Proceedings of the 2009 IEEE International Conference on Systems, Man, and Cybernetics
San Antonio, TX, USA - October 2009

Defensive Dissuasion in Security Risk Management

William L. McGill, *Member, IEEE*

***Abstract*— The purpose of this paper is to explore ways of integrating defensive dissuasion into a probabilistic framework for security risk analysis. Dissuasion influences attacker perceptions and choice with the effect of reducing the probability of occurrence for a particular course of action. Presently, few security risk analysis models offer an approach that explicitly incorporates the dissuasive effect of security in their assessments. This paper offers such an approach based on a simple model of attacker choice. This model suggests a number of alternative strategies for dissuading attackers from acting on a particular opportunity that threatens the interests of a protector. When uncertainty about the attacker is severe, this paper suggests an approach for estimating probability of attack that accounts for the dissuasive effects of countermeasures based on a worst-case attacker whose interests mirror the concerns of the protector. In addition, this paper discusses how an approach that explicitly accounts for dissuasion would enable decision makers to assess the benefits of countermeasures aimed solely at influencing attacker behavior in a manner favorable to the protector. This paper concludes by identifying directions for future research.**

***Index Terms*—defensive dissuasion, security risk management, homeland security, attacker perceptions, threat assessment**

I. Introduction

The objective of any security system is to make progress toward “freedom from danger”. Security risk management (SRM) seeks to allocate limited resources toward this goal in a cost-effective manner, which includes investments in prevention, protection, response and recovery [1]. In practice, typical strategies to reduce security risk focus on vulnerability reduction through wise implementation of engineered interventions, watchful guards, loss control strategies, security policies and training programs. Yet, most security managers would agree that prevention is the main objective of security – an unsuccessful attack is less desirable than no attack. We must ask, then, how and to what extent do protector actions influence the probability of attack?

Admittedly, these questions are very difficult to answer and have yet to be resolved in any rigorous way owing to the severe uncertainties surrounding attacker choice. Traditional approaches to security risk analysis sidestep the issues by assuming either the probability of attack is fixed and uninfluenced by protector actions or that the probability of attack is completely unknowable and thus cannot be assessed. The unstated, though often unintended, assumptions of the former “constant threat” approach suggests statistical attackers that behave much like earthquakes, tornadoes and other naturally occurring events governed by a physical process; this is in stark contrast to the more realistic portrayal of attackers as strategic threats guided by reasoning and preference [2]. The latter “conditional risk” approach avoids the extreme uncertainty regarding adversarial threats by removing the requirement for assessing probability of attack from explicit analysis. Consequently, it is left to subjective opinion how alternative scenarios compare in likeliness when forming aggregate judgments of risk based on assessed vulnerability and impact, an unfortunate situation given the widely held view that unaided human judgment is often flawed, particularly when the decision situation is unfamiliar and shrouded by uncertainty (e.g., see [3]). Explicit consideration of the factors that shape the probability of attack, even if descriptive, would improve decision support for security.

Fortunately, much is already understood about how attackers make choices in a competitive environment. The now classic study by Sandler and Lapan [4] demonstrated analytically that, given two targets of equal value to the attacker, the reduction in probability of success at one target shifts attacker interest toward the less-protected (and hence, relatively more attractive) target. The major insight here is that vulnerability reduction only reduces risk to the point where attacker interest is deflected toward less protected targets; all else being equal, a further reduction in the vulnerability does nothing more to reduce risk (i.e., excess protection yields a zero return on investment). Accordingly, there is some level of protective capability that dissuades a particular attack, which in essence shifts the corresponding probability toward zero in exchange for increasing the probability assigned to other attacker opportunities.

Despite its significance, a major point of contention in the cited study [4] is the assumption that attackers have perfect knowledge of protector capabilities. To remove this assumption means to accept that attackers carry out some sort of reconnaissance activity to draw inferences about protector vulnerabilities and the value of a successful attack, but admits that these inferences may be imperfect and potentially inaccurate. Accordingly, strategies aimed at influencing attacker perceptions in a manner that favorably impacts security risk (e.g., [5] for information and cyber security) can be devised with an understanding of an attacker’s belief system, observational channels and sensitivities [6]. An

W. L. McGill is with the College of Information Sciences and Technology, The Pennsylvania State University, 101P IST Building, University Park, PA 16802 USA (phone: 814-867-0270; fax: 814-865-7559; e-mail: wmcgill@ist.psu.edu).

