{q}_k > 0$, with $\bar{q}_k < q_k$. (Note that (1) The model extends easily to handle multiple options for reducing nondetection probability on an arc, (2) completely closing off an arc k can be handled by setting $\bar{q}_k$ arbitrarily close to zero, (3) an artificial arc k connects each entry point to an artificial source node s, with $q_k = \bar{q}_k = 1$, and, similarly (4) an artificial arc k connects each node representing a completed infiltration to an artificial sink node t, with $q_k = \bar{q}_k = 1$.)

We seek to spend a budget of c' dollars to minimize the maximum probability of nondetection along any path the infiltrator might take. If we assume independence of detection events, this model can be formulated as follows (see the related model in Pan et al. 2003):

Indices and Structural Data

$i \in \mathcal{N}$ nodes of the infiltration network.
$k \in \mathcal{A}$ directed arcs of the infiltration network.
$\mathcal{G} = (\mathcal{N}, \mathcal{A})$ infiltration network.

Variables

$$x_k = \begin{cases} 1 & \text{if the defender upgrades security on arc } k, \\ 0 & \text{otherwise.} \end{cases}$$

$$y_k = \begin{cases} 1 & \text{if the attacker traverses arc } k \text{ when } x_k = 0, \\ 0 & \text{otherwise.} \end{cases}$$

$$\bar{y}_k = \begin{cases} 1 & \text{if the attacker traverses arc } k \text{ when } x_k = 1, \\ 0 & \text{otherwise.} \end{cases}$$

Data

A node-arc incidence matrix for $\mathcal{G}$.
$\mathbf{b}$ node-length vector with $b_s = 1$, $b_t = -1$, and $b_i = 0$ for all $i \in \mathcal{N} \backslash \{s, t\}$.
q_k nominal probability of nondetection on arc k when $x_k = 0$ ($q_k > 0$).
$\bar{q}_k$ probability of nondetection on arc k when $x_k = 1$ ($q_k > \bar{q}_k > 0$).
d_k $\log q_k$ (vector form $\mathbf{d}$ and $D \equiv \text{diag}(\mathbf{d})$).
$\bar{d}_k$ $\log \bar{q}_k$ (vector form $\bar{\mathbf{d}}$ and $\bar{D} \equiv \text{diag}(\bar{\mathbf{d}})$).
c_k cost to upgrade security on arc k ($).
c' total budget for upgrading security ($).

Formulation **See Errata, Note 2**

$$(\mathrm{DA1}_{\mathrm{YUMA}}) \quad \min_{\mathbf{x}\in X} \max_{\mathbf{y},\,\bar{\mathbf{y}}} \prod_{k\in\mathcal{A}} q_k^{(1-x_k)y_k} \bar{q}_k^{\,x_k\bar{y}_k} \tag{24}$$

$$\text{s.t.} \quad A\mathbf{y} + A\bar{\mathbf{y}} = \mathbf{b}, \tag{25}$$

$$\mathbf{y}, \bar{\mathbf{y}} \in \{0,1\}^{|\mathcal{A}|}, \tag{26}$$

where $X = \{\mathbf{x} \in \{0,1\}^{|\mathcal{A}|} \mid \mathbf{c}\mathbf{x} \le c'\}$.

Constraints (25) and (26) ensure that one unit of "unsplittable flow," representing the infiltrator, moves from s to t. Constraints (25) are standard flow-balance constraints, just like those that would model a shortest-path problem in $\mathcal{G}' = (\mathcal{N}, \mathcal{A} \cup \mathcal{A}')$, where $\mathcal{A}'$ duplicates $\mathcal{A}$.

