INTERNATIONAL SERIES IN OPERATIONS RESEARCH AND MANAGEMENT SCIENCE

ADVANCING THE STATE-OF-THE-ART

Volume Contributors
M. Naceur Azaiez
Michael G.H. Bell
Vicki M. Bier
Louis A. (Tony) Cox
Achille Fonzone
Russ Garber
Donald P. Gaver
Kevin D. Glazebrook
Seth D. Guikema
Kjell Hausken
Patricia A. Jacobs
Urszula Kanturska
Paul Kucik
Gregory Levitin
Elisabeth Paté-Cornell
Jan-Dirk Schmöcker
Jun Zhuang

Game Theoretic Risk Analysis of Security Threats

Edited by
Vicki M. Bier | M. Naceur Azaiez

Springer

Game Theoretic Risk Analysis of Security Threats

INT. SERIES IN OPERATIONS RESEARCH & MANAGEMENT SCIENCE

Series Editor: Frederick S. Hillier, Stanford University

Special Editorial Consultant: Camille C. Price, Stephen F. Austin State University

Titles with an asterisk (*) were recommended by Dr. Price

Axsäter/ *INVENTORY CONTROL, 2nd Ed.*

Hall/ *PATIENT FLOW: Reducing Delay in Healthcare Delivery*

Józefowska & Węglarz/ *PERSPECTIVES IN MODERN PROJECT SCHEDULING*

Tian & Zhang/ *VACATION QUEUEING MODELS: Theory and Applications*

Yan, Yin & Zhang/ *STOCHASTIC PROCESSES, OPTIMIZATION, AND CONTROL THEORY APPLICATIONS IN FINANCIAL ENGINEERING, QUEUEING NETWORKS, AND MANUFACTURING SYSTEMS*

Saaty & Vargas/ *DECISION MAKING WITH THE ANALYTIC NETWORK PROCESS: Economic, Political, Social & Technological Applications w. Benefits, Opportunities, Costs & Risks*

Yu/ *TECHNOLOGY PORTFOLIO PLANNING AND MANAGEMENT: Practical Concepts and Tools*

Kandiller/ *PRINCIPLES OF MATHEMATICS IN OPERATIONS RESEARCH*

Lee & Lee/ *BUILDING SUPPLY CHAIN EXCELLENCE IN EMERGING ECONOMIES*

Weintraub/ *MANAGEMENT OF NATURAL RESOURCES: A Handbook of Operations Research Models, Algorithms, and Implementations*

Hooker/ *INTEGRATED METHODS FOR OPTIMIZATION*

Dawande et al/ *THROUGHPUT OPTIMIZATION IN ROBOTIC CELLS*

Friesz/ *NETWORK SCIENCE, NONLINEAR SCIENCE and INFRASTRUCTURE SYSTEMS*

Cai, Sha & Wong/ *TIME-VARYING NETWORK OPTIMIZATION*

Mamon & Elliott/ *HIDDEN MARKOV MODELS IN FINANCE*

del Castillo/ *PROCESS OPTIMIZATION: A Statistical Approach*

Józefowska/*JUST-IN-TIME SCHEDULING: Models & Algorithms for Computer & Manufacturing Systems*

Yu, Wang & Lai/ *FOREIGN-EXCHANGE-RATE FORECASTING WITH ARTIFICIAL NEURAL NETWORKS*

Beyer et al/ *MARKOVIAN DEMAND INVENTORY MODELS*

Shi & Olafsson/ *NESTED PARTITIONS OPTIMIZATION: Methodology and Applications*

Samaniego/ *SYSTEM SIGNATURES AND THEIR APPLICATIONS IN ENGINEERING RELIABILITY*

Kleijnen/ *DESIGN AND ANALYSIS OF SIMULATION EXPERIMENTS*

Førsund/ *HYDROPOWER ECONOMICS*

Kogan & Tapiero/ *SUPPLY CHAIN GAMES: Operations Management and Risk Valuation*

Vanderbei/ *LINEAR PROGRAMMING: Foundations & Extensions, 3rd Edition*

Chhajed & Lowe/ *BUILDING INTUITION: Insights from Basic Operations Mgmt. Models and Principles*

Luenberger & Ye/ *LINEAR AND NONLINEAR PROGRAMMING, 3rd Edition*

Drew et al/ *COMPUTATIONAL PROBABILITY: Algorithms and Applications in the Mathematical Sciences**

Chinneck/ *FEASIBILITY AND INFEASIBILITY IN OPTIMIZATION: Algorithms and Computation Methods*

Tang, Teo & Wei/ *SUPPLY CHAIN ANALYSIS: A Handbook on the Interaction of Information, System and Optimization*

Ozcan/ *HEALTH CARE BENCHMARKING AND PERFORMANCE EVALUATION: An Assessment using Data Envelopment Analysis (DEA)*

Wierenga/ *HANDBOOK OF MARKETING DECISION MODELS*

Agrawal & Smith/ *RETAIL SUPPLY CHAIN MANAGEMENT: Quantitative Models and Empirical Studies*

Brill/ *LEVEL CROSSING METHODS IN STOCHASTIC MODELS*

Zsidisin & Ritchie/ *SUPPLY CHAIN RISK: A Handbook of Assessment, Management & Performance*