978-1-4244-2794-9/09/$25.00 ©2009 IEEE

appearance of protector strength or of lessened value associated with a successful attack, whether real or ruse, coupled with ambiguation or concealment of truth has its role to play in reducing the probability of attack through direct influence of attacker perceptions and choice.

The SRM literature offers a number of alternative approaches for assessing risk and estimating risk reduction in a variety of security contexts, typically by relating security improvements to an explicit reduction in vulnerability (e.g., "constant threat" and "conditional risk" approaches). Few approaches explicitly modify the threat aspect of security risk in response to a risk mitigation strategy despite widespread recognition that an appearance of enhanced security may deflect attacker interest toward less protected or higher valued targets (e.g., [7,8]). This deficiency is more pronounced in those situations where one wishes to assess the benefit of strategies targeting attacker perceptions, such as measures to deny information, distract attention, dull interest and display an exaggerated or ambiguous image of strength. Such deceptive or informational countermeasures have minimal, if any impact on the tangible capabilities of a security system, though in principle they can be as effective at reducing risk. Yet without a suitable behavioral model of attacker choice, it is near impossible to account for the dissuasiveness of perceived security.

The purpose of this article is to explore ways of integrating defensive dissuasion into a probabilistic framework for security risk analysis. We begin our discussion with a high-level introduction to defensive dissuasion followed by a brief discussion on defensive deception. We then develop a conceptual model for attacker choice that offers one take on the mechanisms by which an attacker is dissuaded from pursuing a particular opportunity. Next we return to the fundamental principles of risk analysis to develop a mathematical expression for security risk that explicitly integrates the dissuasion model. This paper concludes with a brief discussion of these ideas and directions for future study.

II. Key Definitions and Concepts

Consider a *security context* defined by an asset, its protector and an attacker [9]. An *asset* is anything of value, to include both tangible and intangible concerns. A *protector* or defender assigns value to an asset and seeks to protect the asset against any and all harms. An *attacker*, unwittingly or wittingly, threatens harm to an asset, and in the most general sense may be any agent that is deliberately responsive to countermeasures (e.g., thieves, hackers, terrorists, rodents and roaches). A *threat* describes an action or event initiated by an attacker against an asset that may cause its protector to suffer harm or loss. From the attacker point of view, threats are viewed as *opportunities*. A *scenario* is a postulated combination of a particular threat and a particular outcome [10]; depending on the specificity of the scenario statement, a single scenario may embody a seemingly infinite number of variations. For the purposes of this paper, *risk* is defined as the potential for harm or loss arising from a complete set of scenarios in a specified security context defined by a protector, a collection of assets and a variety of attackers with varied intentions and capabilities.

It is widely recognized that an appearance of security, whether real or imagined, dissuades would-be-attackers from aggression [11]. In common English usage, to dissuade a person means to talk him or her out of some course of action by argument, reasoning or entreaty [12]. Dissuasion is the opposite of persuasion; whereas persuasion promotes action by another party, dissuasion promotes inaction. In a security context, *defensive dissuasion* persuades would-be attackers against pursuing a particular course of action that cause the protector to suffer harm or loss. The overall *dissuasiveness* of a protected asset or system is a property that emerges from its interaction with a particular attacker that describes the extent to which the system influences attacker choice in favor of the protector's interests.

There are two fundamental ways that a security system might dissuade would-be attackers: by projecting an image where the attacker must expend significant resources to achieve success and by lessening the attacker's perceived benefits of success [13]. *Deterrence* is the inhibition of undesirable behavior through fear of failure (e.g., perceived denial of success) and consequences of retribution (e.g., perceived threat of swift, certain, and severe sanctions) [14]. From the defender point of view, deterrence influences an attacker's perceived probability of success and perceived loss from failure. To compensate, the attacker must strengthen his own capabilities, improve his knowledge of a target weaknesses to maximize his chances of success or be willing to accept a higher risk of failure. In contrast, *devaluation* is the process of lessening the attacker's perceived value of success to the point where the attacker believes that a successful attack would not yield the desired benefit [15].