We then apply a standard logarithmic transformation to the objective function to obtain this equivalent model:

$$(\mathrm{DA2}_{\mathrm{YUMA}}) \quad \min_{\mathbf{x}\in X} \max_{\mathbf{y},\,\bar{\mathbf{y}}\ge\mathbf{0}} \; (\mathbf{1}-\mathbf{x})^T D\mathbf{y} + \mathbf{x}^T \bar{D}\bar{\mathbf{y}}$$

$$\text{s.t.} \quad A\mathbf{y} + A\bar{\mathbf{y}} = \mathbf{b}.$$

Simple nonnegativity restrictions replace constraints (26), because the constraint matrix in (25) is totally unimodular. Indeed, for fixed $\mathbf{x}$, the model defines a shortest-path problem on $\mathcal{G}'$ if one multiplies D and $\bar{D}$ by -1, and switches the maximization to a minimization. This model converts easily to an MILP. (See Case 2 in the Defender-Attacker Models section.)

We use standard search-theory to estimate detection probabilities on arcs. Although absolute statistics are of questionable value, relative results are plausible. The results for four different resources scenarios are summarized below. Note that the results are specified in terms of probability of detection, not nondetection.

Baseline Scenario 1, *no security improvements*. An infiltrator would cross the border and traverse downtown Yuma, exiting the city to the northeast. Probability of detection = 0.04.

Scenario 2, *one check point, one remote observation post, two road patrols, sensors to cover at most* 15 *road segments, and one helicopter, all visible to the infiltrator*. Probability of detection = 0.07.

Scenario 3, *two check points, one remote observation post, two road patrols, sensors to cover at most* 15 *road segments, and one helicopter, all visible to the infiltrator*. Probability of detection = 0.11.

Scenario 4, *surprise interdiction of downtown Yuma infiltration route*. One hidden sensor field and two surprise roadblocks are located optimally. The probability of detection rises to 0.6 because information has been hidden from the infiltrator. This represents an interesting use of a Stackelberg game in which we fool the follower ("infiltrator" or "attacker") into playing one game but evaluate success according to another that is more advantageous to the leader ("defender"). This game will be played many times, however, and the infiltrator will eventually catch on to the ruse. Two-person zero-sum game theory may be needed here.

Electric Power Grids: Defender-Attacker-Defender
We have produced a complete decision-support system called the Vulnerability of Electric Grids Analyzer (VEGA) that uses an AD model to identify critical components in a power grid (Salmerón et al. 2004a, b; Brown et al. 2005a). Criticality of grid components is measured through "disruption," which may be viewed as the penalty for unserved demand, weighted by different customer sectors. (Disruption includes a small factor for actual generation costs, but we ignore that in this paper.) We assume that a group of terrorists, using limited offensive resources, will attack and destroy, i.e., "interdict," grid components to maximize disruption.

In VEGA, a set of standard "optimal DC power-flow submodels" (DCOPFs) comprise D0, the inner, minimizing LP (Wood and Wollenberg 1996, p. 514). Each submodel looks just like the pipeline model ($\mathrm{D0_P}$), constraints (9)–(11), except that (1) the network is an electrical grid instead of a pipeline network; (2) the commodity flowing through the network is electrical current instead of oil; and (3) the model adds linearized admittance constraints for AC lines. This LP approximates the "true," nonlinear AC model, but the industry deems it adequate for security analyses. In fact, an independent system operator may solve a model like this thousands of times per day to ensure that a power grid maintains "$N-1$ security," i.e., can still meet all customer demand after any single component failure. In our case, the submodels represent different system states as demand varies and repairs proceed, over time, after an attack.

Ultimately, we wish to identify the best, budget-limited set of protective measures for the power grid,

i.e., to solve an instance of DAD with VEGA's current model representing the "AD" part of "DAD." We have developed such a DAD model for VEGA, but cannot yet solve full-scale problems as we can for AD. Consequently, the example described below only covers a modest-size test system from the Institute of Electrical and Electronics Engineers (IEEE).

We make a number of assumptions to simplify the presentation: Only power lines can be interdicted, and thus only power lines need defending; all lines require the same amount of time to repair; and load (demand) remains constant. Thus, we concern ourselves only with the instantaneous unserved demand for power and solve only a single DCOPF to evaluate the inner model, D0.