Matsui/ *MANUFACTURING AND SERVICE ENTERPRISE WITH RISKS: A Stochastic Management Approach*

Zhu/*QUANTITATIVE MODELS FOR PERFORMANCE EVALUATION AND BENCHMARKING: Data Envelopment Analysis with Spreadsheets*

Kubiak/ *PROPORTIONAL OPTIMIZATION AND FAIRNESS**

~A list of the early publications in the series is found at the end of the book~

Game Theoretic Risk Analysis of Security Threats

edited by

Vicki M. Bier
Universityo fW isconsin-Madison
UnitedState s

M. Naceur Azaiez
KingS audU niversity
SaudiAr abia

Springer

Editors
Vicki M. Bier
University of Wisconsin-Madison
Madison, WI, USA
bier@engr.wisc.edu

M. Naceur Azaiez
King Saud University
Riyadh, Saudi Arabia
mnazaiez@ksu.edu.sa

ISSN: 08848289
ISBN 978-0-387-877662 ISBN 978-0-387-87767-9
DOI 1007/978-0-387-877662-2

Library of Congress Control Number: 2008935381

Printed on acid-free paper

springer.com

TABLEO FC ONTENTS

Chapter 1

WHY BOTH GAME THEORY AND RELIABILITY THEORY ARE IMPORTANT IN DEFENDING INFRASTRUCTURE AGAINST INTELLIGENT ATTACKS

Vicki M. Bier, Louis A. Cox, Jr., and M. Naceur Azaiez

1. WHY GAME THEORY?

Many countries have multiple critical infrastructures that are potentially vulnerable to deliberate attacks by terrorists or other intelligent adversaries. These include networked infrastructures (oil and natural gas pipelines, electric power grids, transportation routes and facilities, telecommunications networks, water supply), built infrastructures (office buildings, hospitals, convention centers, sports stadiums, storage facilities), information infrastructures (flight control, civil defense, emergency broadcasting), and food-production, processing, and distribution supply chains or networks. Since September 11, 2001, determining how best to protect these and other critical infrastructures against intelligent attacks has become a topic of great concern. Researchers and practitioners have attempted a variety of approaches for dealing with this issue.

One motivation for this book is the belief that methods for guiding resource allocations to defend against intelligent antagonists should explicitly take into account the intelligent and adaptive nature of the threat. As discussed in this chapter, simple approaches to risk assessment that may work well in other contexts (such as protecting against accidents or acts of nature) can fail to correctly anticipate and quantify the risks from persistent, intelligent attackers (Golany et al., to appear). Therefore, a more effective approach is needed. It is natural to turn to game theory for ideas and

principles to help optimize defenses, taking into account that antagonists may adapt their subsequent actions to exploit remaining weaknesses.

Protecting critical infrastructures against intentional attacks is fundamentally different from protecting against random accidents or acts of nature. Intelligent and adaptable adversaries may try different offensive strategies or adapt their tactics in order to bypass or circumvent protective security measures and exploit any remaining weaknesses.

Although engineering risk and reliability analysis are clearly important for identifying the most significant security threats and vulnerabilities to terrorist attacks (particularly in complex engineered systems, whose vulnerabilities may depend on interdependencies that cannot be readily identified without detailed analysis), such analyses do not lead in any straightforward manner to sound recommendations for improvements. In particular, risk and reliability analysis generally assumes that the threat or hazard is static, whereas in the case of security, the threat is adaptive and can change in response to the defenses that have been implemented. Therefore, simply rerunning an analysis with the same postulated threat but assuming that some candidate security improvements have been implemented will in general significantly overestimate the effectiveness of the candidate improvements. (For example, installing anthrax sterilization equipment in every post office in the U.S., if publicly known, might have just caused future attackers to deliver anthrax by Federal Express or United Parcel Service.) The routine application of probabilistic reliability and risk analysis methods developed for system safety and reliability engineering is typically not adequate in the security domain.

Game theory provides one way to account for the actions of intelligent adversaries. However, the use of game theory in this context will generally require probabilities of different consequences (e.g., attack success or failure) for various possible combinations of attacker and defender actions. Quantitative risk assessment and reliability analysis models can provide these consequence probabilities. Thus, security and counter-terrorism analysis could benefit substantially from a combination of reliability analysis and game theory techniques.

Among the available approaches for defending against intelligent attacks, game-theoretic models stand out for their rigor and mathematical depth. They have the striking virtue of attributing intelligence (rather than, for example, random activity) to attackers; in other words, game-theoretic methods anticipate that attackers will attempt to exploit paths of least resistance, rather than acting blindly or randomly in response to defender preparations.

On the other hand, game-theoretic models can be difficult to develop, quantify, apply, and validate. Moreover, some game theory models ascribe

unrealistic levels of hyper-rationality and mathematical or computational sophistication to both attackers and defenders, so that both their predictions for real-world attacks and their prescriptions for real-world defenses may be questionable. In fact, many game-theoretic methods rely on idealized or unrealistic assumptions (such as "common knowledge" of prior beliefs about game structures and payoffs) that may not hold in practice. (These stringent assumptions can sometimes be relaxed when agents interact repeatedly and learn successful strategies by trial and error, but that may not be a realistic model for some types of attack strategies, such as attacks that require substantial advance planning and can be tried only once.)