Deception is the deliberate misrepresentation of reality for the purposes of causing a target audience to take an action that is in the interests of the defender [16]. Accordingly, the *deception context* consists of three elements: the *deceiver*, the *target* of deception, and the deception *objective*. Deception operates on the knowledge, beliefs and perceptions held by the deceiver's target, and is achieved via a communication false cues and denial of true cues (i.e., what the target sees or does not see) that encourages the target to reason toward conclusions (i.e., what the target thinks) that ultimately causes the target to behave in a manner favorable to the deceiver (i.e., what the target does) [17]. *Defensive deception* aims to decrease the probability of attack through information and actions designed to dissuade would-be attackers from either considering certain opportunities or preventing action on those that have been selected for attack. While having no bearing on the actual performance of a security system, measures such as fake cameras, decoy guards, signage and mock targets serve to dissuade attackers by creating the appearance of tight security and thus increases their perceived chances of failure.

III. A Model of Attacker Reasoning

We assume a specific attacker $a \in A$ intent on achieving a particular, albeit often vague, goal G. A goal can take on a seemingly infinite number of forms, such as "cause panic," "kill a lot of people," "gain root access to a server," "steal a lot of money," etc. and any combination thereof. The attacker believes he possesses capabilities C consisting of knowledge, skills, abilities and resources (e.g., time, money, access). We admit that the goals and capabilities of an attacker are dynamic and that an attacker's goals shape his capabilities, which then further shape his goals, *ad infinitum*. For sake of the present discussion we assume a snapshot in time where a fixed G and C coexist and pay no regard to how they evolved, and we consider the deliberation among the options to adopt or not adopt a particular course of action. That is, the present model does not include decisions to pursue a particular course of action in the future, but rather focuses on the course of action an attacker would choose right now if he decided to act.

An open question is the extent to which an attacker considers and discriminates among a variety of attack opportunities. Here we define an *opportunity* as a course of action that can be adopted by an attacker with the aim of achieving a particular goal. An opportunity from the attacker's point of view is equivalent to a *threat* from the protector's point of view. We assume an attacker willing to entertain any imagined opportunity that he believes to be consistent with his goals. Of interest to our discussion is the decision model employed by this attacker. For simplicity, we assume an attacker with two decision rules as follows. Given some imagined opportunity, the attacker asks:

1. Can this opportunity achieve my goals?
2. If so, are my chances of success sufficiently high?

In the exploratory phase, the attacker imagines one or more opportunities to achieve his goals. The extent of an attacker's imagination is bounded by his motivations, experiences, creativity, inspiration and environment [18]. For each opportunity, the attacker leverages his knowledge of how the world works and all information available to him to judge whether the opportunity is viable with some appreciable degree of confidence. This question requires the attacker to make two supporting judgments. First, what is the probability that the potential gain from success g meets or exceeds the attacker's goals G? Second, what is the threshold level of confidence α needed to select this alternative (the decision criterion)? That is, does the probability of $g \geq G$ meet or exceed some threshold α (i.e., does $\pi(g \geq G) \geq \alpha$, where $\pi(X)$ denotes the attacker's perceived probability of event X)? The question of viability is binary, where a "yes" admits the opportunity into the attacker's set of viable alternatives (i.e., a *viable opportunity*).

For each viable opportunity, the second decision rule deals with what the attacker believes to be his chances of successfully achieving his goals. This question requires the adversary to make two judgments. The first judgment concerns the perceived probability of being successful, and considers how the attacker perceives the effectiveness of his capabilities C relative to the protector's capabilities c. We assume here that C is so defined that success is believed to occur if $c < C$ at the time of attack. The second judgment centers on the threshold level of confidence β expressed in terms of a minimum acceptable perceived probability of success at the time of attack. As will be shown below, this value of risk changes depending on the attacker's risk tolerance relative to the perceived consequences of failure and perceived penalties of inaction. Ultimately, the question is whether $\pi(c < C) \geq \beta$.

To summarize the two decision rules defined above in compact form, for a given opportunity:

Q1. $\pi(g \geq G) \geq \alpha$?
Q2. $\pi(c < C) \geq \beta$?

If the attacker's answer to each question above is "yes" for a given opportunity, then we assume that the opportunity is feasible (i.e., a *feasible opportunity*). If the answer is "no" to one or both questions, then the opportunity is no longer considered at that time. However, we note here that these judgments may be difficult to make with confidence, particularly if the protector employs a strategy aimed at denying information on security investments or system design. For example, in the case of pure incertitude, an attacker would only be able to assign a probability of [0,1] for each of $\pi(g \geq G)$ and $\pi(c < C)$.