Under the above assumptions, the following model describes a valid master problem for this DAD, taking the place of (22)–(23):

$$z^* = \min_{\mathbf{w}\in W,\, z} \; z \tag{27}$$

$$\text{s.t. } z \geq f(\hat{\mathbf{x}}_l) + \sum_{k \mid \hat{x}_{lk}=1} \hat{\beta}_{lk} u_k w_k, \quad l \in L, \tag{28}$$

where $f(\hat{\mathbf{x}}_l)$ evaluates the disruption caused by interdiction plan (attack plan) $\hat{\mathbf{x}}_l$, i.e., "load shedding" (unmet demand for electricity) or its cost; u_k denotes the capacity of line k; and $\hat{\beta}_{lk}$ is the optimal dual variable on the capacity constraint that must be enforced when $\hat{x}_{lk} = 1$, namely, $y_k \leq 0$.

The formulation (27)–(28) ignores the fact that an attack not only drops the capacity of a line to zero, but also eliminates one or more admittance constraints that relate phase angles of power flows on interconnected lines. Thus, a partial benefit may actually accrue to the system because of an attack. This means that when interdicted line k is retrospectively defended, i.e., the master problem sets $w_k = 1$ for some $\hat{x}_{lk} = 1$, so that $(\hat{x}_{lk} - w_k)^+ = 0$ (see DAD0), then we should account for the negative benefit accrued by re-enforcing one or more admittance constraints. However, we ignore this effect. The negative benefit could serve to reduce the coefficients $\hat{\beta}_{lk}$ in (28) and thereby strengthen the master problem. But, the master problem remains valid because each constraint in (28) defines a valid lower-bounding function on z^*, and the solution to the final master problem returns the true objective-function value for DAD0 for any explicitly evaluated solution $\widehat{\mathbf{w}}_l \in W$. (This is true because whenever the master problem returns $\widehat{\mathbf{w}}_l \in W$, we will immediately solve for a corresponding optimal interdiction plan $\hat{\mathbf{x}}_l$, with objective value $f(\hat{\mathbf{x}}_l)$, and add a Benders cut (28) to the master problem.)

A hypothetical grid known as "Reliability Test System with Two Areas" (Institute of Electrical and Electronics Engineers 1999) defines our test scenario (Figure 3). This grid comprises 48 buses (nodes), 69 power lines (which allow flow of electricity in both directions), and 10 high-voltage transformers in four substations and 66 generating units. (However, recall that, for simplicity, only power lines may be interdicted.) Also, Equation (26) is slightly modified to account for attacks on physically parallel lines (14 in our example). Specifically, if two lines are mounted on the same towers, an attack on one implies an attack on both. We allow $n = 4$ interdictions (attacker's resource), but assume that we can prevent interdiction on $m = 8$ lines (defender's resource).

Using our AD model, it is relatively easy to find an optimal interdiction plan on the undefended network, i.e., when $\widehat{\mathbf{w}} = \widehat{\mathbf{w}}_0 = \mathbf{0}$. The optimal lines to interdict are $X^*(\widehat{\mathbf{w}}_0) = \{\text{A18, A21, A23, A26}\}$, yielding a cost of \$915,023/hour based on a load-shedding penalty of \$1,000/MWh. Interestingly, the optimal defense plan for eight lines does not cover all four of the lines in $X^*(\widehat{\mathbf{w}}_0)$. In fact, $W^* = \{\text{A18, A23, A32-1, A33-1, B21, B23, B27, B28}\}$, which includes only two lines from the optimal, undefended interdiction plan. With this defense, denoted by the vector $\widehat{\mathbf{w}}^*$, the best interdiction plan becomes $X^*(\widehat{\mathbf{w}}^*) = \{\text{A11, A12-1, B11, B12-1}\}$, and the defended system now costs only \$421,028/hour. (Note that the attacker does not interdict the now-undefended lines {A21, A26}.)

