Adversarial Risk Analysis for Counterterrorism Modeling

Jesus Rios[1] and David Rios Insua[2]

Recent large scale terrorist attacks have raised interest in models for resource allocation against terrorist threats. The unifying theme in this area is the need to develop methods for the analysis of allocation decisions when risks stem from the intentional actions of intelligent adversaries. Most approaches to these problems have a game theoretic flavor although there are also several interesting decision analytic based proposals. One of them is the recently introduced framework for adversarial risk analysis, which deals with decision making problems that involve intelligent opponents and uncertain outcomes. We explore how adversarial risk analysis addresses some standard counterterrorism models: simultaneous defend-attack models, sequential defend-attack-defend models and sequential defend-attack models with private information. For each model, we first assess critically what would be a typical game theoretic approach and then provide the corresponding solution proposed by the adversarial risk analysis framework, emphasizing how to coherently assess a predictive probability model of the adversary's actions, in a context in which we aim at supporting decisions of a defender versus an attacker. This illustrates the application of adversarial risk analysis to basic counterterrorism models that may be used as basic building blocks for more complex risk analysis of counterterrorism problems.

KEY WORDS: adversarial risk analysis, counterterrorism resource allocation, defender-attacker models, intelligent adversary, elicitation of attack probabilities

1. INTRODUCTION

Appropriate responses to terrorism represent one of the key challenges for states in this century[1,2]. Indeed, after recent 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. 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. Recent accounts of various techniques and applications may be seen in Ezell et al.[3], Gutfraind[4] and Wein[5]. Parnell et al.[6] and Enders and Sandler[7] provide outstanding overviews 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 and interdependent. Thus, it is no wonder that much of this research has reminiscent game theoretic and risk analytic flavors. The role of standard risk analysis in the management of terrorism risks has been discussed in Deisler[8]. Also, Garrick[9] points out some of the challenges associated with extending standard risk assessment methods for the analysis of threats from terrorist acts. Dillon et al.[10] describe a decision

[1]Business Analytics and Mathematical Sciences, IBM T.J. Watson Research Center, Yorktown Heights, NY 10598, USA, jriosal@us.ibm.com
[2]Royal Academy of Sciences, Madrid, Spain, david.rios@urjc.es

0272-4332//0100-0001$22.00/1

making framework based on risk analysis principles for allocating anti-terrorism resources using risk scores. Ezell et al.[3] defend the use of traditional probabilistic risk assessment methods like event trees to estimate terrorism risks. These approaches, based on the direct application of conventional risk analysis methods to terrorism risk management, have been criticized by Cox[11] and Brown and Cox[12], who warn about the inappropriateness of modeling the actions from terrorists in essentially in the same way that random adverse events in natural or engineered systems.

On the other hand, there is a rich literature in political sciences and economics regarding game theory and terrorism, though it places little emphasis on risk analysis aspects, see e.g. Siqueira and Sandler[13], Arce and Sandler[14], and Powell[15]. Recent relevant references with a game-theoretic flavor in the OR/MS literature, include Zhuang and Bier[16], 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[17,18,19], who present bilevel (max-min, min-max) and trilevel (min-max-min) optimization models for three stylized counterterrorism models such as defender-attacker, attacker-defender and defender-attacker-defender problems. Kardes[20] surveys various approaches to strategic decision making in the 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. The book edited by Bier and Azaiez[21] contain many papers on the attacker-defender model and several variants and applications. Insights combining risk analysis and game theory can be found in Hausken[22] and Cox[23].

A thread in the above game theoretic approaches is the common knowledge assumption, criticized, e.g. in Raiffa et al.[24]. 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 opponents. 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 are 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 is aggravated as participants try to conceal information.

The other mainstream literature in the field has a decision analytic flavor. Among others, Pinker[25] uses qualitative influence diagrams to assess short and long-term deployment of countermeasures; Merrick and McLay[26] use decision trees and influence diagrams from the point of view of the defender to model the decision of installing radiation sensors to screen cargo containers against terrorist radiological threats; and Parnell et al.[27] describe canonical terrorist multi-objective decision trees and influence diagrams to evaluate bioterrorist threats. Their recurrent issue is the difficulty in assessing the probabilities over the actions of the adversaries, which is the key objection, see Harsanyi[28], to the Bayesian approach to games, introduced by Kadane and Larkey[29] and Raiffa[30,24]. Banks and Anderson[31] provide a numerical comparison of classical and Bayesian approaches to games within a smallpox attack problem.

