Improving Risk Management

RIESGOS-CM

Análisis, Gestión y Aplicaciones

P2009/ESP-1685

Technical Report 2010.01

Adversarial Risk Analysis for Counterterrorism Modeling

Jesus Rios, David Rios Insua

http://www.analisisderiesgos.org

Adversarial Risk Analysis for Counterterrorism Modeling

Abstract

Recent large scale terrorist attacks have raised interest in counterterrorism models. A unifying theme in this area is the need to develop methods for the analysis of decisions when there are intelligent adversaries aiming at increasing our risks. Most previous approaches have a clear game theoretic flavor, although there are also interesting decision analytic based approaches. We have recently introduced a framework for adversarial risk analysis, aimed at dealing with decision making problems with intelligent opponents and uncertain outcomes. We explore here how adversarial risk analysis may cope with several standard counterterrorism models: simultaneous defend-attack models, defend-attack-defend models and sequential defend-attack models with private information. These may be used as basic model building blocks for complex counterterrorism problems.

1 Introduction

As described in English (2009), terrorism, and appropriate responses to terrorism, represent one of the key challenges for states in this century. In response to recent and potential large scale terrorist attacks, multi billion euro investments are being made to increase safety and security. This has stirred public debate about the convenience of such measures, specially in a context of limited resources within a shrinking economy. In turn, this has motivated a great deal of interest in modeling issues in relation with counterterrorism, with varied techniques and tools from fields such as reliability analysis, data mining or complex dynamic systems. Lively accounts of various techniques and applications may be seen in Wein (2009) or Gutfraind (2009). Parnell et al. (2008) provide an outstanding report on strategies, models and research issues in terrorism risk analysis, with challenges cutting across many fields from Political Science to Operations Research and Management Science.

The key feature of these problems is the presence of two or more intelligent opponents who make decisions whose outcomes are uncertain. Thus, it is no wonder that much of this research has reminiscent game theoretic and risk analytic flavors. Indeed, there is a rich literature in

political science regarding game theory and terrorism, though it places little emphasis on risk analysis aspects, see e.g. Siqueira and Sandler (2006), Arce and Sandler (2007), or Powell (2007). Hausken (2002) provides insights combining risk analysis and game theory. Recent relevant references in the OR/MS literature, include Zhuang and Bier (2007), who compute best responses and Nash equilibria as a basis for allocating resources against terrorism, in situations of both simultaneous and sequential play; and various papers by Brown, Carlyle, Wood and Salmeron (e.g., 2005, 2006, 2008), who present bilevel (max-min, min-max) and trilevel (min-max-min) optimization models for three stylised counterterrorism models such as defender-attacker, attacker-defender and defender-attacker-defender problems. Kardes (2005) surveys various approaches to strategic decision making in presence of adversaries, arguing for the use of robust stochastic games to deal with counterterrorism, pointing out the difficulty in assessing what the adversary aims at doing in this context. Bier and Azaiez (2009) provide a book length treatment of the attacker-defender model and several variants and applications.

A common thread in the above game theoretic approaches is the common knowledge assumption, criticized, e.g. in Raiffa et al. (2002). We believe that such criticism is even more relevant in the counterterrorism context. Most versions of game theory assume that the opponents not only know their own payoffs, preferences, beliefs, and possible actions, but also those of their opponent. Moreover, when there is uncertainty in the game, it is assumed that players have common probabilities over the uncertain variables. This strong common knowledge assumption allows a symmetric joint normative analysis in which players try to maximize their expected utilities and expect the other players to do the same. Therefore, their decisions can be anticipated and predated by Nash equilibria concepts. However, in counterterrorism contexts, players will not typically have full knowledge of their opponent's objectives, beliefs and possible moves. This will be aggravated as participants try to conceal information.

From a policy standpoint, standard game theory methods have been used to study social dilemmas. As an example, Heal and Kunreuther (2006) reflect a prisoner's dilemma in relation with the implementation of security measures in interconnected networks. Security would increase with investments in risk reduction by the network members. However, each member is better off if he contributes nothing but enjoys the benefits of investments by the other network members: defection is the selfish optimal strategy. But if everyone defects, the result is worse for each player than if they would all cooperate. For this reason, third party regulators are needed to impose mechanisms to ensure security investments. Creating such mechanisms is difficult, specially when there are multiple Nash equilibria.

The other mainstream literature in the field has a decision analytic flavor. Among others, Pinker (2007) uses qualitative influence diagrams to assess the short and long-term deployment of countermeasures; Parnell et al. (2010) describe canonical terrorist multi-objective decision trees and influence diagrams to evaluate bioterrorist threats. The recurrent problem is the need to assess the probabilities of the action of the adversaries, which is the key issue within the Bayesian approach to games, see Kadane and Larkey (1982) or Raiffa (1982, 2002). Indeed, Harsanyi (1982) objects to the Bayesian approach as contrary to the spirit of game theory, since the assessment of the adversaries' actions should be based on an analysis of their rational behavior[1], and not on data from previous plays or more psychological methods. Banks and Anderson (2006) provide a numerical comparison to classical and Bayesian approaches to games within a smallpox attack problem. Paté-Cornell and Guikema (2002) present an interesting perspective, suggesting to address the problem of assessing the probabilities of possible attacks by modeling the Attacker's problem from the point of view of the Defender, including the assessment of the Attacker's probabilities and utilities, to assess the expected utilities of the Attacker's actions and estimate the probabilities of these actions as proportional to the Attacker's perceived expected utilities. But this proposal does not take proper account of the fact that the (idealized) Attacker is an expected utility maximizer and, thus, would certainly choose the optimal action. Another possibility would be to undertake a sensitivity analysis approach, see Rios Insua and Ruggeri (2000), taking into account our imprecision about the likely actions of our adversary. This is the approach adopted by Von Winterfeldt and O'Sullivan (2006) within a simple decision tree to evaluate Man-Portable Air Defense Systems countermeasures. This approach may be too involved computationally in complex problems.

In Rios Insua et al. (2009), we have recently introduced the framework of Adversarial Risk Analysis (ARA), to cope with risk analysis situations in which one or more adversaries are ready to increase our risks. Our main application in that paper was geared towards auction models. ARA lays somewhat between both approaches mentioned above, with a Bayesian game theoretic flavor. In supporting one of the participants, we view the problem as a decision analytic one, but use principled procedures which employ the game theoretical structure, and other information available, to assess probabilities on the opponents' actions. We assume that opponents are expected utility maximizers. Our uncertainty in our adversaries' actions stem from our uncertainty about their utilities and probabilities when used to analyze their decision making problems. We note here that the common (prior) knowledge assumption is used in

[1]But Harsanyi assumes full and common knowledge, see our comments above and our later discussion

game theory to avoid the potentially infinite analysis of nested decision models that we arrive at when using the ARA approach. However, avoiding this is at the cost of a strong unrealistic assumption, which in turn invalidates the analysis when it comes to this type of applications. We prefer to be realistic and accommodate as much information from intelligence into the analysis as we can through a structure of nested decision models and stop when no more information can be accommodated, ending the recursion of models with a noninformative or reference probability distribution, which will need to pass a sensitivity analysis test.

In this paper, we show the relevance of ARA to support one of the participants, which we shall call the Defender, when analyzing three important stylized counterterrorism models: the simultaneous defend-attack model; the sequential defend-attack-defend model and the sequential defend-attack model with private information. Our choice of these three models is due to the fact that, as we shall discuss, we view them as basic model building blocks for more complex counterterrorism problems on one hand, and, on the other, because they have been studied in some detail from a standard game theoretic perspective.

For each of these three models, we first describe the basic problem with coupled influence diagrams and decision trees. We then assess critically what would be the standard game theoretic approach and, then, provide the corresponding ARA solution. Our emphasis is on how we may coherently assess the probabilities of the Attacker's actions, in a context in which we aim at supporting decisions of the Defender, versus the Attacker. For completeness, we shall alternate between discrete and continuous models. A simple example will illustrate some of the assessment and computational intricacies. We end up with some discussion.