Next, let us show that forcing a defense plan to cover the optimal, undefended interdiction plan $X^*(\widehat{\mathbf{w}}_0)$ would result in a substantial misuse of defensive resources. Such a defense might result from a planner using a natural, defensive rule of thumb: Completely defend against the worst-case interdiction plan, and use your remaining defensive resources as advantageously as possible. To simulate this rule of thumb, we fix variables in the DAD to defend {A18,

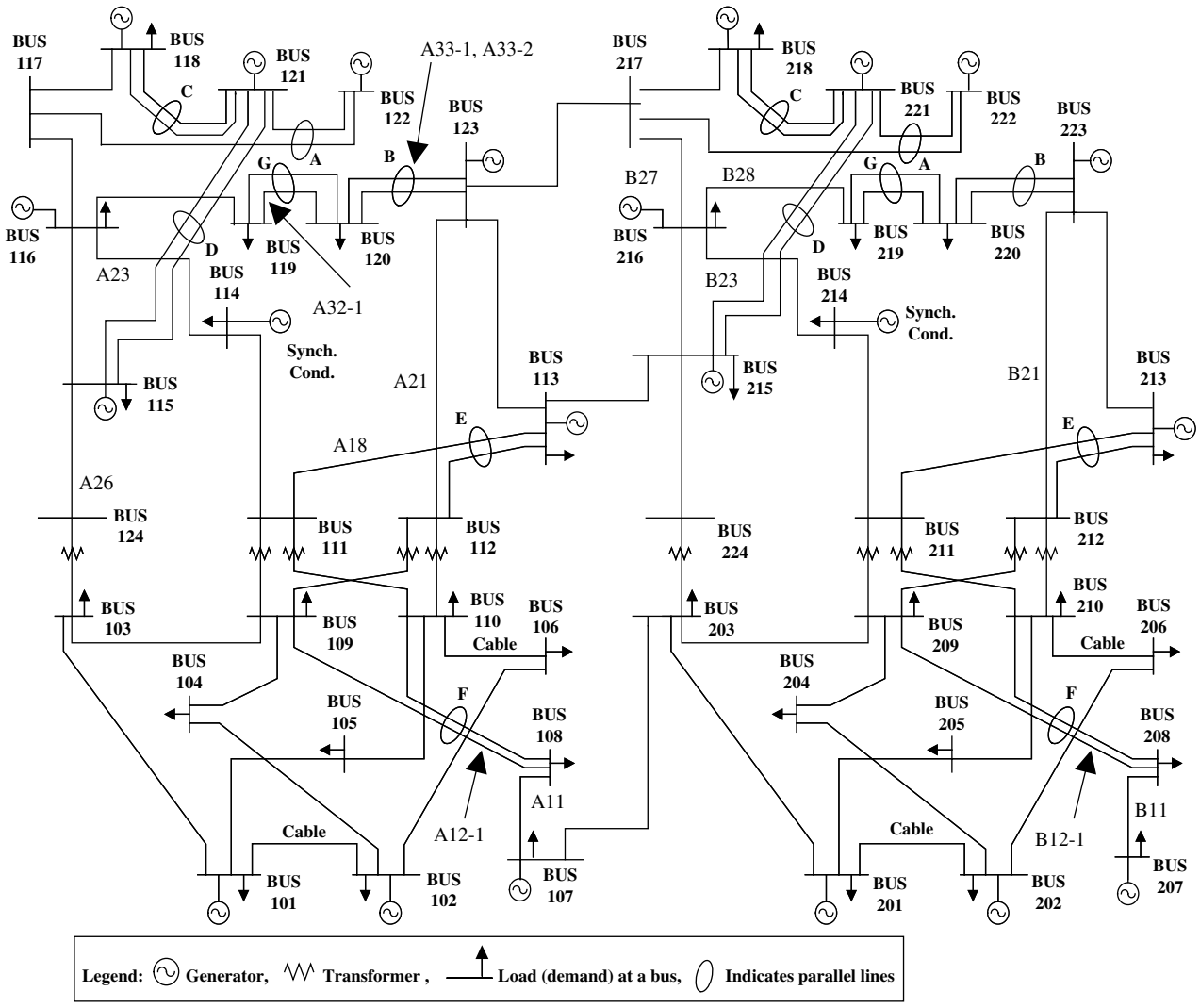

Figure 3: Reliability Test System with Two-Areas (after Institute of Electrical and Electronics Engineers 1999). This "one-line diagram" describes a hypothetical electric power transmission grid used here to illustrate optimal and suboptimal defensive plans evaluated through a formal, trilevel, defender-attacker-defender model.