Paté-Cornell and Guikema[32] 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, based on point estimates of the Attacker's probabilities and utilities. Then, they 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. This approach 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[33], taking into account our imprecision about the likely actions of our adversary. This is the venue adopted by von Winterfeldt and O'Sullivan[34] within a simple decision tree to evaluate Man-Portable Air Defense Systems countermeasures. This may be too involved computationally in complex problems.

We introduced[35] the framework of Adversarial Risk Analysis (ARA) to cope with the risk analysis of situations in which risks stem from the deliberate actions of intelligent adversaries. The main application in that paper was geared towards auctions. ARA lies somewhat between both approaches mentioned above, with a Bayesian game theoretic flavor. In

supporting one of the participants, the problem is viewed as a decision analytic one, but principled procedures which employ the adversarial structure, and other information available, are used to assess probabilities on the opponents' actions. In order to obtain a (probabilistic) descriptive model of how the opponents will behave, incorporating an analysis of how they think about their decision problem, we assume that opponents are expected utility maximizers, but see §5 for a discussion of this assumption. The uncertainty in adversaries' actions stems from the uncertainty about their utilities and probabilities when used to analyze their decision making problems. The potentially infinite analysis of nested decision models arrived at when using ARA is avoided in the game theoretic approach by using the common (prior) knowledge assumption. But, as mentioned, this is at the cost of a strong unrealistic assumption, which would invalidate the analysis when it comes to most counterterrorism applications. We prefer to be realistic and accommodate as much information as we can from intelligence into the analysis, through a structure of nested decision models, and stop when no more information can be accommodated. We then conclude the recursion of decision models with a noninformative probability distribution, which needs to pass some sensitivity analysis test before the whole analysis is considered appropriate.

In this paper, we show the relevance of ARA in supporting one of the participants, which we shall call the Defender, when analyzing three important stylized counterterrorism resource allocation 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 may 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 considerable detail in the literature from a standard game theoretic perspective.

For each of these three models, we first describe the basic problem through coupled influence diagrams and game trees. We then assess critically what would be the typical game theoretic approach for such problem based on a Nash equilibrium concept, or some refinement thereof. We then provide the corresponding ARA solution, emphasizing 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. Simple examples illustrate some of the assessments and computational intricacies for the ARA approach. However, our main interest here is in the novel conceptual framework, leaving aside computational and algorithmic issues for later work. We end up with some discussion and conclusions.

2. SIMULTANEOUS DEFEND-ATTACK MODELS

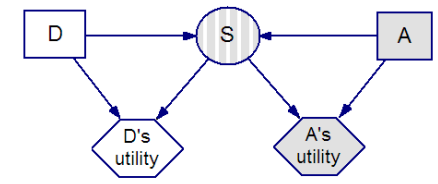
We start by discussing the simultaneous defend-attack model: a Defender (she, D) and an Attacker (he, A) decide their defense and attack, respectively, without knowing the action chosen by each other. See Zhuang and Bier[16] for a related discussion. As an example, imagine a case in which the FAA decides whether to introduce undercover marshals in an airplane that might, or not, be hijacked by terrorists.

We shall assume that the adversaries have discrete alternative sets $\mathcal{D} = \{d_1, d_2, ..., d_m\}$ and $\mathcal{A} = \{a_1, a_2, ..., a_k\}$ of defenses and attacks, respectively. We shall also assume that the only relevant uncertainty is S, denoting the success ($S = 1$) or failure ($S = 0$) of an attack. Each decision maker assesses differently the probability of the result of an attack, which depend on the defense and attack adopted: $p_D(S = s \mid d, a)$ and $p_A(S = s \mid d, a)$. The utility function of the Defender $u_D(d, s)$ depends on her chosen defense and the result of the attack. Similarly, the Attacker's utility function is $u_A(a, s)$. This situation can be represented by two coupled influence diagrams (one for the Defender, one for the Attacker) with a shared uncertainty node associated with the attack success, as in Fig. 1. We also show a game tree for this problem, with just two possible attacks and defenses, to simplify the figure.