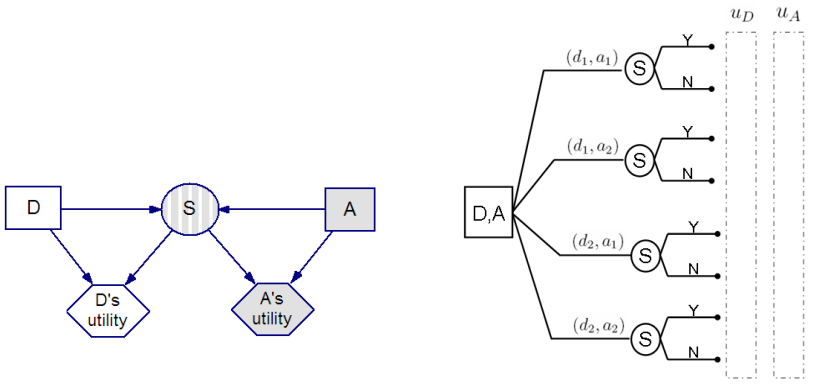
2 Simultaneous Defend-Attack Models

We start by discussing the simultaneous defend-attack model, in which a Defender (she) and an Attacker (he) decide their defense and attack, respectively, without knowing the action chosen by each other. See Zhuang and Bier (2007) for a related discussion. As an example, imagine a case in which the FAA decides whether to introduce undercover marshals in an airplane that might, or not, be hijacked by Al Qaeda.

We shall assume that the adversaries have, respectively, discrete alternative sets $\mathcal{D} = \{d_1, d_2, ..., d_m\}$ and $\mathcal{A} = \{a_1, a_2, ..., a_k\}$ and the only relevant uncertainty is S, marking the success ($S = 1$) or failure of the attack ($S = 0$). Each participant assesses differently the probability of success of the attack, which we describe through $p_A(S \mid d, a)$ and $p_D(S \mid d, a)$. The utility function of the Defender $u_D(d, s)$ depends on her decision and the result of the attack

and, similarly, for the attacker's utility function $u_A(a, s)$. This situation can be represented by two coupled influence diagrams (one for the Defender, one for the Attacker, with a shared uncertainty node in relation with the success of attack), as in Figure 1. We also show a game tree for this problem, with just two possible attacks and defenses, for simplicity.

Figure 1: The simultaneous Defend-Attack model



(a) Influence diagram (b) Game tree ($m = k = 2$)

2.1 The Standard Game Theory Analysis

Under common knowledge assumptions, preferences and beliefs (u_A, p_A, u_D, p_D) are disclosed. Therefore, each adversary knows the expected utility that each pair $(d, a) \in \mathcal{D} \times \mathcal{A}$ would provide to both of them, computed through

$$\psi_A(d, a) = p_A(S = 0 \mid d, a)\ u_A(a, S = 0)\ +\ p_A(S = 1 \mid d, a)\ u_A(a, S = 1),$$

and, similarly, for $\psi_D(d, a)$. A Nash equilibrium solution (d^*, a^*) for this game satisfies

$$\psi_D(d^*, a^*) \geq \psi_D(d, a^*)\ \ \forall\, d \in \mathcal{D} \quad \text{and} \quad \psi_A(d^*, a^*) \geq \psi_A(d^*, a)\ \ \forall\, a \in \mathcal{A}.$$

Finding them may require the use of randomized strategies, see Gibbons (1992). There could be several equilibria with no unambiguous criteria to further discern among them, see Raiffa et al. (2002).

If utilities and probabilities are not common knowledge among the adversaries, the standard game-theoretic approach would model the game as one with incomplete information (Harsanyi, 1967), introducing the notion of player's types. Each player will be of a certain type. This is known to him, but not to his opponent: a player's type represents the private information he may have which is not common knowledge. The Defender's possible types $\tau_D \in T_D$ determine

her utility $u_D(d, s, \tau_D)$ and probability $p_D(S \mid d, a, \tau_D)$, and, similarly, for the Attacker's types $\tau_A \in T_A$. Harsanyi proposes the solution concept of Bayes-Nash equilibrium, still under a strong common knowledge assumption: the adversaries' beliefs about the opponent's types are common knowledge and, therefore, modeled through a common prior distribution $\pi(\tau_D, \tau_A)$. Moreover, it is assumed that the players' beliefs about other uncertainties in the problem are also common knowledge. Then, the solution is computed as follows.

Define, first, the notion of strategy functions for the participants. These associate a decision with each type, that is, $d : \tau_D \to d(\tau_D) \in \mathcal{D}$ and $a : \tau_A \to a(\tau_A) \in \mathcal{A}$. The Defender's expected utility associated with a pair of strategy functions, given any of her privately known types $\tau_D \in T_D$, is

$$\psi_D(d(\tau_D), a, \tau_D) = \int \left[\sum_{s \in S} u_D(d(\tau_D), s, \tau_D)\, p_D(S = s \mid d(\tau_D), a(\tau_A), \tau_D) \right] \pi(\tau_A \mid \tau_D)\, \mathrm{d}\tau_A.$$

Similarly, we can compute the Attacker's expected utility $\psi_A(d, a(\tau_A), \tau_A)$ for a pair of strategy functions (d, a), given any of his privately known types $\tau_A \in T_A$. Then, a Bayes-Nash equilibrium, is a pair (d^*, a^*) of strategy functions, respectively, for the Defender and the Attacker satisfying

$$\begin{aligned} &\psi_D(d^*(\tau_D), a^*, \tau_D) \geq \psi_D(d(\tau_D), a^*, \tau_D) \quad \forall\, d : \tau_D \to d(\tau_D) \quad \text{and} \\ &\psi_A(d^*, a^*(\tau_A), \tau_A) \geq \psi_A(d^*, a(\tau_A), \tau_A) \quad \forall\, a : \tau_A \to a(\tau_A). \end{aligned}$$

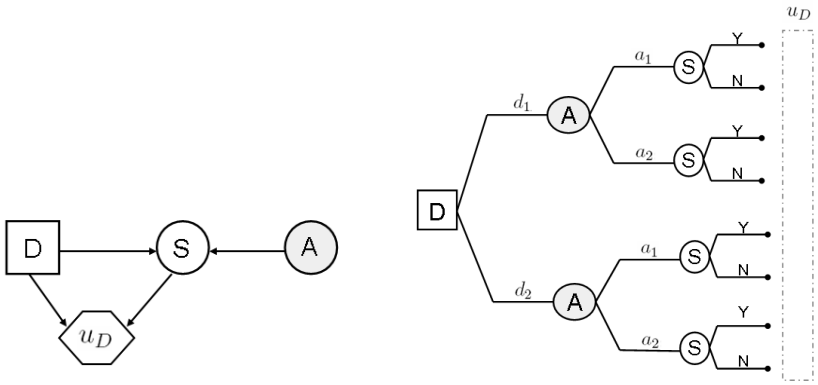
We believe that the underlying common prior knowledge assumption is still counterintuitive and unrealistic, specially in the context of counterterrorism: it implies that players need to disclose, inter alia, their true beliefs about their opponent's type, as well as their private probabilistic assessments in order to be able to compute a Bayes-Nash equilibrium.

2.2 The ARA Approach

More realistically, we now weaken the common (prior) knowledge assumption and support the Defender solving the simultaneous Defend-Attack model. As reflected in Figure 2, the Defender has to choose a defense $d \in \mathcal{D}$, whose consequences depend on the success of the attack simultaneously chosen by the Attacker, which is, therefore, uncertain for the Defender at the time she makes her decision.

By standard decision theory, see French and Rios Insua (2000), the Defender should maximize her expected utility. The Defender knows her utility function $u_D(d, s)$ and her probability assessment p_D over S conditional on (d, a). However, she does not know the Attacker's decision

Figure 2: The Defender's decision analysis



(a) Influence diagram (b) Decision tree ($m = k = 2$)

at node A. She expresses her uncertainty through a probability distribution $\pi_D(A = a)$, over all $a \in \mathcal{A}$. Then, the optimization problem she should solve is

$$d^* = \text{argmax}_{d \in \mathcal{D}} \sum_{a \in \mathcal{A}} \left[\sum_{s \in \{0,1\}} u_D(d, s)\, p_D(S = s \mid d, a) \right] \pi_D(A = a). \tag{1}$$

The Defender needs to assess $\pi_D(A)$. To do so, suppose she thinks of the Attacker as an expected utility maximizer who tries to solve the decision problem shown in Figure 3. The Attacker would look for the attack $a \in \mathcal{A}$ providing him maximum expected utility:

$$a^* = \text{argmax}_{a \in \mathcal{A}} \sum_{d \in \mathcal{D}} \left[\sum_{s \in \{0,1\}} u_A(a, s)\, p_A(S = s \mid d, a) \right] \pi_A(D = d). \tag{2}$$

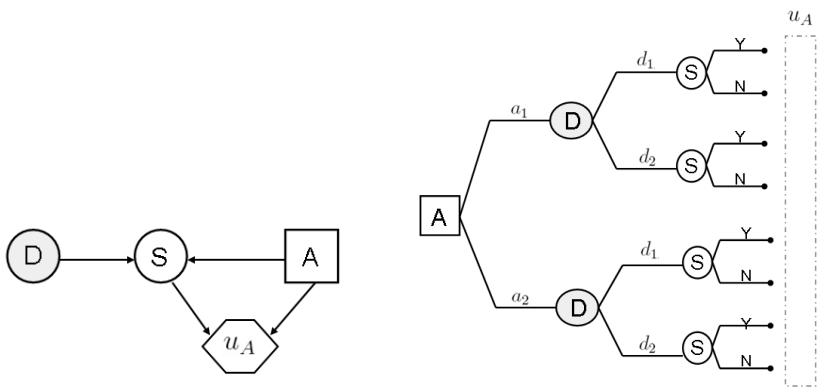
In general, the Defender will not know the Attacker's true utility function and probabilities (u_A, p_A, π_A) required to solve such problem.