A21, A23, A26}, and allow the model to select optimally the remaining four defended lines. The full, suboptimal defense plan becomes $W' = \{\text{A18, A21, A23, A26, A27, A28, B21, B28}\}$, also denoted by the vector $\widehat{\mathbf{w}}'$. The attacker counters $\widehat{\mathbf{w}}'$ by interdicting lines $X^*(\widehat{\mathbf{w}}') = \{\text{A12-1, B18, B23, B26}\}$, yielding a cost of \$538,192/hour—almost 28 percent higher than optimal. This percentage would likely be even higher in the real world: Presumably, a planner who suboptimally forces defense of $X^*(\widehat{\mathbf{w}}_0)$, will not optimally allocate his remaining defensive resources, either.

Supply Chains and Other Systems

Supply chains, i.e., physical-distribution systems, are a key infrastructure of companies that manufacture or distribute goods. Supply chains are critical to our nation's well-being despite their omission from the Department of Homeland Security (2002) list of critical infrastructure. For example, Wein and Liu (2005) describe how thousands of people could be killed by the introduction of botulinum toxin at various points in a milk production, transportation, and processing chain.

Strategic supply chain design for reducing costs and improving service levels has a long and successful record in the United States. Unfortunately, efficient supply chains are highly vulnerable to attack. In fact, after scrupulously investing exactly the right amount of money in a supply chain, on exactly the right bottlenecks, the resulting product-flow channels resemble one or more spanning trees. However, a spanning tree is maximally fragile: Breaking any link disconnects the network.

Brown et al. (2003a, b, 2004) and INSIGHT (2006) address supply chain vulnerability. Our most important "result" is an observation: We still encounter considerable confusion in the private sector between random acts of nature—these have been studied by insurance actuaries for centuries—and belligerent acts of an intelligent attacker who observes defensive preparations and acts to maximize damage. We strongly suggest remedying this confusion before proceeding with any analysis.

Sometimes, one can reduce vulnerability substantially with simple planning and with only a modest investment in new physical infrastructure: Strategically relocating surge capacity may provide benefit at virtually no cost. This contrasts with the high cost of adding redundant capacity, or hardening components, in other types of infrastructure.

We have learned to model competitors and dissatisfied labor unions as attackers because they seek to maximize damage inflicted (e.g., to market share, profit, or reputation). For instance, the labor dispute that resulted in denial of access to west coast ports in the United States in 2002 was no less damaging than the anthrax attacks of 2001 that closed eastern postal services.

We have presented our findings to numerous companies and have received enthusiastic responses. American companies now have senior executives focused on "corporate continuity." These positions were originally motivated by threats to information systems. Thus, back-up computer facilities and doubly backed-up data have become ubiquitous. Now, these same companies are realizing that they must also back up their physical operations to handle attacks on their own infrastructure (e.g., equipment, warehouses) as well as attacks on the public infrastructure they use (e.g., roads, communications networks).

Our work has also led to new military and diplomatic planning models; two have already been incorporated into comprehensive decision-support systems. One system helps plan theater ballistic-missile defense (Brown et al. 2005a). The embedded defender-attacker model optimally locates anti-missile platforms (ships and ground-based units supplied with antimissile missiles) while assuming the attacker can see some or all of our defensive preparations. The other system identifies optimal actions (e.g., embargoes of key materials, economic sanctions, military strikes) to delay a nuclear weapons program (Skroch 2005, Harney et al. 2006). In this attacker-defender model, we are the attacker. This model applies to any complex industrial project that can be delayed by a competitor.