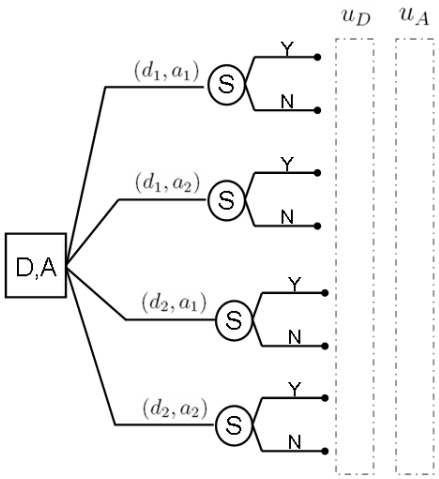
2.1 A Game Theoretic Analysis

Under the common knowledge assumption, preferences and beliefs from both the Defender and the Attacker, (u_D, p_D) and (u_A, p_A) respectively, 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

$$\begin{aligned}\psi_D(d, a) =& p_D(S = 0 \mid d, a)\, u_D(d, S = 0) + \\ & p_D(S = 1 \mid d, a)\, u_D(d, S = 1),\end{aligned}$$



(a) Influence diagram



(b) Game tree ($m = k = 2$)

Fig. 1. The simultaneous Defend-Attack model

and, similarly,

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

A Nash equilibrium (d^*, a^*) for this game would satisfy

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

Finding Nash equilibria may require the use of randomized strategies[36]. There could be several equilibria with no unambiguous criteria to further discern among them[24].

If utilities and probabilities are not common knowledge among the adversaries, a game-theoretic approach proceeds by modeling the game as one with incomplete information[37], introducing the notion of player types: each player will be of a certain type which is known to him but not to his opponent. Thus, a player's type represents the private information he may have. The Defender's type $\tau_D \in T_D$ determines her utility $u_D(d, s, \tau_D)$ and probability $p_D(S = s \mid d, a, \tau_D)$. Similarly, for the Attacker's types $\tau_A \in T_A$. Harsanyi proposes the Bayes-Nash equilibrium as a solution concept, under a still strong common knowledge assumption: the adversaries' beliefs about the opponent's types are common knowledge and 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, $d : \tau_D \to d(\tau_D) \in \mathcal{D}$ 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

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

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

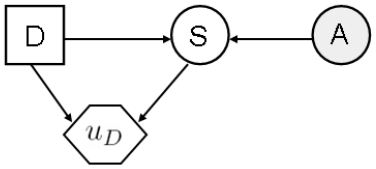
$$\psi_D(d^*(\tau_D), a^*, \tau_D) \geq \psi_D(d(\tau_D), a^*, \tau_D),\ \forall\, \tau_D \ \text{ and}$$
$$\psi_A(d^*, a^*(\tau_A), \tau_A) \geq \psi_A(d^*, a(\tau_A), \tau_A),\ \forall\, \tau_A$$

for every d and every a, respectively. Again randomized strategies might be required to possibly find an equilibria.

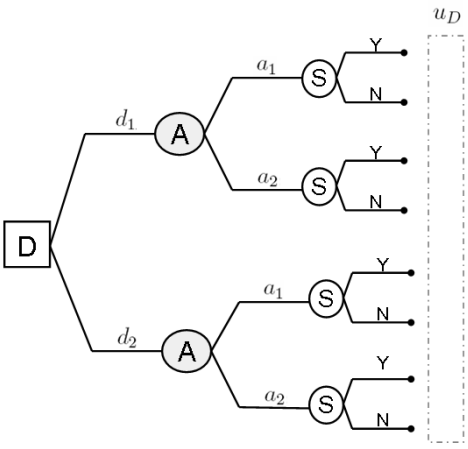
We believe that the underlying common prior (knowledge) assumption is still counterintuitive and unrealistic, specially in the context of counterterrorism: it implies that players need to disclose, inter alia, their true beliefs about their opponent's type, as well as their private probabilistic assessments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

2.2 The ARA Approach

More realistically, we weaken the common (prior) knowledge assumption. We assume that we support the Defender in solving the simultaneous Defend-Attack model. As reflected in Fig. 2, the Defender has to choose a defense $d \in \mathcal{D}$, whose consequences depend on the success of an attack $a \in \mathcal{A}$ simultaneously chosen by the Attacker, which is,