Suppose now that we may model all information available to the Defender about (u_A, p_A, π_A) through a probability distribution (U_A, P_A, Π_A). Then, mimicking the argument in (2), we may propagate such uncertainty to compute the following probability distribution, which aids us in assessing $\pi_D(A)$

$$A \mid D \sim \text{argmax}_{a \in \mathcal{A}} \sum_{d \in \mathcal{D}} \left[\sum_{s \in \{0,1\}} U_A(a, s)\, P_A(S = s \mid d, a) \right] \Pi_A(D = d). \tag{3}$$

Note that although (U_A, P_A) could be elicited from the Defender at this step, the elicitation of $\Pi_A(D)$ may require further analysis leading to a next level of recursive thinking, in which the Defender would need to think about how the Attacker analyzes her problem. This is why

Figure 3: The Attacker's decision analysis, as seen by the Defender



(a) Influence diagram (b) Decision tree ($m = k = 2$)

we condition above by (the distribution of) D. Indeed, $\Pi_A(D)$ incorporates two sources of uncertainty:

- the Attacker's uncertainty about the Defender's choice, represented through his beliefs $\pi_A(D)$, and
- the Defender's uncertainty about the probabilistic model π_A used by the Attacker to predict what the Defender will choose, assessed from the Defender's perspective through $\pi_A \sim \Pi_A$.

In the above, the Defender may presume that the Attacker thinks she is an expected utility maximizer trying to solve a decision problem as in Figure 2. Therefore, in order for the Defender to assess the distribution (3), she will elicit $(U_A, P_A) \sim F$ from her viewpoint, and assess $\Pi_A(D)$ through the analysis of her decision problem as thought by the Attacker, mimicking problem (1). This reduces the assessment of $\Pi_A(D)$ to the computation of the distribution

$$D \mid A^1 \sim \text{argmax}_{d \in \mathcal{D}} \sum_{a \in \mathcal{A}} \left[\sum_{s \in \{0,1\}} U_D(d,s)\ P_D(S = s \mid d, a) \right] \Pi_D(A^1 = a), \tag{4}$$

conditional on the Defender being able to assess $\Pi_D(A^1)$, where A^1 represents the Attacker's decision within the Defender's second level of recursive thinking: the nested decision model used by the Defender to predict the Attacker's analysis of the Defender's decision problem. To assess the distribution (4), the Defender needs to elicit $(U_D, P_D) \sim G$, representing her probabilistic knowledge about the Attacker's estimation of her utility function $u_D(d,a)$ and her probability p_D over $S \mid d, a$, when she analyzes how the Attacker thinks about her decision problem. Again,

the elicitation of $\Pi_D(A^1)$ might require further recursive thinking from the Defender. This would lead to the recursive assessments:

`Repeat from` $i = 1$

`Find` $\Pi_{D^{i-1}}(A^i)$ `by solving`

$$A^i \mid D^i \sim \texttt{argmax}_{a \in \mathcal{A}} \sum_{d \in \mathcal{D}} \left[\sum_{s \in \{0,1\}} U_A^i(a,s)\, P_A^i(S = s \mid d, a) \right] \Pi_{A^i}(D^i = d)$$
$$\texttt{with } (U_A^i, P_A^i) \sim F^i$$

`Find` $\Pi_{A^i}(D^i)$ `by solving`

$$D^i \mid A^{i+1} \sim \texttt{argmax}_{d \in \mathcal{D}} \sum_{a \in \mathcal{A}} \left[\sum_{s \in \{0,1\}} U_D^i(d,s)\, P_D^i(S = s \mid d, a) \right] \Pi_{D^i}(A^{i+1} = a)$$
$$\texttt{with } (U_D^i, P_D^i) \sim G^i$$

$i = i + 1$

To simplify the discussion, we have assumed that the recursive decision models used to assess A^i and D^i are a reflection of each other and have the same structure as in Figures 3 and 2, respectively. Moreover, the choice sets for the Defender and the Attacker are the same in all the recursive models: $\mathcal{D}$ and $\mathcal{A}$, respectively. Note that more asymmetries in the models of the hierarchy, including the possibility of different choice sets for the Defender and the Attacker's models, could be easily implemented without extra complications in the methodology.

This hierarchy of nested models would stop at a level in which the Defender lacks the information necessary to assess the distribution F^i or G^i associated with the decision analysis of A^i and D^i, respectively. At this point, the Defender would holistically assign an unconditional probability distribution over A^i or D^i, respectively, without going deeper in the hierarchy, summarizing all remaining information she might have through the direct assessment of $\Pi_{D^{i-1}}(A^i)$ or $\Pi_{A^i}(D^i)$, as might correspond. Of course, should she feel that she has no information available to do so, she could assign a noninformative probability distribution, see French and Rios Insua (2000).

We illustrate the ARA approach to the simultaneous Defend-Attack model with a simple numerical example.

Example. The DHS (the Defender) is considering whether to use (d_1) or not (d_2) undercover marshals in all flights over the US territory to prevent terrorists from hijacking airplanes. The

terrorists (the Attacker) will not know the action chosen by the Defender in their analysis about whether to try (a_1) or not (a_2) to hijack an airplane. The Defender's analysis incorporates the increase in security, the costs as well as the political and social consequences in her utility function. Assume that we are able to assess from the Defender:

- Her utility function $u_D(d, s)$ and probability distribution $p_D(S \mid d, a)$ associated with her decision problem (Figure 2)

$u_D(d, s)$

	$s = 1$	$s = 0$
d_1	50	80
d_2	0	100

$p_D(S = 1 \mid d, a)$

	a_1	a_2
d_1	0.1	0
d_2	0.9	0

- She considers that the threat may come from two different kinds of Attackers: Class I with probability 0.8 and Class II with 0.2. She also presumes that terrorists, whatever their class, will face a decision problem like the one described in Figure 3. The Defender assesses that the utilities and probabilities of a Class I Attacker in (3) are $(U_{A_I}, P_{A_I}) \sim F_I$:

$U_{A_I}(a, s)$

	$s = 1$	$s = 0$
a_1	$Tri(20, 100, 100)$	$Tri(0, 20, 100)$
a_2	100	$Tri(0, 40, 100)$

$P_{A_I}(S = 1 \mid d, a)$

	a_1	a_2
d_1	$\mathcal{U}[0, 1]$	0
d_2	$Tri(0.5, 1, 1)$	0

and Class II Attacker's are $(U_{A_{II}}, P_{A_{II}}) \sim F_{II}$:

$U_{A_{II}}(a, s)$

	$s = 1$	$s = 0$
a_1	$\mathcal{U}[0, 100]$	$Tri(0, 20, 100)$
a_2	100	$Tri(40, 80, 90)$

$P_{A_{II}}(S = 1 \mid d, a)$

	a_1	a_2
d_1	$Tri(0, 0, 1)$	0
d_2	$Tri(0, 1, 1)$	0

where $Tri(min, mode, max)$ and $\mathcal{U}[min, max]$ stand, respectively, for the triangular and uniform distributions.

- Based on the information available, the Defender thinks that a Class I Attacker is smart enough to analyze her problem as in Figure 2. She estimates that a Class I Attacker's

beliefs about her utilities and probabilities in (4) are $(U_{D_I}, P_{D_I}) \sim G_I$:

$U_{D_I}(d,s)$	$s=1$	$s=0$
d_1	$Tri(0,0,40)$	$\mathcal{U}[50,100]$
d_2	$Tri(0,0,40)$	$\mathcal{U}[50,100]$

$P_{D_I}(S=1 \mid d,a)$	a_1	a_2
d_1	$Tri(0,0,0.5)$	0
d_2	$\mathcal{U}[0,1]$	0

The Defender's confidence in these assessments leads her to elicit $\Pi_{A_I}(D_I = d_1)$ as a beta distribution with mean $\pi_{A_I}(D_I = d_1)$ and precision 10, that is $\Pi_{A_I}(D_I = d_1) \sim \mathcal{B}e(\alpha, 10-\alpha)$, where $\alpha = \pi_{A_I}(D_I = d_1) \times 10$.