One 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 an adversary's infrastructure. For instance, hiding the location of a defensive asset could cause an attacker to strike an essentially invulnerable target. When dealing with a suicide bomber, such an outcome could be desirable.

Even though this work is relatively new, there is already a large body of unclassified publications, including about 70 red-team case studies, over 20 graduate theses, and numerous journal papers from our research team and others. The topics include those discussed in this paper as well as rail networks, domestic water-distribution systems, sea routes, attacks on public events, and others. Furthermore, several decision-support tools have been built and are actively being extended.

What We Have Learned

The answers are not obvious. The most damaging coordinated attacks, and the most effective defenses, can be nonintuitive. The United States infrastructure is enormous and complex. Analysis of such a large infrastructure deserves rigorous, optimizing, decision-support tools to formalize the notion of a transparent, two-sided conflict.

High-fidelity models are achievable. We can formulate, find data for, and quickly solve high-fidelity

models of critical-infrastructure systems. Simpler, aggregated models may appeal, but unless verified by high-fidelity models, their answers will always be suspect and insights may be lost.

Heuristics and rules of thumb are useful, but not for identifying vulnerability. If we can evaluate vulnerability precisely, we can create a reasonable heuristic to identify good, budget-limited sets of vulnerability-reducing defensive actions. However, using a surrogate measure of vulnerability (e.g., node degree and basic connectivity indices in a network) leads to sensible defensive plans only if the system is very simple, or an attacker plans attacks using that surrogate. If we base defensive measures on heuristically identified, "near-optimal" attacks, we risk an attack by an aggressor who is smarter than our heuristic.

Reliability is not the answer. We must protect collections of critical components in our infrastructure systems, rather than backing up the least-reliable components.

Malicious, coordinated attacks can be more damaging than random acts of nature.

The attacker has the advantage. This is the reverse of classical military theory and occurs, in part, because of the asymmetric nature of this conflict: The defender must protect a huge, dispersed target set, while the attacker need only focus on a small set of targets chosen to maximize damage. The attacker also has an advantage in terms of information.

The data are available to everyone. Governmental agencies have produced Web sites that offer much useful information to citizens and terrorists alike. While many Web sites have been redesigned to reduce access to potentially dangerous information, exceptions abound. We advise any owner of a public Web site to appoint an independent red team to analyze that site with intent to cause harm. There must be a proper balance between the public's right to know and advertising our vulnerabilities.

Some systems are naturally robust, while others are not. Our road networks are remarkably robust; fuel pipeline-and-storage systems are highly fragile; most other systems lie somewhere in the middle.

Hardening infrastructure from attack can be expensive. However, if we understand the nature of the most damaging attacks, we can improve a system's robustness for a given budget. Critical infrastructure has been built to be cost-effective with little concern for belligerent attacks; economic incentives to mitigate this situation are lacking. This requires (1) subsidies, changes to tax codes, and regulatory reform, and/or (2) proving the secondary economic benefit of the necessary expenditures (e.g., spare electric transmission capacity could provide new, profitable trading opportunities).

However, there is at least one exception to the "can-be-expensive" rule:

An appropriate level of redundancy or reorganization could be inexpensive. Some types of infrastructure, e.g., supply chains, will benefit, at little expense, by adding a few alternate shipping paths, or by relocating surge capacity wisely.

Secrecy and deception can be valuable. Two-person zero-sum games (e.g., Owen 2001, pp. 11–31) have secrecy at their core, and are likely to be useful in this arena, too.

Conclusion

We face a determined, intelligent enemy who seeks to cause us maximum harm. Worst-case analysis using optimization is crucial to a credible assessment of infrastructure vulnerability and for planning mitigating actions.

Acknowledgments

We thank the Air Force Office of Scientific Research, the Office of Naval Research, the US Department of Homeland Security, the US Department of Energy, and every US uniformed military service for their sustained research support. We also thank INSIGHT, Inc., for helping us study how to protect businesses against hostile threats.