Game-theoretic models are also sometimes criticized for ignoring important psychological and behavioral factors that may drive real-world behaviors. For example, experimental evidence shows that game-theoretic analyses of many standard problems—such as bargaining games, iterated Prisoner's Dilemma, ultimatum games, escalation games, and others—have only limited predictive power in practice (Shermer, 2008).

Certainly, game-theoretic models of attack and defense can provide useful concepts and computational tools for thinking about risks and allocating resources to defend infrastructure targets against intelligent attackers. The crucial insight from game theory, that rational players base their actions in part on what they expect others to do, is too important to ignore. Yet, we believe that an approach that is more effective and practical than pure game-theoretic analysis is needed. To obtain realistic risk assessments and useful guidance for resource allocation, it is essential to take into account an adversary's possible adaptive behaviors, but without necessarily descending into the mathematical quagmire of full game-theoretic modeling.

The purpose of this book is to delineate some elements of a more effective approach. In the past decade, attack-defense models have been formulated that avoid many of the computational complexities and recursive belief difficulties familiar in full game-theoretic models, while still allowing for the key feature that attackers are assumed to exploit weaknesses left by defenders. Such models typically allow the attacker to formulate an optimized attack, taking into account the defender's preparations. Knowing this, the defender prepares accordingly. The additional sophistication allowed in full game-theoretic treatments (e.g., attackers taking actions based on imperfect information but rational conjectures about what the defender may have done, taking into account the attacker's beliefs about the defender's beliefs about the attacker's priorities) is truncated in favor of a simpler approach (analogous to a Stackelberg game in economics), in which the defender acts first and the attacker responds. The power and utility of this approach are illustrated in several of the following chapters.

More generally, this book discusses how to apply a variety of game-theoretic approaches for defending complex systems against knowledgeable and adaptable adversaries. The results yield insights into the nature of optimal defensive investments to best balance the costs of protective investments against the security and reliability of the resulting systems.

2. WEAKNESSES OF RISK AND RELIABILITY ANALYSIS WITHOUT GAME THEORY

This section explores further some of the weaknesses of current non game-theoretic approaches to risk and reliability analysis in setting priorities for protecting infrastructure against terrorist attacks. For example, approaches that attempt to directly assess probabilities for the actions of intelligent antagonists, without modeling how they depend on the defender's choices (which in turn may depend on assumptions about attack probabilities), are liable to produce ambiguous and/or mistaken risk estimates that are not suitable for guiding resource allocations in practice (National Academy of Sciences, 2008).

Many non game-theoretic applications of risk analysis to security and infrastructure protection rely, either explicitly or implicitly, on the following basic formula:

$$Risk = Threat \times Vulnerability \times Consequence$$

For example, this is the basis of the Risk Analysis and Management for Critical Asset Protection (RAMCAP™) methodology used by the Department of Homeland Security (RAMCAP™ Framework, 2006) for chemical facilities. It has also been applied (at least implicitly) to agricultural terrorism (see for example Linacre et al., 2005), computer security (Computer Science and Telecommunications Board, 2002), and to protection of secure facilities (Indusi, 2003).

RAMCAP™ defines "risk" as "The potential for loss or harm due to the likelihood of an unwanted event and its adverse consequences." Similarly, "threat is based on the analysis of the intention and capability of an adversary to undertake actions that would be detrimental to an asset or population"; "vulnerability" is defined as "Any weakness in an asset's or infrastructure's design, implementation or operation that can be exploited by an adversary"; and "consequence" accounts for "human casualties, monetary and economic damages and environmental impact, and may also include less tangible and therefore less quantifiable effects, including political ramifications, decreased morale, reductions in operational effectiveness or

other impacts." Finally, "conditional risk" takes into account "consequences, vulnerability and adversary capabilities, but excludes intent" (for use in situations where intent would be extremely difficult to assess, such as long-range planning) by simply assuming that an attack takes place.

Systems like RAMCAP™ attempt to model the actions of rational adversaries holistically. Facility owners and operators are encouraged to use their knowledge (or perceptions and judgments) to assess the "attractiveness" of each facility to terrorists. They are asked to estimate potential adverse consequences of attacks using a "reasonable worst case" approach, considering (using what might be termed a "conversational game theory" approach) "that the adversary is intelligent and adaptive and will attempt to optimize or maximize the consequences of a particular attack scenario…" (RAMCAP™ Framework, p. 28).

However, in lieu of formal game-theoretic models, RAMCAP™ proposes two options for risk assessment, one qualitative and one quantitative. The "qualitative" (or semi-quantitative) approach uses tables to categorize and score consequences using ordinal scales. Vulnerability is assessed similarly, using an ordinal scale for "likelihood of attack success." Finally, a "conditional risk matrix" assigns overall conditional-risk scores to pairs of consequence and vulnerability scores, via the formula:

$$\textit{Conditional Risk Score} = \textit{Consequence Score} + \textit{Vulnerability Score}$$

The additive formula (rather than a multiplicative formula) for conditional risk reflects the fact that scales for consequence and vulnerability are logarithmic rather than linear (with each additional increment on the scale reflecting a multiplicative increase in consequence or vulnerability, respectively).