Assuming a rational attacker, one could in principle characterize attacker preferences over a set of feasible opportunities $e \in E_a$ by a utility function U_a as follows:

$$U_a(e) = U_a(g,l,c,C) \tag{1}$$

where g, c and C were defined previously and l is the attacker's perceived consequences of a failed attempt. We assume here a determined attacker that will give up only if interdicted and neutralized by the protector, or rather that a loss from failure is assured. Assuming that (1) takes the form of a simple expected utility model with monotonic functions for g and l, we have:

$$U_a(e) = \pi(c < C)g - \pi(c \geq C)l \tag{2}$$

where we recognize that $\pi(c \geq C) = 1 - \pi(c < C)$. Here we assume an attacker that is immune to the sunk-cost fallacy, that is, the attacker does not consider the cost of the capabilities he already possesses and the effort it took to acquire them. This assumption largely applies to those one-shot attacks executed by an individual or group of individuals (e.g., suicide bombing), but we admit that this assumption does not hold as well for the case of a terrorist or criminal organization that seeks to appropriate resources in a manner that maximizes overall gain from multiple attacks. Again, the focus here is on an attacker that already possesses fixed capabilities C and intends to act on a single viable opportunity.

For the attacker to choose to act on a viable opportunity, we require that the expected utility in (2) meets or exceeds the perceived disutility or costs of inaction u (i.e., $U_a \geq u$), which may be larger in magnitude than the gain from success. From (2) this requires:

$$\pi(c < C) \geq \min\left(\max\left(\frac{1-\gamma}{1+\delta},0\right),1\right) \tag{3}$$

where $\gamma = \frac{u}{l}$ is the inaction/loss ratio and $\delta = \frac{g}{l}$ is the potential gain/loss ratio. According to this formulation and comparison with Q2, we have:

$$\beta = \min\left(\max\left(\frac{1-\gamma}{1+\delta},0\right),1\right) \tag{4}$$

where the maximum and minimum functions are used to ensure that $\beta \in [0,1]$ as required by the axioms of probability. Equation (4) illustrates how perceptions of sanctions from failure and the costs of inaction shape an attacker's risk tolerance: a higher consequence of loss relative to the absolute gain increases the acceptance threshold for an opportunity. A higher perceived cost of inaction relative to gain from success decreases the acceptance threshold. The protector's goal is to decrease an attacker's tolerance for risk by decreasing u and increasing l to minimize γ, by decreasing g and increasing l minimize δ, and by adding sufficient noise to the system to make it difficult to pass judgment on whether $g \geq G$ or $c < C$.

In light of the above discussion and the two attacker decision questions Q1 and Q2, we can identify five defensive dissuasion strategies as follows:

I. Decrease the perceived potential gain from success (g) through devaluation measures, real or ruse; this may decrease the perceived probability of achieving his goals given success in Q1 and thus prevent him from identifying viable opportunities; such actions may also increase δ which in turn increases β

II. Increase the perceived ambiguity about whether $g \geq G$ through information denial or inconsistency; this may make it hard to pass judgment on Q1 and thus prevent an attacker from identifying viable opportunities

III. Increase the perceived magnitude and extent of protector capabilities (c) through deterrence measures, real or ruse; this may lower the perceived probability of success in Q2 and thus prevent him from identifying feasible opportunities

IV. Increase the perceived ambiguity about whether $c < C$ through information denial or inconsistency; this may make it difficult for the attacker to pass judgment on Q2 and thus prevent him from identifying feasible opportunities

V. Increase the perceived penalties for failure (l) through deterrence strategies, real or ruse; this may decrease l which in turn increases β

Similarly, the two decision questions also suggest four additional dissuasion strategies as follows:

VI. Increase the desired goals G; this may influence the judgment for Q1 in favor of the protector

VII. Decrease the attacker's perceived strength of his capabilities C; this may influence the judgment for Q2 in favor of the protector

VIII. Increase the personal goal certainty criterion α; this may influence the judgment for Q1 in favor of the protector

IX. Decrease the perceived cost of inaction u; this may increase β

Since the focus of this paper is on how protector/defender actions influence the probability of attack against their interests, we emphasize defensive dissuasion (i.e., types I-V). Defensive deception, then, seeks to enhance dissuasion types I-V through a combination of measures aimed at increasing ambiguity (A-Type deception to enhance dissuasion types II and IV) and misleading attackers toward false conclusions (M-Type deception to enhance dissuasion types I, III and V) [17].