References

Ahuja, R., T. Magnanti, J. Orlin. 1993. *Network Flows*. Prentice Hall, Englewood Cliffs, NJ.

Benders, J. 1962. Partitioning procedures for solving mixed integer variables programming problems. *Numerische Mathematik* **4** 238–252.

Benedetto, M., J. Bridges, D. Doyle, G. Spitz. 2005. Strategic Petroleum Reserve (SPR) interdiction. Red Team Report, OA4202, Naval Postgraduate School, Monterey, CA.

Brown, G., M. Carlyle, J. Salmerón, K. Wood. 2005a. 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, 102–123.

Brown, G., M. Carlyle, J. Diehl, J. Kline, K. Wood. 2005b. How to optimize theater ballistic missile defense. *Oper. Res.* **53** 263–275.

Brown, G., M. Carlyle, T. Harrison, J. Salmerón, K. Wood. 2003a. How to attack a linear program. Presented at 71st Military Operations Research Society Symposium, Quantico, VA, June 10–12.

Brown, G., M. Carlyle, T. Harrison, J. Salmerón, K. Wood. 2003b. Tutorial: How to build a robust supply chain or harden the one you have. Presented at INFORMS Annual Meeting, Atlanta, GA, October 19–22.

Brown, G., M. Carlyle, T. Harrison, J. Salmerón, K. Wood. 2004. Designing robust supply chains and hardening the ones you have. Presented at INFORMS Conf. on OR/MS Practice, Cambridge, MA, April 24–27.

Brown, P. 2005. Optimizing the long-term capacity expansion and protection of Iraqi oil infrastructure. Master's thesis, Operations Research Department, Naval Postgraduate School, Monterey, CA.

Cormican, K., D. Morton, K. Wood. 1998. Stochastic network interdiction. *Oper. Res*. **46** 184–197.

Department of the Army (DOA). 2000a. *Army Field Manual FM 3-01.11*, Appendix A: ADA employment principles, guidelines, and priorities. Retrieved May 12, 2006 http://www.globalsecurity.org/military/library/policy/army/fm/3-01-11/appa. htm.

Department of the Army (DOA). 2000b. *Army Field Manual FM 44-100*, Chapter 4. Fundamentals of army air and missile defense operations. Retrieved May 12, 2006 http://www.globalsecurity.org/space/library/policy/army/fm/44-100/ch4.htm.

Department of Homeland Security (DHS). 2002. National strategy for homeland security. Retrieved May 12, 2006 http://www.whitehouse.gov/homeland/book.

Federation of American Scientists (FAS). 2006. Al Qaeda training manual. Federation of American Scientists. Retrieved August 1, 2006 http://www.fas.org/irp/world/para/aqmanual.pdf.

Garcia, M. L. 2001. *The Design and Evaluation of Physical Protection Systems*. Butterworth-Heinemann, Woburn, MA.

General Accounting Office (GAO). 2004. Border security: Agencies need to better coordinate their strategies and operations on federal lands. Report to Congressional requesters GAO-04-590, US General Accounting Office, Washington, D.C., June.

Golden, B. 1978. A problem in network interdiction. *Naval Res. Logist. Quart.* **25** 711–713.

Harney, R., G. Brown, M. Carlyle, E. Skroch, K. Wood. 2006. Anatomy of a project to produce a first nuclear weapon. *Science and Global Security*. Forthcoming.

INSIGHT. 2006. Strategic analysis of integrated logistics systems (SAILS). Manassas, VA. Retrieved May 15, 2006 http://www.insight-mss.com/data/SAILS_Product_Description1.pdf.

Institute of Electrical and Electronics Engineers (IEEE). 1999. The IEEE reliability test system—1996. *IEEE Trans. on Power Systems* **14** 1010–1020.

Israeli, E., K. Wood. 2002. Shortest-path network interdiction. *Networks* **40** 97–111.