Unfortunately, this qualitative approach to risk rating does not necessarily provide adequate information for guiding resource allocation, since ultimately, quantitative numbers of dollars or other resources must be allocated to risk reduction. Also, Cox et al. (2005) note that qualitative risk rankings can in some cases lead to ranking reversals (with larger risks receiving smaller qualitative ratings).

Quantitative risk assessment in RAMCAP™ is also based on the formula:

$$\textit{Risk} = \textit{Threat} \times \textit{Vulnerability} \times \textit{Consequence}$$

but using approximate vulnerability and consequence estimates. The RAMCAP™ Framework states that, using this approach, "The risk associated with one asset can be added to others to obtain the aggregate risk

for an entire facility… [and] can be aggregated and/or compared across whole industries and economic sectors."

However, even the basic multiplicative formula, *Risk = Threat × Vulnerability × Consequence*, can be incorrect if the different components on the right-hand side are correlated with each other. In particular, positive correlations may arise, for example, if intelligent attackers are more likely to attack targets with high *Vulnerability* and *Consequence* values. In that case, if the *Risk = Threat × Vulnerability × Consequence* formula is applied to expected values of threat, vulnerability, and consequence, the result may substantially underestimate the true risk.

3. SOME LIMITATIONS OF RISK = THREAT × VULNERABILITY × CONSEQUENCE

Ordinarily, *Threat* is intended to reflect the probability of an attack (e.g., in a stated interval of time). However, when intelligent attackers use intelligence about the defender's own beliefs, values, and defenses to plan their attacks, no such probability may exist. For example, assigning a high enough *Threat* value to a facility may justify sufficient defensive investment or vigilance to guarantee that the facility will not be attacked. Thus, threat estimates can be "self-defeating": estimating a threat as high can make the true threat low, and vice versa. This may be viewed as a game in which the defender estimates threats and allocates resources, and the attacker attacks where the estimated threat is low and few defenses have been implemented. Such a game may have no pure-strategy equilibrium (although it may have mixed-strategy equilibrium solutions). This suggests that the concept of threat as a static probability to be estimated may be fundamentally inadequate for protecting against informed, intelligent attackers, since the threat estimate itself may affect the nature of the threat.

"Vulnerability" can also be ambiguous and difficult to calculate via risk-analysis methods such as event trees. Standard event-tree modeling that ignores adaptive, goal-oriented decision-making by the attacker, by treating attacker choices as random events, may underestimate risks. Treating attackers instead as optimizers who calculate their best responses to different possible defenses may lead to different resource allocations, and larger risk reductions (based on deterring attacker actions), than could be achieved using models that ignore the ability of intelligent attackers to adapt their plans as information becomes available before (and during) the course of an attack. Best-response models, in which the defender calculates the attacker's best responses to various conditions and then chooses defensive investments to minimize the damage from the attacker's best response (see for example

Brown et al., 2006), may be substantially easier to formulate and solve than full game-theoretic analyses, yet do a much better job than static vulnerability analyses (e.g., using event trees or other traditional tools of risk and reliability analysis) at capturing important features of likely attacker behavior.

This general approach, mathematically reminiscent of principal-agent games, has been developed by military operations-research experts into a powerful alternative to a full game-theoretic analysis that appears to be practical for many counter-terrorism and infrastructure-protection applications (e.g., Brown et al., 2006). Such hierarchical optimization approaches dispense with the concept of *Threat* as a single number to be estimated, and also go beyond simplistic estimates of vulnerability in the *Risk = Threat* × *Vulnerability* × *Consequence* framework. Instead, they focus on predicting and controlling attacker behaviors (via strategically chosen defensive investments). However, this is less sophisticated than a fully game-theoretic framework, since for example, in Brown et al. (2005), the attacker is assumed to be unaware even of the defender *options* when the defender chooses to keep the defenses secret, while a typical endogenous model usually assumes only that the attacker is unaware of the specific choice made by the defender, and has full knowledge of the defender's options and preferences. This intermediate level of modeling and analysis has proven to be useful in practice, and avoids both the (potentially unrealistic) sophistication and idealizations of game theory, and the extreme simplicity (and concomitant limitations) of the *Risk = Threat* × *Vulnerability* × *Consequence* approach.

Thus, *modeling an intelligent attacker's intelligent (adaptive) behavior can lead to different recommendations and risk estimates from traditional risk analysis.* The lesson is not that risk analysis cannot be used at all. In fact, once optimal defense and attack strategies have been determined, it may then be possible to represent them by event trees or other risk-analysis models. However, risk analysis by itself is not sufficient to represent and solve the key decision problems that defenders confront.