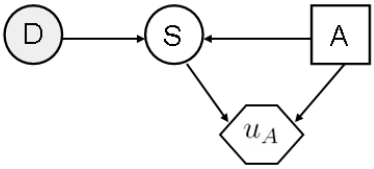


(a) Influence diagram

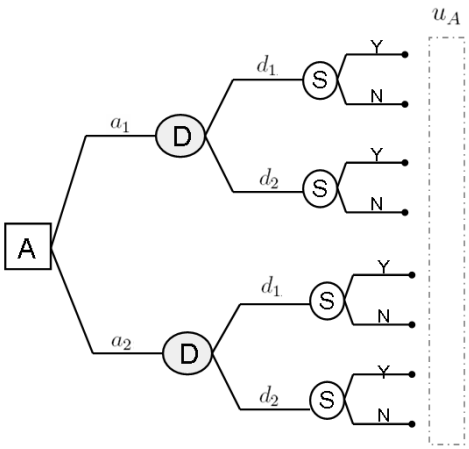


(b) Decision tree ($m = k = 2$)

Fig. 2. The Defender's decision analysis



(a) Influence diagram



(b) Decision tree ($m = k = 2$)

Fig. 3. The Attacker's decision analysis, as seen by the Defender

therefore, uncertain for the Defender at the time she makes her decision.

By standard decision theory, the Defender should maximize her expected utility[38]. 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 a at node A. She expresses her uncertainty through a probability distribution $\pi_D(A = a)$. Then, the optimization problem she should solve is

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

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

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

In general, the Defender will be uncertain about the Attacker's utility function and probabilities (u_A, p_A, π_A) required to solve such problem. Suppose that we 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, and this will aid us in assessing $\pi_D(A)$, mimicking the argument in (2), we propagate such uncertainty to compute the following probability distribution

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

Note now that (U_A, P_A) could be directly elicited from the Defender. However, the elicitation of $\Pi_A(D)$ may require further analysis leading to the next level of recursive thinking: the Defender would need to think about how the Attacker analyzes her problem. This is why we condition in (3) by (the distribution of) D. Note that $\Pi_A(D)$ 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 presumes that the Attacker thinks she is an expected utility maximizer trying to solve a decision problem like the one described in Fig. 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 the resolution of problem (1) from the Attacker's perspective. This reduces the assessment of $\Pi_A(D)$ to the computation of the distribution

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

assuming the Defender is 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 her decision problem. To assess the distribution (4), the Defender needs to elicit $(U_D, P_D) \sim G$, representing her probabilistic knowledge about how the Attacker may estimate her utility function $u_D(d,a)$ and her probability p_D over $S|d,a$, 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`

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

`with` $(U_A^i, P_A^i) \sim F^i$

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

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

`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 Figs. 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.

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 distribution[38].

We illustrate the ARA approach to this 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. We assume that we are able to assess from the Defender:

Table I . Defender's assessments

(a) $u_D(d, s)$

	$s=1$	$s=0$
d_1	0	80
d_2	10	100

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

	a_1	a_2
d_1	0.1	0
d_2	0.8	0

(c) $U_{A_I}(a, s)$

	$s=1$	$s=0$
a_1	$Tri(50, 100, 100)$	0
a_2	100	$Tri(0, 0, 50)$

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

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

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

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

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

	a_1	a_2
d_1	$Tri(0, 0, 0.5)$	0
d_2	$Tri(0.5, 1, 1)$	0

(g) $U_{D_I}(d, s)$

	$s=1$	$s=0$
d_1	$\mathcal{U}(0, 40)$	$\mathcal{U}(60, 100)$
d_2	$\mathcal{U}(0, 40)$	$\mathcal{U}(60, 100)$

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

	a_1	a_2
d_1	$\mathcal{U}(0, b)$	0
d_2	$b \sim \mathcal{U}(0, 1)$	0

Note: $Tri(min, mode, max)$ and $\mathcal{U}(min, max)$ stand, respectively, for triangular and uniform distributions.