However, the Defender has no information to assess how a Class II Attacker would analyze her problem, but believes that this attacker estimates that she is more likely to choose d_1, specifically that $\Pi_{A_{II}}(D_{II} = d_1) \sim \mathcal{B}e(75, 25)$.

- Finally, she assigns a noninformative unconditional distribution on what a Class I Attacker thinks to be her beliefs about what he will choose. Thus, $\Pi_{D_I}(A_I^1 = a_1) \sim \mathcal{U}[0,1]$.

To solve the Defender's decision problem, which amounts to finding her maximum expected utility defense in (1), she needs to assess $\pi_D(A = a_1)$, her predictive distribution about what the terrorists will do, where A is the mixture $0.8\ A_I + 0.2\ A_{II}$. Thus,

$$\pi_D(A = a_1) = 0.8\ \pi_D(A_I = a_1) + 0.2\ \pi_D(A_{II} = a_1).$$

A_I represents the Defender's beliefs about what attack in $\mathcal{A} = \{a_1, a_2\}$ a Class I terrorist will choose. Based on (3) and (4), $\pi_D(A_I = a_1)$ could be estimated through Monte Carlo simulation as follows:

1. `For` $k = 1, \ldots, n$, `repeat`

   `Draw` $\pi_{D_I}^k \sim \Pi_{D_I} = \mathcal{U}[0,1]$.

   `Draw` $(u_{D_I}^k, p_{D_I}^k) \sim (U_{D_I}, P_{D_I}) = G_I$

   `Compute`

$$d_I^k = \texttt{argmax}_{d \in \mathcal{D}} \sum_{a \in \mathcal{A}} \left[\sum_{s \in \{0,1\}} u_{D_I}^k(d,s)\ p_{D_I}^k(S = s \mid d, a) \right] \pi_{D_I}^k(A_I^1 = a)$$

2. `Approximate` $\pi_{A_I}(D_I = d_1)$ `through` $\hat{\pi}_{A_I}(D_I = d_1) = \#\{d_I^k = d_1\}/n$.

   `Set` $\alpha = \hat{\pi}_{A_I}(D_I = d_1) \times 10$.

   `Set` $\hat{\Pi}_{A_I}(D_I = d_1) \sim \mathcal{B}e(\alpha, 10-\alpha)$,

3. For $k = 1, \ldots, n$, repeat

Draw $\hat{\pi}^k_{A_I} \sim \hat{\Pi}_{A_I}$.

Draw $(u^k_{A_I}, p^k_{A_I}) \sim (U_{A_I}, P_{A_I}) = F_I$

Compute

$$a^k_I = \texttt{argmax}_{a \in \mathcal{A}} \sum_{d \in \mathcal{D}} \left[\sum_{s \in \{0,1\}} u^k_{A_I}(a, s)\, p^k_{A_I}(S = s \mid d, a) \right] \hat{\pi}^k_{A_I}(D_I = d)$$

4. Approximate $\pi_D(A_I = a_1)$ through $\hat{\pi}_D(A_I = a_1) = \#\{a^k_I = a_1\}/n$.

Similarly, $\pi_D(A_{II} = a_1)$ can be estimated by Monte Carlo simulation as follows, where A_{II} represents the Defender's beliefs about what attack in $\mathcal{A} = \{a_1, a_2\}$ a Class II terrorist will choose.

1. For $k = 1, \ldots, n$, repeat

Draw $\pi^k_{A_{II}} \sim \Pi_{A_{II}} = \mathcal{B}e(75, 25)$.

Draw $(u^k_{A_{II}}, p^k_{A_{II}}) \sim (U_{A_{II}}, P_{A_{II}}) = F_{II}$

Compute

$$a^k_{II} = \texttt{argmax}_{a \in \mathcal{A}} \sum_{d \in \mathcal{D}} \left[\sum_{s \in \{0,1\}} u^k_{A_{II}}(a, s)\, p^k_{A_{II}}(S = s \mid d, a) \right] \pi^k_{A_{II}}(D_{II} = d)$$

2. Approximate $\pi_D(A_{II} = a_1)$ through $\hat{\pi}_D(A_{II} = a_1) = \#\{a^k_{II} = a_1\}/n$.

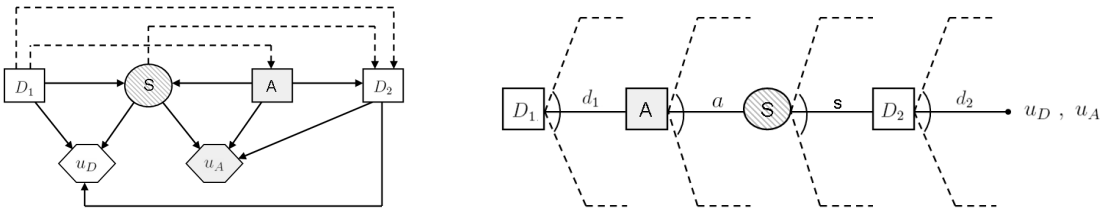
In a run with $n = 1000$, we got the approximations $\hat{\pi}_D(A_I = a_1) = 0.97$ and $\hat{\pi}_D(A_{II} = a_1) = 0.82$. Hence, $\pi_D(A = a_1)$ can be approximated by $\hat{\pi}_D(A = a_1) = 0.8\ \hat{\pi}_D(A_I = a_1) + 0.2\ \hat{\pi}_D(A_{II} = a_1) = 0.94$. The Defender can now solve her decision problem in (1), obtaining that her best defense choice is $d^* = d_1$ with (Monte Carlo estimated) expected utility 77.2, against d_2 with 15.4. △

3 Defend-Attack-Defend Models

We deal now with the sequential defend-attack-defend model, see Brown et al. (2006) for various examples. In it, the Defender first deploys defensive resources. Then, the Attacker, having observed such decision, performs an attack and, finally, the Defender tries to recover, as best as she can, from the attack. Figure 4 shows coupled influence diagrams, with a shared uncertainty node S, and a game tree representing this situation, where nodes D_1 and D_2 correspond to the Defender's first and second decisions, respectively, and node A represents the Attacker's decision. The respective choices will be in $\mathcal{D}_1$, $\mathcal{A}$ and $\mathcal{D}_2$, which we shall assume continuous. Again, the

only relevant uncertainty will be the outcome S of the attack, which depends probabilistically on $(d_1, a) \in \mathcal{D}_1 \times \mathcal{A}$. We shall assume that the consequences for the Defender and the Attacker will depend, respectively, on (d_1, s, d_2) and (a, s, d_2).

Figure 4: The Defend-Attack-Defend model



(a) Influence diagram (b) Game tree

3.1 The Standard Game Theory Analysis

The standard game-theoretic approach requires the Defender to know the Attacker's true utilities and probabilities and the Attacker to know her true utilities and probabilities, and, furthermore, that all this will be common knowledge. Let these utility functions be $u_D(d_1, s, d_2)$ and $u_A(a, s, d_2)$, respectively, and their probability assessments about $(S \mid d_1, a)$ be $p_D(S \mid d_1, a)$ and $p_A(S \mid d_1, a)$. Then, we may compute a solution using backward induction as follows.