4. SUMMARY OF NON-GAME-THEORETIC MODELS FOR ALLOCATING DEFENSIVE RESOURCES

The preceding discussion suggests that the concepts of "*Threat*" and "*Vulnerability*" as static numbers that experts can estimate for use in calculating risk may be inadequate when risks result from the actions of intelligent actors. Defining "*Threat*" as "Probability of an attack in a stated

period of time" begs the key question of how an attacker makes and revises attack decisions in response to intelligence about the defender's actions and beliefs. In fact, the use of static threat estimates can even be self-defeating, if attackers use intelligence about the defender's own threat estimates to help them decide where and when to attack (for example, choosing to attack targets that the defender believed were not at risk). Conversely, the mere fact of defending a target can make it more attractive to attackers (for example, by increasing the "prestige value" of a successful attack), and thereby make it more likely to be attacked.

Similarly, the probability that an attack succeeds if it is attempted may depend on the attacker's knowledge, contingency plans, and ability to dynamically adapt when obstacles are encountered. The information needed to predict what an intelligent attacker will do and how likely it is to succeed must include such contingency plans, and therefore is better represented by game theory than by traditional risk and reliability analysis. Thus, attempting to assess vulnerability by standard techniques of risk analysis (e.g., event trees or fault trees), without explicit analysis of the attacker's best responses to candidate defenses, can produce misleading risk estimates and poor risk-management recommendations.

An alternative approach (as illustrated, for example, in Brown et al., 2006) is to avoid calculation of *Threat*, *Vulnerability*, and *Consequence* in isolation, and to concentrate instead on optimizing allocation of defensive resources, assuming that attackers will then adopt "best responses" to those allocations. This leads to hierarchical optimization as a framework for guiding risk-management decisions, with a game-theoretic flavor (even if not fully endogenous). The examples given in Brown et al. (2006) suggest that this approach is satisfactory in many applications.

The chapters in this book employ varying levels of game-theoretic sophistication. Some of the chapters adopt the spirit of Brown et al. (2006), and provide game-theoretic problem formulations, without necessarily investigating equilibrium solutions. Others exploit the classical game-theoretic approach more fully. Finally, some chapters could be characterized as implementations of "conversational" game theory—thinking carefully about attacker goals and motivations without actually computing best responses to attacker actions.

5. OVERVIEW OF THE BOOK

The following chapters offer a variety of models and frameworks for more realistic analysis and prevention of intentional attacks. Chapter 2 (by Guikema) provides a state-of-the-art review of the use of game theory in

reliability. It focuses mainly on combining game theory with reliability to model attack-defense situations. The adequacy of game theory in this context is critically discussed, and several non-game theoretic approaches for modeling reliability problems in the presence of intelligent attackers are also reviewed.

Chapter 3 (by Levitin) considers an attack-defense model in which the attacker attempts to maximize the expected damage to a complex multi-state series-parallel system. The defender uses both separation and protection of system elements, as well as deployment of false targets, to reduce the risk of a successful attack. An optimization algorithm is suggested based on assessing the losses due to reduction in system performance, and applying a genetic algorithm for determining optimal defense strategies.

Chapter 4 (by Hausken et al.) considers a game where the defender of a particular asset plays against one active player (a terrorist) and one passive player (nature). The defender can deploy defenses that protect against: (a) terrorist attack only; (b) natural disaster only; or (c) both. A variety of simultaneous and sequential decision situations are considered, with the defender seeking to maximize expected utility.

Chapter 5 (by Azaiez) emphasizes the strategic role of information in a model of attack-defense strategies. The attacker has only partial information about the survivability of the targeted system. The defender, who moves first, attempts to deter the attack through modifications or improvements of the targeted system, by making the chance of a successful attack "sufficiently" low and/or the attacker's estimate of that chance "sufficiently" unreliable. Intermediate levels of partial information are also discussed.

Chapter 6 (by Gaver et al.) deals with some specific games where the defender (a counter-terrorist) seeks to detect a hostile individual (terrorist) early on and neutralize it. The attacker invests time to identify an "attractive" target (e.g., a crowd of people). If not neutralized, the attacker attacks the first identified "attractive" target. The defender faces two types of errors: namely, false identification of the hostile individual within a population of non-hostile individuals; and failure to identify the hostile individual before an attack is launched. The probability of correct identification increases with the observation time of the defender, but of course this increases the probability that an attack is launched before the hostile individual is identified.

Chapter 7 (by Paté-Cornell et al.) applies a combination of game theory and probabilistic risk analysis to investigate links among the actions of different parties (e.g., the government on one side, and potential terrorists on the other) and the resulting state of a system. The results involve the probability of different outcomes (e.g., the chances of different attack scenarios) and the risks of various failure types.

Chapter 8 (by Cox) surveys recent developments in designing resilient networks to protect telecommunications infrastructure against deliberate attacks. Resilient networks are designed to provide enough flexibility, redundancy, and rapid recovery capability that any attempt to disrupt traffic results in automatic re-routing of traffic and uninterrupted service. The focus is on design of network topologies and methods of traffic routing and switching to make communications among nodes resilient to both link and node failures.

Chapter 9 (by Kanturska et al.) considers how to improve the reliability of transportation networks through multi-path routing and link defense in a game-theoretic setting. An illustration of an attacker-defender model shows the importance of mixed route strategies, and illustrates how critical links in the network can be identified. The approach is extended to a defender-attacker-defender game to investigate the optimal set of infrastructure to protect in advance. Varying degrees of visibility of the protection are considered; the results show that the visibility of the protective measures significantly affects their expected benefit.