Luft, G., A. Korin. 2003. Terror's next target. Institute for the Analysis of Global Security. Retrieved May 10, 2006 http://www.iags.org/0111041.htm.

Moore, J., J. Bard. 1990. The mixed integer linear bilevel programming problem. *Oper. Res.* **38** 911–921.

O'Neill, R. 1976. Nested decomposition of multistage convex programs. *SIAM J. Control Optim.* **14** 409–418.

Owen, G. 2001. *Game Theory*, 3rd ed. Academic Press, San Diego, CA.

Pan, F., W. Charlton, D. Morton. 2003. A stochastic program for interdicting smuggled nuclear material. D. L. Woodruff, ed. *Network Interdiction and Stochastic Integer Programming*. Kluwer Academic Publishers, Dordrecht, The Netherlands, 1–20.

Pulat, H. 2005. A two-sided optimization of border patrol interdiction. Master's thesis, Operations Research Department, Naval Postgraduate School, Monterey, CA.

Roberts, N., W. Vesely, D. Haasl, F. Goldberg. 1981. *Fault Tree Handbook*. NUREG-0492, US Nuclear Regulatory Commission, Washington, D.C.

Salmerón, J., K. Wood, R. Baldick. 2004a. Optimizing electric grid under asymmetric threat (II). Technical Report NPS-OR-04-001, Naval Postgraduate School, Monterey, CA.

Salmerón, J., K. Wood, R. Baldick. 2004b. Analysis of electric grid security under terrorist threat. *IEEE Trans. on Power Systems* **19** 905–912.

Skroch, E. 2005. Interdicting a nuclear weapons project. Master's thesis, Operations Research Department, Naval Postgraduate School, Monterey, CA.

von Stackelberg, H. 1952. *The Theory of the Market Economy* (translated from German). William Hodge & Co., London, UK.

Wein, L. M., Y. Liu. 2005. Analyzing a bioterror attack on the food supply: The case of botulinum toxin in milk. *Proc. National Acad. Sci.* **102**(28) 9984–9989.

Wood, A., B. Wollenberg. 1996. *Power Generation, Operation, and Control*, 2nd ed. John Wiley and Sons, New York.

Wood, K. 1993. Deterministic network interdiction. *Math. Comput. Model.* **17** 1–18.

Errata

Brown, Carlyle, Salmerón, and Wood: *Defending Critical Infrastructure*, Interfaces 36(6), pp. 530–544.

Note 1: Page 533, column 2

In (AD1-MILP),

$$\text{s.t. } A^T\boldsymbol{\theta} + F^T\boldsymbol{\beta} - F^T P\mathbf{x} \le \mathbf{c}$$

should be

$$\text{s.t. } A^T\boldsymbol{\theta} + F^T\boldsymbol{\beta} - F^T P\mathbf{x} \le \mathbf{c}^T$$

Note 2: Page 539, column 1

(DA1_{YUMA}) should be

$$(\text{DA1}_{\text{YUMA}}) \quad \min_{\mathbf{x}\in X} \max_{\mathbf{y}} \prod_{k\in\mathcal{A}} (q_k^{(1-x_k)} \bar{q}_k^{\,x_k})^{y_k} \qquad (24)$$

$$\text{s.t. } A\mathbf{y} = \mathbf{b} \qquad (25)$$

$$\mathbf{y} \in \{0,1\}^{|\mathcal{A}|} \qquad (26)$$

where $X = \left\{ \mathbf{x} \in \{0,1\}^{|\mathcal{A}|} \,\middle|\, \mathbf{c}\mathbf{x} \le c' \right\}$.

And, (DA2_{YUMA}) should be

$$(\text{DA2}_{\text{YUMA}}) \quad \min_{\mathbf{x}\in X} \max_{\mathbf{y}\ge\mathbf{0}} \left(\mathbf{d} + \mathbf{x}^T(\bar{D} - D)\right)\mathbf{y}$$

$$\text{s.t.} \qquad A\mathbf{y} = \mathbf{b}.$$