At node D_2 of the game tree in Figure 4, the Defender's best response to each observed $(d_1, s) \in \mathcal{D}_1 \times S$ is

$$d_2^*(d_1, s) = \text{argmax}_{d_2 \in \mathcal{D}_2} u_D(d_1, s, d_2). \tag{5}$$

Under the common knowledge assumption, the Defender's behavior at D_2 can be anticipated by the Attacker. Thus, at node S the Defender's expected utility associated with each $(d_1, a) \in \mathcal{D}_1 \times A$,

$$\psi_D(d_1, a) = \int u_D(d_1, s, d_2^*(d_1, s))\; p_D(S = s \mid d_1, a)\; \mathrm{d}s, \tag{6}$$

and the Attacker's

$$\psi_A(d_1, a) = \int u_A(a, s, d_2^*(d_1, s))\; p_A(S = s \mid d_1, a)\; \mathrm{d}s,$$

are known to both of them. Then, the Attacker can find his best attack decision at node A, after observing the Defender's first move $d_1 \in \mathcal{D}_1$, by solving

$$a^*(d_1) = \text{argmax}_{a \in \mathcal{A}} \psi_A(d_1, a).$$

Knowing this, the Defender can find her best decision at node D_1

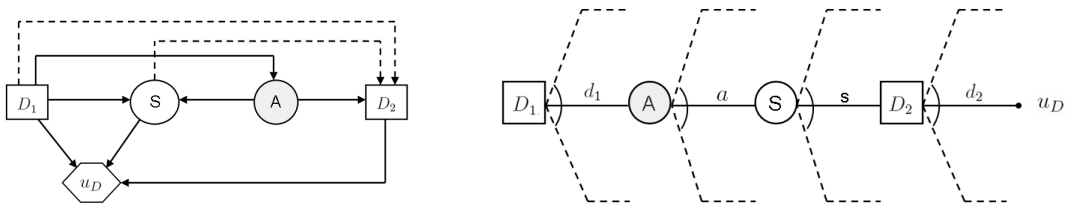
$$d_1^* = \text{argmax}_{d_1 \in \mathcal{D}_1} \psi_D(d_1, a^*(d_1)).$$

Therefore, under common knowledge, game theory predicts that the Defender will choose $d_1^* \in \mathcal{D}_1$ at node D_1; then, the Attacker will respond choosing attack $a^*(d_1^*) \in \mathcal{A}$ at node A; and, finally, the Defender, after observing $s \in S$, will choose $d_2^*(d_1^*, s) \in \mathcal{D}_2$ at node D_2.

3.2 The ARA Analysis

We now give up the strong common knowledge assumption and provide an ARA analysis to support the Defender. For this, we treat the Attacker's behavior at node A as uncertain from the Defender's viewpoint and model her uncertainty. This is reflected in the influence diagram and the decision tree in Figure 5, where the Attacker's decision node has been converted into a chance node, by replacing $\boxed{\text{A}}$ with $\bigcirc\!\!\!\!\text{A}$. Thus, the Defender needs to elicit $p_D(A \mid d_1)$, her predictive distribution about what attack the Attacker will choose at node A against each $d_1 \in \mathcal{D}_1$, besides her standard assessments $u_D(d_1, s, d_2)$ and $p_D(S \mid d_1, s)$.

Figure 5: The Defender's decision problem



(a) Influence diagram (b) Decision tree

Given these assessments, the Defender can solve her decision problem working backwards the tree in Figure 5. At node D_2, she can compute her maximum utility action $d_2^*(d_1, s)$ for each $(d_1, s) \in \mathcal{D}_1 \times S$ as in (5). Afterwards, she will obtain at node S her expected utility $\psi_D(d_1, a)$ for each $(d_1, a) \in \mathcal{D}_1 \times A$ as in (6). At this point, she can use her probabilistic assessment about what the Attacker will do, $p_D(A \mid d_1)$, to compute her expected utility at node A for each $d_1 \in \mathcal{D}_1$,

$$\psi_D(d_1) = \int \psi_D(d_1, a)\, p_D(A = a \mid d_1)\, \text{d}a.$$

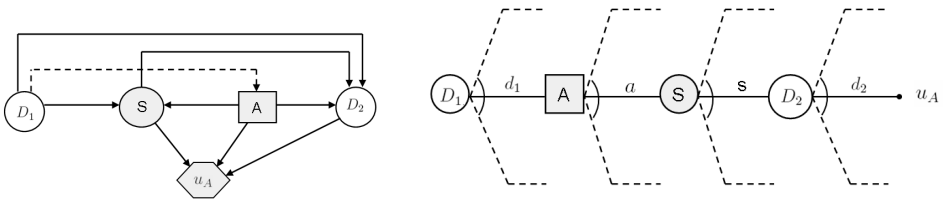
Finally, she can find her maximum expected utility decision at node D_1

$$d_1^* = \text{argmax}_{d_1 \in \mathcal{D}_1} \psi_D(d_1).$$

Based on this approach, the Defender's best strategy is to choose first $d_1^* \in \mathcal{D}_1$ at node D_1, and later, after observing $s \in S$, choose $d_2^*(d_1^*, s) \in \mathcal{D}_2$ at node D_2 .

Thus, the key issue is the assessment of $p_D(A \mid d_1)$. To do so, the Defender could use a standard statistical elicitation method if historical data on the Attacker's behavior in prior similar situations is available, together with judgements incorporating expert opinion. However, we propose to model the Defender's uncertainty about the Attacker's decision assuming he is an expected utility maximizer and taking into account that the Defender's uncertainty stems from her uncertainty about the Attacker's probabilities and utilities associated with his decision problem: the assessment of $p_D(A \mid d_1)$ is related with the analysis of the Attacker's decision problem as seen by the Defender, shown in Figure 6. The assessment of the Attacker's probabilities and utilities from the Defender's perspective will be based on all the information available to her, which may include both hard probabilistic data and expert opinions. An advantage of structuring the Attacker's problem is that the Defender can isolate different uncertainty and value parts of his problem and accommodate different expertise levels. Again, should this kind of information not be available to the Defender, she could use a noninformative distribution to describe $p_D(A \mid d_1)$.

Figure 6: The Defender's view of the Attacker's decision problem



(a) Influence diagram (b) Decision tree

Therefore, to elicit $p_D(A \mid d_1)$, the Defender needs to assess $u_A(a, s, d_2)$, $p_A(S \mid d_1, a)$ as well as $p_A(D_2 \mid d_1, a, s)$. In general, she will not know these true quantities, but she may acknowledge her uncertainty about them through a probability distribution $F = (U_A(a, s, d_2), P_A(S \mid d_1, a), P_A(D_2 \mid d_1, a, s))$ and solve the perceived Attacker's decision problem using backward induction over the decision tree in Figure 6 as follows.

- At chance node D_2, compute

$$(d_1, a, s) \rightarrow \Psi_A(d_1, a, s) = \int U_A(a, s, d_2)\, P_A(D_2 = d_2 \mid d_1, a, s)\, \mathrm{d}d_2.$$

- At chance node S, compute

$$(d_1, a) \rightarrow \Psi_A(d_1, a) = \int \Psi_A(d_1, a, s)\ P_A(S = s \mid d_1, a)\ \mathrm{d}s.$$

- At decision node A, solve

$$d_1 \rightarrow A^*(d_1) = \mathrm{argmax}_{a \in \mathcal{A}} \Psi_A(d_1, a).$$

Then, the Defender's predictive density $p_D(A \mid d_1)$ over attacks, conditional on her first defense decision, is given by

$$\int_0^a p_D(A = x \mid d_1)\mathrm{d}x = \Pr(A^*(d_1) \leq a).$$

This distribution could be approximated by Monte Carlo as follows

1. For $i = 1, ..., n$, repeat

   Draw

$$\left(u_A^i(a, s, d_2), p_A^i(S \mid d_1, a), p_A^i(D_2 \mid d_1, a, s)\right) \sim F$$

   At chance node D_2, compute

$$(d_1, a, s) \rightarrow \psi_A^i(d_1, a, s) = \int u_A^i(a, s, d_2)\ p_A^i(D_2 = d_2 \mid d_1, a, s)\ \mathrm{d}d_2$$

   At chance node S, compute

$$(d_1, a) \rightarrow \psi_A^i(d_1, a) = \int \psi_A^i(d_1, a, s)\ p_A^i(S = s \mid d_1, a)\ \mathrm{d}s$$

   At decision node A, compute

$$d_1 \rightarrow a_i^*(d_1) = \mathtt{argmax}_{a \in \mathcal{A}} \psi_A^i(d_1, a)$$

2. For each a, approximate

$$\int_0^a p_D(A = x \mid d_1)\mathrm{d}x \approx \#\{a_i^*(d_1) \leq a\}/n.$$

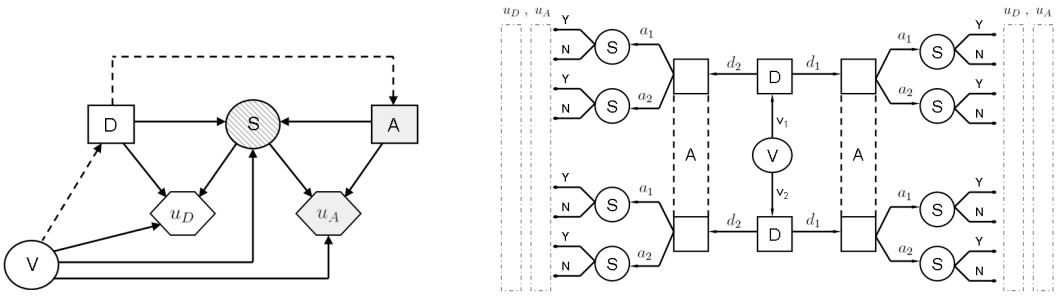
We have seen how the assessment of $p_D(A \mid d_1)$ is straightforward after the Defender's elicitation of F. However, the assessment of $P_A(D_2 \mid d_1, a, s)$ in F could be problematic, as the Defender may want to exploit information available to her about how the Attacker is analyzing her decision problem. Of course, if there is no information that the Defender can use, she will put a noninformative distribution over $P_A(D_2 \mid d_1, a, s)$. The Defender may continue this recursive analysis to model the thinking-about-what-the-other-is-thinking-about levels of analysis, until eventually she has no more information to analyze the next level of the hierarchy of recursive decision models, much as described in Section 2.2. The recursive analysis will always stop at some point, maybe after some simplification leading to an heuristic distribution for modeling an adversary's thinking at some step of the recursive analysis, as illustrated in Rios Insua et al. (2009) for an auction problem.