In summary, the theoretical and applied work in the following chapters shows how defense against an intelligent, adaptive attacker can be designed rationally within the frameworks provided by attacker-defender models. Alternative approaches (both game-theoretic and non game-theoretic) are addressed briefly, but the main message is that attacker-defender models are now well enough developed to provide a new generation of infrastructure-defense planning and optimization algorithms that can substantially improve on earlier approaches.

REFERENCES

Bertsimas, D., and Nino-Mora, J., 1996, Conservation laws, extended polymatroids and multiarmed bandit problems: Polyhedral approach to indexable systems, *Mathematics of Operations Research* **21**(2):257–306.

Brown, G., Carlyle, M., Salmeron, J., and Wood, K., 2006, Defending critical infrastructure. *Interfaces* **36**(6):530–544.

Computer Science and Telecommunications Board, 2002, *Cybersecurity Today and Tomorrow: Pay Now or Pay Later,* National Academy Press, Washington, D.C., http://www.nap.edu/catalog.php?record_id=10274

Cox, L. A., Jr., Babayev, D., and Huber, W., 2005, Some limitations of qualitative risk rating systems, *Risk Analysis* **25**(3):651–662.

Denardo, E. V., Rothblum, U. G., and van der Heyden, L., 2004, Index policies for stochastic search in a forest with an application to R&D project management, *Mathematics of Operations Research* **29**(1):162–181.

GAO, 1999, *Combating Terrorism: Threat and Risk Assessments Can Help Prioritize and Target Program Investments,* Washington, D.C., http://www.gao.gov/archive/1998/ns98074.pdf

Golany, B., Kaplan, E., Mamur, A., and Rothblum, U. G., To appear. Nature plays with dice—terrorists do not: Allocating resources to counter probabilistic and strategic risks, *European Journal of Operational Research.*

Indusi, J. P., 2003, *Terrorist Protection Planning Using a Relative Risk Reduction Approach.* Brookhaven National Laboratory. Upton, New York. BNL-71383-2003-CP. http://www.pubs.bnl.gov/documents/25368.pdf

Infanger, G., 2006, Dynamic asset allocation strategies using a stochastic dynamic programming approach, Chapter 5 in: S. A. Zenios, and W. T. Ziemba, eds., *Handbook of Asset and Liability Management, Volume 1,* North Holland. New York, New York.

Linacre, N. A., Koo, B., Rosegrant, M. W., Msangi, S., Falck-Zepeda, J., Gaskell, J., Komen, J., Cohen, M. J., and Birner, R., 2005, Security Analysis for Agroterrorism: Applying the Threat, Vulnerability, Consequence Framework to Developing Countries. International Food Policy Research Institute, Environment and Production Technology Division, EPT Discussion Paper 138. http://www.ifpri.org/divs/eptd/dp/papers/eptdp138.pdf

RAMCAP™ Framework, 2006, www.asme-iti.org/RAMCAP/RAMCAP_Framework_2.cfm

Sethuraman, J., and Tsitsikilis, J. N., 2004, Stochastic search in a forest revisited, *Mathematics of Operations Research* **32**(3):589–593.

Shermer, M., 2008, *The Mind of the Market: Compassionate Apes, Competitive Humans, and Other Tales From Evolutionary Economics,* Times Books, New York.

Chapter 2

GAME THEORY MODELS OF INTELLIGENT ACTORS IN RELIABILITY ANALYSIS

An Overview of the State of the Art

Seth D. Guikema

Abstract: Much research has been done to develop game theoretic methods for modeling intelligent actors in reliability analysis, especially in the context of modeling intentional attacks against systems. This work has provided strong, and at times unexpected, insights into how to best defend systems against intelligent attacks. General principles of how to best defend systems of specific types against intelligent attacks are emerging that can help system managers allocate resources to best defend their systems. However, there is still work to be done to improve these models. The current game theoretic models for intelligent threats are based on a number of assumptions, and the implications and appropriateness of these assumptions in different situations have not been fully explored. Further research to better characterize how well these assumptions match reality and what impacts changing these assumptions would have on both modeling difficulty and the suggested decisions would be fruitful.

Key words: game theory; risk analysis; reliability; terrorism; rationality; utility functions; sensitivity analysis

1. INTRODUCTION

As discussed in the first chapter of this book, game theory provides a powerful framework for modeling strategic interactions between those defending a system and those attempting to harm that system. Game theory has also been used in modeling the reliability of complex systems and for suggesting the optimal allocation of defensive resources. In this case, the analysis builds from the usual definition of reliability as the probability of successful operation of a system under specified conditions for a defined

period of time to incorporate the effects of intelligent agents on system reliability. The usual definition of reliability still holds, but the nature of the initiating events that may lead to system failure and the types of risk management measures to be considered change substantially. However, the use of game theory as the basis for modeling system reliability in the face of intelligent threats is not uniformly accepted.