- Her utility function $u_D(d, s)$, which incorporates the increase in security, the costs, as well as other possible consequences, and her probability distribution $p_D(S = s|d, a)$ associated with her decision problem (Fig. 2), shown in Tables I (a) and I (b) respectively.
- She considers that the terrorist 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 will face a decision problem as described in Fig. 3. The Defender assesses that the utilities and probabilities of a Class I Attacker in (3) are $(U_{A_I}, P_{A_I}) \sim F_I$, see Tables I (c) and I (d), and those of a Class II Attacker are $(U_{A_{II}}, P_{A_{II}}) \sim F_{II}$, see Tables I (e) and I (f).
- Based on the information available, the Defender thinks that a Class I Attacker is capable of analyzing her problem as in Fig. 2. She estimates that a Class I Attacker's beliefs about her utilities and probabilities in (4) are $(U_{D_I}, P_{D_I}) \sim G_I$, shown in Tables I (g) and I (h). 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 $\alpha = \pi_{A_I}(D_I = d_1) \times 10$. The Defender has no information to assess how a Class II Attacker would analyze her problem. However, she 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 over his choice of action: $\Pi_{D_I}(A_I^1 = a_1) \sim \mathcal{U}(0, 1)$.

To solve the Defender's decision problem, we need to assess $\pi_D(A = a_1)$, her predictive distribution about what the terrorists will do, where A is the mixture $0.8\, A_I + 0.2\, A_{II}$, with A_I representing the Defender's beliefs about what attack in $\mathcal{A} = \{a_1, a_2\}$ a Class I terrorist will choose, and similarly for A_{II}. Thus,

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

Based on (3) and (4), $\pi_D(A_I = a_1)$ could be estimated through Monte Carlo simulation as follows:

1. For $j = 1, \ldots, n$, repeat
   Draw $\pi^j_{D_I} \sim \Pi_{D_I} = \mathcal{U}(0,1)$.
   Draw $(u^j_{D_I}, p^j_{D_I}) \sim (U_{D_I}, P_{D_I}) = G_I$
   Compute

$$d^j_I = \arg\max_{d \in \mathcal{D}} \sum_{a \in \mathcal{A}} \left[\sum_{s \in \{0,1\}} u^j_{D_I}(d,s) \; p^j_{D_I}(S = s \mid d, a) \right] \pi^j_{D_I}(A^1_I = a)$$

2. Approximate $\pi_{A_I}(D_I = d_1)$ through
   $\hat{\pi}_{A_I}(D_I = d_1) = \#\{1 \le j \le n : \; d^j_I = d_1\}/n$
   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 $j = 1, \ldots, n$, repeat
   Draw $\hat{\pi}^j_{A_I} \sim \hat{\Pi}_{A_I}$
   Draw $(u^j_{A_I}, p^j_{A_I}) \sim (U_{A_I}, P_{A_I}) = F_I$
   Compute

$$a^j_I = \arg\max_{a \in \mathcal{A}} \sum_{d \in \mathcal{D}} \left[\sum_{s \in \{0,1\}} u^j_{A_I}(a,s) \; p^j_{A_I}(S = s \mid d, a) \right] \hat{\pi}^j_{A_I}(D_I = d)$$

4. Approximate $\pi_D(A_I = a_1)$ through
   $\hat{\pi}_D(A_I = a_1) = \#\{1 \le j \le n : \; a^j_I = a_1\}/n.$

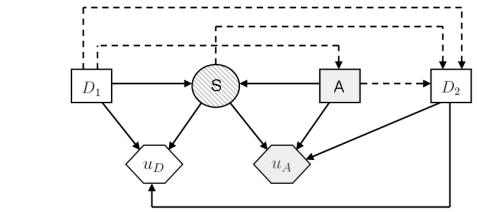
Similarly, $\pi_D(A_{II} = a_1)$ can be estimated by Monte Carlo simulation as follows.

1. For $j = 1, \ldots, n$, repeat
   Draw $\pi^j_{A_{II}} \sim \Pi_{A_{II}} = \mathcal{B}e(75, 25)$.
   Draw $(u^j_{A_{II}}, p^j_{A_{II}}) \sim (U_{A_{II}}, P_{A_{II}}) = F_{II}$
   Compute