4 Sequential Defend-Attack Models with Private Information

We end up with the sequential Defend-Attack model with Defender's private information, i.e. which she does not want the Attacker to know. This is the case when e.g. the Defender wants to keep secrecy about vulnerabilities of sites she is trying to protect, as this information can be used by the Attacker to increase the chances of success and impact of an attack. In this model, the Defender moves first by choosing a defense and, then, having observed it, the Attacker moves by choosing an attack. An example of this situation is given by an Attacker who gets to observe how the Defender allocates her resources among the sites she wants to protect before deciding his attack. Note how the Defender decision allocating resources to protect different sites might signal to the Attacker about the sites' vulnerability and importance for the Defender, which is precisely the kind of information she wants to keep secret. This kind of applications, with the corresponding standard game theoretic analysis, has been considered by Powell (2007).

Assume that the Defender and the Attacker have, respectively, sets $\mathcal{D}$ and $\mathcal{A}$ of possible defenses and attacks. Again, we shall also assume that the level of success S of an attack is uncertain. The private information (e.g. vulnerabilities) is represented by V, which value is known by the Defender, but not by the Attacker. This affects the chances of success of an attack, as well as its impact. Finally, for both adversaries, the consequences depend, in addition, on the success of this attack and their own action.

Figure 7: The sequential Defend-Attack model with Defender's private information v



(a) Influence diagram

(b) Game tree

Figure 7 depicts the problem graphically. The coupled influence diagrams show explicitly that the uncertainty associated with the success of an attack S is probabilistically dependent on the actions of both the Attacker and the Defender, as well as on v: $S \mid d, a, v$. For example, if v represents a site's vulnerability, this probability will be higher as vulnerability gets higher, the rest of factors staying the same. The utility functions over the consequences for the Defender

and the Attacker are, respectively, $u_D(d, s, v)$ and $u_A(a, s, v)$, reflecting that the consequences are dependent on $V = v$. The arc in the influence diagram from the the Defender's decision node to the Attacker's reflects that the Defender's choice is observed by the Attacker. The arc from (V) to [D] reflects that v is known by the Defender at the time she makes her decision. The lack of arc from (V) to [A] indicates that v is not known by the Attacker at the time he makes his decision.

We also show the corresponding game tree, with only two actions per adversary: $\mathcal{D} = \{d_1, d_2\}$ and $\mathcal{A} = \{a_1, a_2\}$; the possible outcomes of an attack being only success or failure: $S \in \{Y, N\}$; and two possible values for V: v_1 and v_2. We reflect the sequential nature and asymmetric information for the problem. The fact that the Attacker does not know what is the value v at the time he must move, is displayed using information sets (drawn as dashed lines), a standard element of games with imperfect information.

4.1 The Standard Game Theoretic Analysis

We briefly describe how standard game theory solves this model with private and asymmetric information. This is an example of a signalling game, see Aliprantis and Chakrabarti (2000) for details. The game-theoretic approach requires the probability assessment over S, conditional on (d, a, v). As the Defender and the Attacker may have different assessments, these will be represented by $p_D(S|d, a, v)$ and $p_A(S|d, a, v)$. We assume that their utility functions are $u_A(a, s, v)$ and $u_D(d, s, v)$. The Attacker's prior beliefs about the Defender's private information V are represented through the probability distribution $\pi_A(v)$. All these probabilities and utilities are common knowledge. Then, the solution proceeds as follows.

Define, first, the strategy functions for each player. As the Defender observes the value of V, her strategy function is $v \to d(v) \in \mathcal{D}$. Because the Attacker makes his decision knowing the Defender's, his strategy function is $d \to a(d) \in \mathcal{A}$. We compute the expected utilities of both players at node (S) of the tree in Figure 7, when the decisions are $(d, a) \in \mathcal{D} \times \mathcal{A}$ and $V = v$:

$$\psi_D(d, a, v) = \int u_D(d, s, v)\, p_D(S = s \mid d, a, v)\, \mathrm{d}s \tag{7}$$

$$\psi_A(d, a, v) = \int u_A(a, s, v)\, p_A(S = s \mid d, a, v)\, \mathrm{d}s. \tag{8}$$

The Attacker's best response against a defense d is

$$a^*(d) = \mathrm{argmax}_{a \in \mathcal{A}} \int \psi_A(d, a, v)\, \pi_A(V = v \mid d)\, \mathrm{d}v,\ \forall d \in \mathcal{D}, \tag{9}$$

where $\pi_A(v \mid d)$ represents the Attackers' beliefs about the Defender's private information after observing her defense action. We will show how to determine this probability distribution below,

but before we need to solve the problem as if it would be known. Thus, under the assumption that the Defender knows how the Attacker will solve his problem, the Defender's maximum expected utility decision, given that she knows the value of $V = v$, is

$$d^*(v) = \text{argmax}_{d \in \mathcal{D}} \, \psi_D(d, a^*(d), v).$$

As it is standard in game theory, we allow for randomized strategies. Assuming sets $\mathcal{D}$ and $\mathcal{A}$ are continuous, we define

$$\Pi_{\mathcal{D}} = \left\{ \pi : \ \pi(d) \geq 0 \ \forall d \in \mathcal{D} \text{ and } \int_{\mathcal{D}} \pi(d) \mathrm{d}d = 1 \right\}$$

and

$$\Pi_{\mathcal{A}} = \left\{ \pi : \ \pi(a) \geq 0 \ \forall a \in \mathcal{A} \text{ and } \int_{\mathcal{A}} \pi(a) \mathrm{d}a = 1 \right\}$$

as their associated sets of mixed strategies. Hence, $d^*(v)$ and $a^*(d)$ have associated probability distributions $\pi_{d^*(v)}(d \mid v) \in \Pi_{\mathcal{D}}$ and $\pi_{a^*(d)}(a \mid d) \in \Pi_{\mathcal{A}}$, respectively.

We now show how probability distribution $\pi_{d^*(v)}(d \mid v)$ is related to $\pi_A(v \mid d)$. Under the assumption that the Attacker knows how the Defender will solve her problem, he can update his prior knowledge about V after observing a defense d, by using Bayes' rule:

$$\pi_A(v \mid d) \propto \pi_A(v) \, \pi_{d^*(v)}(d \mid v),$$

which is the probability distribution needed to compute (9).

A game theoretic solution can be determined, then, by finding a pair of strategies $(d^*(v), a^*(d))$ which are a fixed point solution of the equations

$$\begin{cases} \pi_{d^*(v)} = \text{argmax}_{\pi \in \Pi_{\mathcal{D}}} \displaystyle\int_{\mathcal{D}} \left[\int_{\mathcal{A}} \psi_D(d, a, v) \, \pi_{a^*(d)}(a \mid d) \, \mathrm{d}a \right] \pi(d) \, \mathrm{d}d \qquad \forall v \in V \\ \pi_{a^*(d)} = \text{argmax}_{\pi \in \Pi_{\mathcal{A}}} \displaystyle\int_{\mathcal{A}} \left[\int_{V} \psi_A(d, a, v) \, \pi_A(v) \, \pi_{d^*(v)}(d \mid v) \, \mathrm{d}v \right] \pi(a) \, \mathrm{d}a \qquad \forall d \in \mathcal{D} \end{cases}$$

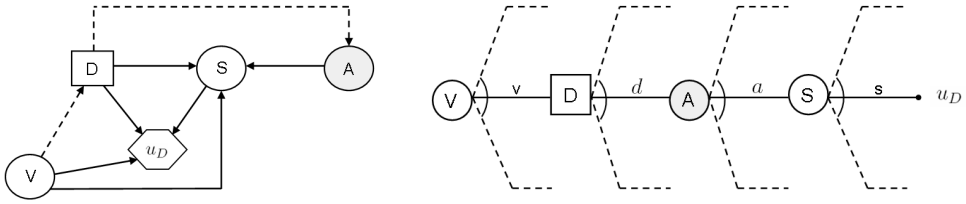
A fixed point solution of the above equations is a Nash equilibrium. However, in general, not all Nash equilibria for the sequential Defend-Attack model with Defender's private information need to be a solution of it. Thus, this is a refinement of the concept of Nash equilibria. Note also that, in general, a Nash equilibrium can be defined using any Attacker's beliefs over v consistent with the equilibrium. In addition, we have imposed a rational learning behavior on the Attacker based on Bayes' rule.