This chapter first provides a general overview of the use of game theory in modeling terrorist-defender interactions. This forms the historical and theoretical backdrop against which more recent advances in the use of game theory in reliability analysis have developed. The chapter then reviews the use of game theory in reliability engineering. A discussion of foundational issues and assumptions in game theory and an overview of alternate approaches follow the review of the use of game theory. Finally, the current state of the art in the use of game theory in reliability analysis is summarized.

2. GENERAL OVERVIEW OF GAME THEORY IN MODELING TERRORIST-DEFENDER INTERACTIONS

While there are a number of situations in which strategic interactions are of interest in reliability analysis such as modeling design teams, modeling vandalism, etc., much of the current interest in the use of game theory in risk and reliability analysis stems from increased concerns about terrorist attacks on western interests. Because the use of game theory in reliability analysis has built from game theoretic analyses of terrorist threats, it is important to summarize this past work as a basis for discussing the use of game theory in reliability analysis.

Game theoretic work on modeling terrorist actions has focused on modeling high-level strategic questions such as the effects of different national policies intended to reduce the likelihood or impact of terrorist attacks (Sandler, 2003; Sandler et al., 1983; Sandler and Scott, 1987; Enders and Sandler, 1993, 1995; Rosendorff and Sandler, 2004). Among other conclusions, this past work has suggested that vigorous, national anti-terrorism campaigns may induce terrorists to switch to more spectacular attacks rather than attacks that are less costly to the defender (Rosendorff and Sandler, 2004), particularly if anti-terrorism policy depends primarily on deterrence (Frey and Luechinger, 2002). Past game theoretic and related work also suggests that there is a substitution effect in which terrorists switch from one form of attack (e.g., hijackings) to another form of attack (e.g., assassinations) in response to national policies deterring a specific type

of action (e.g., Sandler and Scott, 1987; Sandler and Enders, 2004), and that deterrence and pre-emption are both generally under-utilized for global terrorism (Sandler and Siqueira, 2006). Other past work has focused on modeling the cycles observed in terrorist activity on the basis of leader-follower games (Faria, 2003) and system dynamics models (Udwadia et al., 2006). Finally, one modeling effort has combined probabilistic risk analysis, Bayesian influence diagrams, and basic game theory for analyzing high-level, schematic terrorist risk (Paté-Cornell and Guikema, 2002). While this model is more closely related to the use of game theory in reliability analysis than the more global terrorist game theoretic models, it still focused on a high-level analysis rather than the detailed analysis of system protection options of interest in reliability analysis.

While the focus of these models is not on the detailed reliability analysis central to this book, past game theoretic models for terrorism do provide the backdrop against which game theoretic reliability analysis models are being developed. They also share three key characteristics. First, past game theoretic models for terrorist activity have assumed that both attackers and defenders follow the rational actor paradigm. That is, they assume that attackers and defenders act to maximize their own utility, given what they know about their opponent's objectives and knowledge. Second, past and current game theoretic models explicitly model the strategic interactions between attackers and defenders. That is, the response of one party to the actions of the others is forecast based on a predictive model of behavior. Other modeling approaches exist, and these will be summarized later in this chapter. Finally, most past game theoretic analyses have treated attacker behavior as deterministic with given targets being deterministically attacked based on game theoretic equilibriums. This does not need to be the case, and both mixed-strategy equilibriums and variations on standard game theory can be used to model stochastic strategies. The past work using game theory to model terrorism issues provides a basis on which the use of game theory in reliability analysis has been built, and similar assumptions are often made in game theory models in reliability analysis. The role and impacts of these assumptions in using game theory for reliability analysis will be discussed later in this chapter.

3. THE USE OF GAME THEORY IN RELIABILITY ANALYSIS

Until relatively recently, game theory was not a major component of reliability analysis. However, in the last few years, analysts have developed game theoretic models for both examining the impacts of strategic

interaction on the reliability of systems as well as initial methods for optimizing system reliability in the face of intelligent threats. These models have progressed from relatively simple, one-shot game theoretic models to models for repeated strategic interactions. The insights provided by the models have correspondingly grown in sophistication. However, there is still work to be done in improving and testing game theoretic models in reliability analysis.

Early work in using game theory in reliability analysis focused on linking probabilistic risk analysis models with basic game theoretic models to incorporate the effects of strategic interactions into reliability analysis. For example, Hausken (2002) develops general models that treat system reliability as a public good and incorporate game theory into reliability analysis. These models focus on the case in which different players value system reliability differently and conflict arises in resource allocation decisions.

In Hausken's model, each component of a system is matched with an individual player, and the reliability of component i, $p_i(x_i,s_i)$, is assumed to depend on both that component's player's strategy s_i (e.g., effort level) and on the technical characteristics of the component x_i (Hausken, 2002). The reliability of the system components determines the overall system reliability p(x,s). Hausken uses a multi-attribute utility function to model a player's utility as a function of their component's reliability, the system's reliability, their strategy, the technical state of their component, and any reimbursements and costs to the player. The utility function is assumed to be exogenously given.