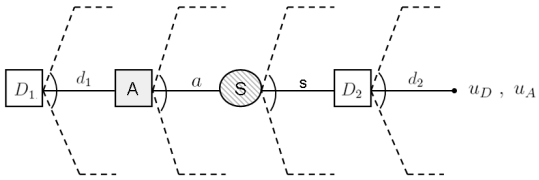
$$a^j_{II} = \arg\max_{a \in \mathcal{A}} \sum_{d \in \mathcal{D}} \left[\sum_{s \in \{0,1\}} u^j_{A_{II}}(a,s) \; p^j_{A_{II}}(S = s \mid d, a) \right] \pi^j_{A_{II}}(D_{II} = d)$$

2. Approximate $\pi_D(A_{II} = a_1)$ through
   $\hat{\pi}_D(A_{II} = a_1) = \#\{1 \le j \le n : \; a^j_{II} = a_1\}/n.$

In a run with $n = 10,000$, we got the approximations $\hat{\pi}_D(A_I = a_1) = 0.84$ and $\hat{\pi}_D(A_{II} = a_1) = 0.39$. Hence, $\pi_D(A = a_1)$ can be approximated by $\hat{\pi}_D(A = a_1) = 0.8\ \hat{\pi}_D(A_I = a_1) + 0.2\ \hat{\pi}_D(A_{II} = a_1) = 0.75$. The Defender can now solve her decision problem in (1), obtaining that her maximum expected utility defense is $d^* = d_1$ with (Monte Carlo estimated) expected utility 74.0, against d_2 whose expected utility is 45.9. △



(a) Influence diagram



(b) Game tree

Fig. 4. The Defend-Attack-Defend model

3. SEQUENTIAL DEFEND-ATTACK-DEFEND MODELS

We deal now with the sequential defend-attack-defend model, see Brown et al.[18] or Parnell et al.[27] for various examples. In it, the Defender first deploys defensive resources. Then, the Attacker, having observed such decision, performs an attack. Finally, the Defender tries to recover from the attack as best as she can. Fig. 4 shows coupled influence diagrams, with a shared uncertainty node S, and a game tree representing this model. 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 $d_1 \in \mathcal{D}_1$, $a \in \mathcal{A}$ and $d_2 \in \mathcal{D}_2$, which we shall assume continuous. Again, we shall assume that the only relevant uncertainty is the success level 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), the effort in implementing her protective and recovery actions and the mitigated result of the attack, and on (a, s, d_2), the effort in implementing his attack and the result of the attack, mitigated by the recovery action of the Defender.

3.1 A Game Theoretic Analysis

A game-theoretic approach requires the Defender to know the Attacker's utilities and probabilities, the Attacker to know the Defender's, and, furthermore, that all this is 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 the success of attack be $p_D(S = s \mid d_1, a)$ and $p_A(S = s \mid d_1, a)$. Then, we may compute a solution using backward induction as follows.

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

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

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 optimal attack decision at node A, after observing the Defender's first move $d_1 \in \mathcal{D}_1$, by solving

$$a^*(d_1) = \arg\max_{a \in \mathcal{A}} \psi_A(d_1, a).$$

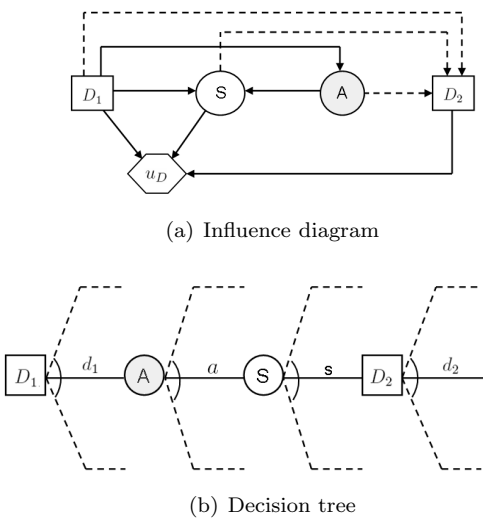
Knowing this, the Defender can find her maximum expected utility decision at node D_1 through

$$d_1^* = \arg\max_{d_1 \in \mathcal{D}_1} \psi_D(d_1, a^*(d_1)).$$

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 by 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 Approach

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 decision at node A as uncertain from the Defender's viewpoint and model such uncertainty. This is reflected in the influence diagram and the decision tree in Fig. 5, where the Attacker's decision node has been converted into a chance node, by replacing $\boxed{\text{A}}$ with $\textcircled{\text{A}}$. The Defender needs to assess $p_D(A|d_1)$, her predictive distribution about



(a) Influence diagram

(b) Decision tree

Fig. 5. The Defender's decision problem

what attack the Attacker will choose at node A against each $d_1 \in \mathcal{D}_1$, besides the (more standard) assessments $u_D(d_1, s, d_2)$ and $p_D(S \mid d_1, a)$.