IV. Linking Dissuasion with Security Risk

The following discussion offers a conceptual model for thinking about security risk and the roles dissuasion and deception play in mitigating risk. Appealing to the first principles of risk analysis [19], security risk analysis focuses on reasoning toward answers to the following triplet of questions (slightly modified from the original trio):

1. How can I be attacked?
2. What is probability of a successful attack?
3. What is the impact of a successful attack?

Mathematically, security risk is described by a complete set of answers to these questions of the form $<s,\Pr(s),V(s)>$, where s is a scenario of interest (e.g., successful attack of a particular type), $\Pr(s)$ is the probability of the scenario, and $V(s)$ is the valuation of the scenario expressed as a utility or probability distribution over some natural loss dimension. Letting each $s \in S$ be as unique pairing of an initiating event $e \in E$ and outcome $o \in O$, the probability of a successful attack can be expressed as:

$$\Pr(s) = \Pr(e)\Pr(o \mid e) \tag{5}$$

where $\Pr(e)$ is the probability of occurrence for the initiating event e (e.g., probability of attack attempt) and $\Pr(o|e)$ is the probability that the outcome o follows from this event. Note that in the more general sense, o can represent a specific system state far right of "boom" on the attack timeline, such as number of immediate fatalities, impact on GDP, etc. following the explosion of a device, successful hack attempt, or theft of sensitive material. As we are presently focused on a security problem, the emphasis is on protection and prevention and and the only two outcomes of concern are "successful attack" and its complement "unsuccessful attack."

The probability distribution in (5) constructed over the set of all s conveys an "objective" unvalued assessment of risk in that it does not assign a value or utility to any particular s. By including a valuation of s, one can express risk as an expected utility $\bar{V}$:

$$\bar{V} = \sum_{(e,o)\in E\times O} \Pr(e)\Pr(o \mid e)V(e,o) \tag{6}$$

where the summation is taken over all scenarios. Equation (6) is consistent with the prototypical "*Risk* = *Threat* x *Vulnerability* x *Consequence*" model often discussed in the security literature, where "threat" is the probability of an initiating security event, "vulnerability" is the conditional probability of attacker success outcome given attempt, and "consequence" is the valuation of the scenario expressed as a probability distribution on some natural dimension of loss, as a personal utility or preference measure, or combination thereof. Note, however, vulnerability is interpreted here in the narrow sense of "security vulnerability," and it can be readily expanded or narrowed further depending on the interpretation of outcome.

A. Determining Probability of Attack Attempt

The following supposes that a utility function exists for each attacker $a \in A$. We assume a non-empty set of feasible opportunities P_a containing elements p (i.e., $p \in P_a$) that may or may not span the interests of multiple protectors. If P_a consists of two or more opportunities (i.e., $|P_a| \geq 2$), we assume the attacker pursues the one with the highest perceived expected utility at the time the choice is made, p_a, or:

$$p_a = \sup_{p\in P_a}\left(U_a(p)\right) \tag{7}$$

In a world of perfect information on who the attackers are and what are their capabilities, knowledge and preferences, we would leverage (7) to prioritize protection efforts to the point where no opportunities remain for each would-be attacker. However, in practice protectors are uncertain about what each particular attacker knows or thinks, which opportunities are on his mind and how he values each opportunity. Accordingly, at best we can use our limited knowledge to construct a probability distribution over all opportunities of concern to a particular protector as follows:

$$\Pr(p_a \mid A) = \sum\nolimits_{a\in A} \Pr(p_a \mid P_a, a, A)\Pr(P_a \mid a, A)\Pr(a \mid A) \tag{8}$$

where the summation is taken over all attacker types $a \in A$, each with their own value system, awareness of opportunities, etc. The protector's strategic objective is to minimize $\Pr(p_a \mid A)$ for all p_a belonging to his set of perceived threats E (i.e., maximize probability of no attempt). According to (7) and (8), this goal can be achieved by discouraging interest in an opportunity completely (e.g., minimizing the chances $p_a \in E$) or at least relative to other less risky opportunities (e.g., minimizing the chances p_a such as by deflecting attacker interest to $p_a \notin E$). Both of these objectives insist on taking steps to influence dissuasion types I–V described in the previous section.