Based on his basic model, Hausken (2002) then addresses a number of examples. The first example is of a series system. In particular, Hausken examines a stylized, perfectly circular island protected by dikes. Each citizen must individually decide whether or not to build and then maintain their dike, incurring a personal cost. The entire island floods (deterministically) if any individual does not build and maintain their section of the dike. If the marginal benefit of a perfectly effective dike system outweighs the cost of building the dikes, there are two Nash equilibriums. In the first, no dikes are built, and all players suffer loss due to a flood. In the second equilibrium, all players build and maintain their dikes, making all players better off. Note that the outcomes are assumed to occur deterministically given the actions of the individuals. The second of Hausken's examples involves a purely parallel system. In the example, each citizen of a fictional country is responsible for operating an anti-ICBM interceptor site, but operating the site is costly. If any individual shoots down the incoming ICBM all players benefit. The equilibriums in this case involve either no-one shooting down the ICBM or a single player shooting down the ICBM at the cost of

obtaining the lowest utility. The other players effectively become free riders at the expense of the single shooter. Note that this model does not consider uncertainty in either the accuracy of the ICBM or the interceptor. The third example is one of mutual assistance among private companies. Each company benefits if other companies agree to provide mutual assistance, but the temptation to be the hold-out and benefit without agreeing to provide costly assistance is strong. There are multiple equilibriums to this game, depending on the relative costs and benefits of providing and receiving assistance.

Hausken's work is significant because it gives the first organized linkage in the reliability and risk analysis literature between reliability analysis based on probabilistic risk analysis and game theory in which system reliability is treated as a public good. While the systems addressed (series, parallel, and summation) are simple from a reliability analysis point of view, the linkage with game theory highlights the importance of individual decision-making in many situations. Since the work of Hausken, a number of other researchers have expanded on this problem, especially in analyzing risk to technical systems due to possible terrorist attacks.

Major (2002) presents a relatively simple game theoretic analysis of target protection decisions. A fundamental assumption throughout much of the paper is that the attackers and defenders value the same thing (e.g., economic losses or lives lost), leading to a zero-sum game. This leads Major to concentrate on minimax defense strategies in which the defender minimizes their maximum possible loss. In a zero-sum game with symmetric and diametrically opposite utility functions, this leads to the optimal defensive strategy, but, as Major (2002) acknowledges, this strategy is generally not optimal when attackers and defenders value different things.

As discussed earlier in this chapter, Paté-Cornell and Guikema (2002) presents a model linking probabilistic risk analysis, influence diagrams, and game theory for analyzing terrorist risk. While the focus was not strictly on reliability analysis for an individual system, the results are applicable to reliability analysis problems. Paté-Cornell and Guikema (2002) developed an influence diagram to model U.S. anti-terrorism decisions and a separate influence diagram to model terrorists' decisions of what target, if any, to attack with what weapon. These two detailed influence diagrams were linked with a relatively simple game theoretic model. However, Paté-Cornell and Guikema (2002) modified traditional game theoretic equilibrium concepts and assumed that terrorists would choose attacks stochastically, with attack probabilities proportional to the expected utility of each attack to them. While this was a fairly arbitrary assumption in the original work, it does correspond to a version of bounded rationality in which players choose

strategies with probabilities proportional to the utilities of the strategies (see Rubinstein, 1997).

In work that is similar in spirit to the approach used by Paté-Cornell and Guikema, Koller and Milch (2003) developed "multi-agent influence diagrams" or MAIDS to capture dependencies between decisions as well as dependencies between uncertainties in multi-player games. The focus of Koller and Milch (2003) was on more general modeling issues, not reliability analysis. However, the approach may prove particularly useful for reliability analysts dealing with large-scale systems because it offers the promise of increased computational efficiency relative to standard methods for solving game theoretic problems.

Bier and her collaborators published the first papers combining game theory and reliability analysis that dealt directly with allocating protective resources among the components of a technical system (Bier and Abhichandani, 2004; Bier et al., 2005). Their work builds from the approach of Hausken (2002) but focuses on a more detailed reliability model of the system. Bier et al. (2005) assume that an attacker acts rationally to maximize the expected amount of damage from an attack, that attacks on different system components succeed or fail independently of one another, and that the timing of failures is not important. Bier et al. (2005) examine the allocation of resources to protect the components of both series and parallel systems assuming both perfect attacker knowledge of defensive actions and no attacker knowledge of defensive actions. The results from Bier et al. (2005) show, for example, that for a single attack against a series system when the attacker has perfect knowledge of defensive actions, the optimal defensive strategy is to try to equalize the probability of failure of each of the components. However, if the same attacker with perfect information can attack more than once, the optimal allocation also depends on the relative marginal costs of decreasing the probability of failure for each of the components. For parallel systems, which are easier than series systems to defend in many cases, the optimal allocation of resources depends the cost-effectiveness of defensive investments, the available budget, and the probability of an attack against the system. In both the series and parallel cases, the results of Bier et al. (2005) suggest that decreasing the probability of system failure is more difficult if the attacker has knowledge of the system defenses than if they do not.

Zhuang and Bier (2007) build from the earlier work of Bier and Abhichandani (2004) and Bier et al. (2005) to model decisions of how best to protect against both natural disasters and terrorism simultaneously. Unlike most earlier work except for Major (2002), Zhuang and Bier (2007) modeled continuous attacker effort rather than discrete attack/no attack decisions. This yields a more realistic model for the many situations in which attackers