4.2 The ARA Analysis

Following a more realistic approach, we weaken the common knowledge assumption. We thus consider the Defender's decision problem as a standard decision analysis problem, illustrated

in Figure 8, with the Attacker's decision node perceived as a random variable. Similarly, her decision tree denotes uncertainty about the Attacker's decision by replacing [A] with (A), and including a reference only to her utility function.

Figure 8: The Defender's decision problem



(a) Influence diagram (b) Decision tree

By observing the influence diagram, note that in order to solve her decision problem, the Defender has already assessed $p_D(S|d,a,v)$ and $u_D(d,s,v)$. She also needs $p_D(A|d)$, which is her assessment of the probability that the Attacker will choose attack $A = a$, after observing that the Defender has chosen defense d. This assessment requires the Defender to analyze the problem from the Attacker's perspective. As we shall see, although the Defender does not actually know (p_A, u_A), she has beliefs about them which will be relevant in her analysis of the Attacker's decision problem. After assessing $p_D(A|d)$, she can obtain her maximum expected utility defense by solving the tree in Figure 8 using backwards induction as follows:

- At chance node S, compute $\psi_D(d,a,v)$ for each (d,a,v) as in (7).
- At chance node A, compute

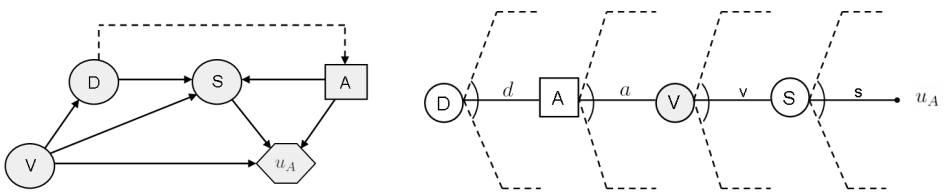
$$(d,v) \rightarrow \psi_D(d,v) = \int \psi_D(d,a,v)\, p_D(A = a \mid d)\, \mathrm{d}a. \tag{10}$$

- At decision node D, solve

$$v \rightarrow d^*(v) = \mathrm{argmax}_{d \in \mathcal{D}} \psi_D(d,v). \tag{11}$$

To assess $p_D(A|d)$, the Defender must place herself in the Attacker's shoes, and solve his decision problem from her perspective. Figure 9 represents the Attacker's problem, as seen by the Defender. Note that the Defender's decision node is represented as a random variable in the Attacker's analysis, as it is not under his control. The arrow from (D) to [A] indicates that the Defender's decision will be known at the time he has to decide. As the Attacker does not know the value of the Defender's private information v, his uncertainty is represented through a

Figure 9: The Defender's analysis of the Attacker's decision



(a) Influence diagram (b) Decision tree

probability distribution: $p_A(V)$, representing the Attacker's (prior) beliefs about the Defender's private information. We assume that the Defender analyzes the Attacker's decision problem considering that he is an expected utility maximizer and that he uses the Bayes rule to learn about the Defender's private information from the observation of her defense decision. Thus, the arrow from (V) to (D), which represents probabilistic dependence, can be inverted using Bayes rule, to obtain the Attacker's (posterior) beliefs about v: $p_A(V|D = d)$. However, in order to apply Bayes rule, we need to assess $p_A(D|v)$ first.

Should the Defender know the Attacker's utility function $u_A(a, s, v)$ and his probabilities $p_A(S|d, a, v)$ and $p_A(V|d)$, she would be able to anticipate his decision $a^*(d)$ by solving backwards the tree in Figure 9 and computing his expected utility ψ_A as follows:

- At chance node S, compute $\psi_A(d, a, v)$ for each (d, a, v) as in (8).
- At chance node V, compute for each (d, a)

$$\psi_A(d, a) = \int \psi_A(d, a, v)\ p_A(V = v \mid d)\ \mathrm{d}v. \tag{12}$$

- At decision node A,

$$v \to a^*(d) = \mathrm{argmax}_{a \in \mathcal{A}} \psi_A(d, a).$$

However, the Defender does not know the true (p_A, u_A): instead of using point estimates for p_A and u_A to find an estimate of the Attacker's optimal decision $a^*(d)$, the Defender's uncertainty about the Attacker's decision should derive from her uncertainty about the Attacker's $(P_A, U_A) \sim F$. This distribution will induce distributions $\Psi_A(d, a, v)$ and $\Psi_A(d, a)$ on the Attacker's expected utilities defined in (8) and (12), so that, respectively,

$$\Psi_A(d, a, v) = \int U_A(a, s, v)\ P_A(S = s \mid d, a, v)\ \mathrm{d}s$$

and

$$\Psi_A(d, a) = \int \Psi_A(d, a, v)\ P_A(V = v \mid d)\ \mathrm{d}v$$

for $(P_A, U_A) \sim F$. Then, assuming the Attacker is an expected utility maximizer, the Defender's predictive distribution about the Attacker's response to her defense choice d is defined through

$$p_D(A = a|d) = \mathbb{P}_F[a = \text{argmax}_{x \in \mathcal{A}} \Psi_A(d, x)], \quad \forall a \in \mathcal{A}.$$

The Defender may use Monte Carlo simulation to approximate $p_D(A|d)$ by drawing n samples $\left\{\left(p_A^i, u_A^i\right)\right\}_{i=1}^n$ from F, which produce $\{\psi_A^i\}_{i=1}^n \sim \Psi_A$, and approximating $p_D(A = a|d)$ by

$$\hat{p}_D(A = a|d) = \#\{a_i^*(d) = a\}/n, \quad \forall a \in \mathcal{A},$$

when $A \mid d$ is discrete, or

$$\hat{p}_D(A \leq a|d) = \#\{a_i^*(d) \leq a\}/n, \quad \forall a \in \mathcal{A},$$

when $A \mid d$ is absolutely continuous.

Thus, the elicitation of $F = (\ P_A(S|d, a, v),\ U_A(a, s, v),\ P_A(V|d)\ )$ allows the Defender to solve her problem of assessing $p_D(A|d)$. Although the Defender may have enough information and judgment available to her to directly assess the Attacker's probabilities $P_A(S|d, a, v)$ and utilities $U_A(a, s, v)$, the assessment of $P_A(V|d)$ requires of a deeper analysis as it has a strategic component. Specifically, given that the Attacker has prior knowledge over V, modeled through $p_A(V)$, and assuming his learning follows Bayes' rule, we can expect that his posterior beliefs about V, after he observes $D = d$, become:

$$p_A(V = v|d) \propto p_A(V = v)\ p_A(D = d|v),$$

where $p_A(D = d|v)$ models the Attacker's probabilistic assessment of what defense she would choose conditional on each possible value of her private information. The elicitation of this probabilistic model requires an analysis of how the Attacker analyzes the Defender's decision problem. Assuming he thinks that she is an expected utility maximizer, and that the decision problem she tries to solve is as in Figure 8, the Defender's elicitation of a probability distributions $G = (\ U_D(d, s, v),\ P_D(S|d, a, v),\ P_D(A^1|d)\ )$ representing the Attacker's assessments of her utilities and probabilities, allows her to solve her problem of assessing $p_A(D|v)$ by solving the tree in Figure 8 as follows:

- At chance node S, compute for each (d, a, v)

$$\Psi_D(d, a, v) = \int U_D(d, s, v)\ P_D(S = s \mid d, a, v)\ \mathrm{d}s.$$

- At chance node A, compute

$$(d, v) \rightarrow \Psi_D(d, v) = \int \Psi_D(d, a, v)\ P_D(A^1 = a \mid d)\ \mathrm{d}a.$$

- At decision node D, solve

$$v \rightarrow p_A(D = d|v) = \mathbb{P}_G[d = \mathrm{argmax}_{x \in \mathcal{D}} \Psi_D(x, v)], \quad \forall d \in \mathcal{D}.$$