B. The "Mirror Image" Assumption

When uncertain or ignorant about attacker interests, a protector can adopt a mirror-image approach that assumes all perceived threats represent feasible attacker opportunities, the attacker is fully aware of protector capabilities, and that the attacker's perceived value for an opportunity is equal in magnitude to its impact on the protector. According to (7) and (8), the mirror-image assumption describes a "worst-case attacker" that pursues the path of greatest opportunity, or:

$$e_a = \sup_{e\in E} \sum\nolimits_{o\in O} \Pr(o \mid e)V(e,o) \tag{9}$$

We admit here the possibility of more than one solution to the optimization problem in (9) up to the limiting case where all opportunities are equally valued by the attacker despite quantitative differences in the valuation. Thus we define $E_a \subseteq E$ as the set of all solutions to (9). Accordingly, the probability of any particular security event $e \in E_a$ is:

$$\Pr(e) = \begin{cases} \dfrac{1}{|E_a|} & \text{if } e \in E_a \\ 0 & \text{otherwise} \end{cases} \tag{10}$$

A more general approach under the mirror-image assumption that includes (10) as a special case follows [10]:

$$\Pr(e) = \begin{cases} \dfrac{\left[U_a(e)\right]^b}{\sum_{p\in E}\left[U_a(p)\right]^b} & \text{if } e \in E \\ 0 & \text{otherwise} \end{cases} \tag{11}$$

where $b \in [0,\infty)$ is a bias parameter that weights the importance of relative differences in an attacker's perceived utility $U_a(e)$ for each $e \in E$. Considering the extreme cases, a value $b = 0$ results in a uniform probability distribution across all e (i.e., attacker indifference among options) whereas a value $b = \infty$ reduces (11) to (10). Under the mirror-image assumption, the expression for U_a in (2) can be rewritten as:

$$U_a(e) = \sum\nolimits_{o\in O} \Pr(o \mid e)V(e,o) \tag{12}$$

where it is assumed that the attacker's perceptions of success match the protector's assessed security vulnerability (i.e., $\Pr(c < C) = \Pr(o \mid e)$), the attacker's gain from success matches the protector's valuation of harm (i.e., $g = V(e,o)$), the attacker perceives no loss from failure (i.e., the "fanatical" attacker per

[4]), and the cost of inaction is infinite. This final assumption admits all possible security events into consideration regardless of the polarity of the utility. An alternative assumption might set the cost of inaction at zero, which then considers only those events for which the utility is non-negative.

V. Discussion and Conclusions

Throughout this paper, we assumed the simple utility model in (2) as the basis for conceptualizing dissuasion in a security context. In truth, according to (8), we require utility functions for all sorts of attackers to arrive at a meaningful estimate of probability of attack. At the extreme where each individual potential attacker has a unique utility function, the required utility expressions may number in the billions. For this reason, we admit that the true utility function of any particular attacker is, and perhaps will always remain, elusive. A focus of future research could be on the development of prototypical attacker types (e.g., "opportunists" vs. "explorers" vs. "hackers" such as described in [20] for cyber assaults) that compensate for individual variation via appropriate uncertainty distributions. Despite this, the simplified function presented in (2) is a useful first attempt at explaining how attacker perceptions influence security risk. Future extensions to this model include admitting a multi-shot attacker that seeks to conserve resources for later yet-to-be conceived attacks (e.g., add cost into the utility expression) and stepping backward along the attacker timeline to a point where the attacker decides whether to pursue particular capabilities or information in anticipation of acting on a particular future opportunity (e.g., adding a decision whether to pursue a capability that is prerequisite to deciding whether and where to attack).

Meanwhile, echoing the views of many security practitioners, oftentimes just thinking about a problem using a suitable conceptual framework yields helpful insights. While this paper does not advocate strict quantification of dissuasion to guide defensive resource allocations, it does insist that a formal conception of deterrence and devaluation will enhance a protector's ability to reason about security risk and the overall effectiveness of protective countermeasures. Absent reasonably accurate models of attacker behavior and beliefs for all relevant attackers, the mirror-image assumption provides an approach that assumes a "worst-case attacker" bent on pursuing those opportunities whose disutility is greatest from the protector point of view. Unfortunately, the assessed amount of risk reduction under this assumption may differ from the true benefits and perceptions of a particular attacker, and thus defensive allocations may be suboptimal with respect to the actual threat environment. An open question remains as to whether and to what extent the mirror-image approach is conservative for security risk assessment under extreme attacker uncertainty.