Note that $p_A(V)$ represents the Attacker's prior knowledge about her private information. As the Defender does not have access to this distribution representing the Attacker's beliefs, we will directly elicit it from the Defender's perspective: $P_A(V)$ represents what she believes to be $p_A(V)$, with the probabilistic model P_A acknowledging her confidence on her assessment of p_A. Thus, we have

$$P_A(V = v|d) \propto P_A(V = v)\ p_A(D = d|v).$$

The only difficulty for the Defender at this step is her assessment of what she thinks to be the Attacker's assessment of the probability model used by her to predict his attack as a response to her chosen defense: $P_D(A^1|d)$ in G. This distribution is necessary to compute $p_A(D|v)$. Thus, we may go deeper in the hierarchy of nested decision models and try to support the Defender in the assessment of $P_D(A^1|d)$ through the analysis of how the Attacker, in his analysis of her decision problem, thinks the Defender will analyze his decision problem. However, if no information is available at this level, we can always end the hierarchy of analysis with a reference distribution over $P_D(A^1|d)$. This would allow the computation of a recommendation for action to the Defender. Clearly, should this recommendation be sensitive to the reference distribution, this would indicate that there is still relevant information that needs to be elicited before reading a robust enough recommendation. Thus, in such case, it would be desirable to collect more data and/or judgement through intelligence.

5 Discussion

We have provided an account of how the framework of Adversarial Risk Analysis can support a Defender against an intelligent adversary, the Attacker: the Defender assesses the probabilities of the adversaries' actions before computing her maximum expected utility defenses. We have assumed that the Attacker is an expected utility maximizer, and that the Defender's uncertainty about the Attacker' decision stems from her uncertainty about his decision analysis, specifically his probabilities and utilities.

One could possibly question the hypothesis of the adversary being rational, but the recent psychology of terrorism literature tends to support such hypothesis, see Schaefer (2006) for a gentle introduction, in the sense that terrorists will tend to use their limited offensive resources to cause significant damage with a high probability of success. Note that we could assume other optimizing models within terrorists, but our arguments could be easily translated.

The models we have discussed are relatively simple, but retain the essence of counterterrorism. Real problems are much more complex. For example, they involve hundreds of possible decisions, many more uncertainties including those associated with the goals and resources of the terrorists, and more complex dynamic interactions which would require more complex analysis. But we expect the methodology to stay essentially the same. Indeed, we would expect to deploy more complex coupled influence diagrams with time partitioned in sequences of defend-attack-defend moves, periods of simultaneous defend-attack moves and periods of sequential defend-attack moves with private information. Thus, we view the three models treated here as basic model building blocks for complex problems. Note also that we have emphasized discrete models, paying little attention to the numerical intricacies associated with the need to optimize resources at the decision nodes.

Extensions of the methodology to the case in which there are more than one Attacker, and more than one defender need to be explored. In this case, we would expect negotiations among the defenders to share risks, possibly as described in Rios and Rios Insua (2010).

Finally, we note that the ARA framework might find applications in other contexts. Areas such as marketing and cybersecurity seem relevant. Note that in this case we might be facing a large and uncertain number of attackers. Our earlier Rios Insua et al. (2009) referred to some simple auctions. This is in line with the recent debate between using decision analysis or game theory models for competing situations, well reflected in papers such as Rothkopf (2007) or van Bingsbergen and Marx (2007).

References

Arce, D. and T. Sandler (2007) Terrorist signalling and the value of intelligence, *British Journal Political Science*, 37, 573-586.

Aliprantis, C. and S. C hakrabarti (2000) *Games and Decision Making*, Oxford U.P.

Banks, D. and S. Anderson (2006) Game theory and risk analysis in the context of the smallpox threat, in *Statistical Methods in Counterterrorism*, A. Wilson, G. Wilson and D. Olwell,

9-22.

Bier, V. and N. Azaiez (2009) *Game Theoretic Risk Analysis of Security Threats*, Springer.

Brown, G., M. Carlyle, J. Salmeron and K. Wood (2005) Analyzing the vulnerability of critical infrastructure to attack and planning defenses, *Tutorials in Operations Research*, INFORMS, 102-123.

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

Brown, G., W.M. Carlyle, and R. Wood (2008). "Optimizing Department of Homeland Security Defense Investments: Applying Defender-Attacker(-Defender) Optimization to Terror Risk Assessment and Mitigation," Appendix E, National Academies Press, Washington, D.C.

English, R. (2009) *Terrorism: How to Respond*, Oxford University Press.

French, S. and D. Rios Insua (2000) *Statistical Decision Theory*, Arnold.

Gibbons, R. (1992). *A Primer in Game Theory*, Pearson Education Ltd.:Harlow.

Gutfraind, A. (2009) Terrorism as a mathematical problem, *SIAM News*, 10, 12.

Harsanyi, J. (1967) Games with incomplete information played by Bayesian players, I-III. Part I. The basic model. *Management Science*, 14, 3, 159-182.

Harsanyi, J. (1982) Subjective probability and the theory of games: Comments on Kadane and Larkey's paper, *Management Science*, 28, 2, 120-124.

Hausken, K. (2002) Probabilistic risk analysis and game theory, *Risk Analysis*, 22, 17-27.

Heal, G. and H. Kunreuther (2006) You can only die once: interdependent security in an uncertain world, in *The Economic Impacts of Terrorist Attacks*, edited by H.W. Richardson, P. Gordon, and J.E. Moore III, Northampton, MA: Edward Elgar Publishers.

Kadane, J.B. and P.D. Larkey (1982) Subjective probability and the theory of games, *Management Science*, 28, 2, 113-120; reply: 124.

Kardes, E. (2005) Robust stochastic games and applications to counter-terrorism strategies, *CREATE report.*

Parnell, G., D. Banks, L. Borio, G. Brown, L. A. Cox, J. Gannon, E. Harvill, H. Kunreuther, S. Morse, M. Pappaioanou, S. Pollack, N. Singpurwalla, and A. Wilson (2008). *Report on Methodological Improvements to the Department of Homeland Security's Biological Agent Risk Analysis*, National Academies Press.

Parnell, G., Smith, C. and D. Moxley (2010) Intelligent adversary risk analysis: a bioterrorism risk management model, *Risk Analysis*, 30, 1, 32-48.

Pate-Cornell, E. and S. Guikema (2002) Probabilistic modeling or terrorist threats: a systematic analysis approach to setting priorities among countermeasures, *Military Operations Research*, 7, 5-23.

Pinker, E.J. (2007) An analysis of short-term responses to threats of terrorism, *Management Science*, 53, 6, 865-880.

Powell, R. (2007) Allocating defensive resources with private information about vulnerability, *American Political Science Review*, 101, 799-809.

Raiffa, H. (1982) *The Art and Science of Negotiation*, Harvard University Press: Cambridge, Massachusetts.

Raiffa, H, Richardson, J., Metcalfe, D. (2002) *Negotiation Analysis*, Harvard University Press: Cambridge, Massachusetts.

Rios, J. and D. Rios Insua (2010) Balanced increment and concession methods for negotiation support, *Journal of the Spanish Royal Academy of Sciences (RACSAM)*, 104, 41-56.

Rios Insua, D., J. Rios, D. Banks (2009) Adversarial risk analysis, *Journal of the American Statistical Association*, 104, 486, 841-854.

Rios Insua, D. and F. Ruggeri (2000) *Robust Bayesian Analysis*, Springer.

Rothkopf, M. (2007) Decision Analysis: The right tool for auctions, *Decision Analysis*, 4, 167-172.

Schaefer, A. (2006) Inside the Terrorist Mind, *Mind*, 18, 6, 72-79.

Siqueira, K. and T. Sandler (2006) Terrorists vs the Government: Strategic intervention, support and sponsorship, *Journal of Conflict Resolution*, 50, 878-898.

van Bingsbergen, J.H., Marx, L.M. (2007) Exploring relations between decision analysis and game theory, *Decision Analysis*, 4, 32-40.

von Winterfeldt, D. and T.M. O'Sullivan (2006) Should we protect commercial airplanes against surface-to-air missile attacks by terrorists? *Decision Analysis*, 3, 2, 63-75.

Wein, L. (2009) Homeland security: from mathematical models to policy implementation, *Operations Research*, 57, 801-811.

Zhuang, J. and V. Bier (2007) Balancing Terrorism and Natural Disasters. Defensive Strategy with endogenous attack effort, *Operations Research*, 55, 5, 976-991.