Mindful of the nine types of dissuasion described in this paper, an interesting line of future research might examine the mechanisms by which the display or concealment of different measures aimed at strengthening the protective capabilities of a system also shape attacker perceptions of loss, gain and probability of success. For example, such work can build on the research areas highlighted by a Defense Nuclear Agency study from the early 1980's [21] or integrate recent game theory results [22]. As a whole, we should seek to understand, and hopefully measure, the dissuasive value of a particular security system configuration in a variety of security contexts. We should also seek to understand the extent to which deceptive measures aimed purely at manipulating attacker perceptions influence risk. Such insights may provide us with techniques for assessing the cost-effectiveness of defensive deception.

References

[1] Purpura, P. P. (2007). *Security and Loss Prevention*. 5th Ed. Butterworth-Heinemann.

[2] Golany, B., Kaplan, E. H., Marmur, A. and Rothblum, U. G. (2009). "Nature Plays With Dice – Terrorists Do Not: Allocating Resources to Counter Strategic versus Probabilistic Risks." *European Journal of Operations Research*, 192(1): 198-208.

[3] Gilovich, T. (1993). *How We Know What Isn't So: The Fallability of Human Reason in Everyday Life*. Free Press.

[4] Sandler, T. and Lapan, H. E. (1988). "The Calculus of Dissent: An Analysis of Terrorists' Choice of Targets." *Synthese*, 76(2): 245-261.

[5] Rowe, N. C. and Goh, H. C. (2007). "Thwarting Cyber-Attack Reconnaissance with Inconsistency and Deception." *Information Assurance and Security Workshop*, 20-22 June 2007, 151-158.

[6] Whaley, B. (1969). *Strategem: Deception and Surprise in War*. Center for International Studies, Massachusettes Institute of Technology.

[7] McGill, W. L., Ayyub, B. M. and Kaminskiy, M. P. (2007). "Risk Analysis for Critical Asset Protection." *Risk Analysis*, 27(5): 1265-1281.

[8] Pate-Cornell, M. E. and Guikema, S. D. (2002). "Probabilistic Modeling of Terrorist Threats: A Systems Analysis Approach to Setting Priorities Among Countermeasures." *Military Operations Research*, 7(3): 5-23.

[9] Manunta, G. (1999). "What is Security?" *Security Journal*, 12(3): 57-66.

[10] McGill, W. L. (2008). *Critical Asset and Portfolio Risk Analysis for Homeland Security*. PhD. Dissertation, Reliability Engineering Program, University of Maryland, College Park.

[11] Fuqua, P. and Wilson, J. (1977). *Terrorism: An Executive's Guide to Survival*. Butterworth-Heinemann.

[12] Answers.com (2009). "Dissuade." http://www.answers.com/dissuade.

[13] Lutes, C. (2004). "The Role of Dissuasion in Combating Weapons of Mass Destruction." *Strategic Insights*, III(10).

[14] Roberts, B. (2007). *Deterrence and WMD Terrorism: Calibrating its Potential Contributions to Risk Reduction*. IDA Paper P-4231, Institute for Defense Analyses, Alexandria, VA.

[15] Kennett, M., Letvin, E., Chipley, M. and Ryan, T. (2005). *Risk Assessment: A How-To Guide to Mitigate Potential Terrorist Attacks Against Buildings*. FEMA-452, FEMA Risk Management Series.

[16] Daniel, D. C. and Herbig, K. L. (1982). *Strategic Military Deception*. Pergamon Press.

[17] Bennett, M. and Waltz, E. (2007). *Counterdeception: Principles and Applications for National Security*, Artech House Publishers.

[18] Gigerenzer, G. and Selten, R. (2002). *Bounded Rationality*. Cambridge University Press.

[19] Kaplan, S. and Garrick, B. J. (1981). "On the Quantitative Definition of Risk." *Risk Analysis*, 1(1): 11-23.

[20] Dantu, R., Kolan, P., Akl, R. and Loper, K. (2007). "Classification of Attributes and Behavior in Risk Management Using Bayesian Networks." *Presented at the 2007 IEEE Conference Intelligence and Security Informatics*, 23-24 May 2007, 71-74.

[21] Lapinsky, G. W. and Goodman, C. (1980). "Psychological Deterrents to Nuclear Theft: An Updated Literature Review and Bibliography." National Bureau of Standards Report NSBIR 80-1038.

[22] Zhuang, J., V.M. Bier, and O. Alagoz. "Modeling Secrecy and Deception in a Multiple-period Attacker-Defender Signaling Game," submitted. http://www.eng.buffalo.edu/~jzhuang/Papers/SDD_peer.pdf.