Given these, the Defender can solve her decision problem working backwards the tree in Fig. 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 will use her probabilistic assessment about what the Attacker will do, $p_D(A|d_1)$, to compute her expected utility at node A for each $d_1 \in \mathcal{D}_1$,

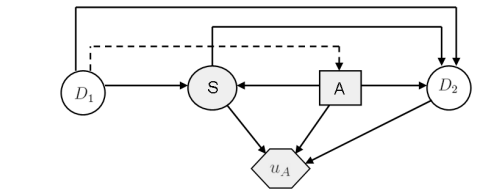
$$\psi_D(d_1) = \int \psi_D(d_1, a)\, p_D(A = a \mid d_1)\, \mathrm{d}a.$$

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

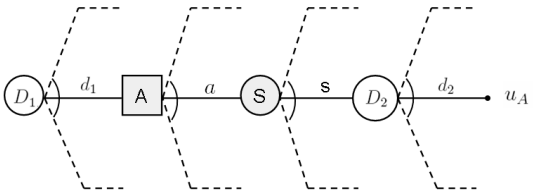
$$d_1^* = \arg\max_{d_1 \in \mathcal{D}_1} \psi_D(d_1).$$

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

Let us discuss now the assessment of $p_D(A \mid d_1)$. Alternatively to the standard risk analysis approach as in Ezell et al. [3], we propose in ARA 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 analysis of the Attacker's decision



(a) Influence diagram



(b) Decision tree

Fig. 6. The Defender's view of the Attacker's decision problem

problem, as seen by the Defender, is shown in Fig. 6, where the Attacker's probabilities and utilities need to be assessed from the Defender's perspective, based on all the information available to her. Again, should this kind of information not be available to the Defender, she could use a noninformative distribution to describe $p_D(A \mid d_1)$.

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

- 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, compute

$$d_1 \rightarrow A^*(d_1) = \arg\max_{a \in \mathcal{A}} \Psi_A(d_1, a).$$

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

   $$d_1 \rightarrow a_i^*(d_1) = \arg\max_{a \in \mathcal{A}} \psi_A^i(d_1, a)$$

2. `For any` a, `approximate` $\int_0^a p_D(A = x \mid d_1)\mathrm{d}x$ `through` $\#\{1 \leq i \leq n : a_i^*(d_1) \leq a\}/n$.

We have seen how the assessment of $p_D(A|d_1)$ is straightforward after the Defender's elicitation of F. However, the assessment of $P_A(D_2|d_1, a, s)$ within F could be problematic, as the Defender may want to exploit information available to her about how the Attacker analyzes her decision problem. Of course, if there is no information that the Defender can use, she will put a noninformative distribution over $P_A(D_2|d_1, a, s)$. The Defender may continue this recursive analysis, until eventually she has no more information to analyze the next level of the hierarchy of recursive decision models, much as described in §2.2. The recursive analysis will always stop at some point, perhaps after some simplifications leading to an heuristic distribution to model an adversary's thinking at some step of the recursive analysis, as illustrated in Rios Insua et al.[35] for an auction problem.

An illustration of the ARA approach to this model can be found in Sevillano et al.[39], where a sequential defend-attack-defend model is used by a ship owner to manage risks from piracy acts around the coast of Somalia when sailing through the Gulf of Aden.

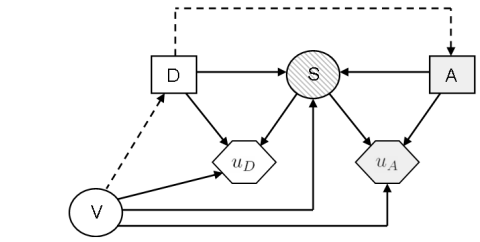
4. SEQUENTIAL DEFEND-ATTACK MODELS WITH PRIVATE INFORMATION

Our final basic model will be the sequential Defend-Attack model with Defender's private information, i.e. information that 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 the expected 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. Note that the Defender's decision allocating resources to protect different sites might signal to the Attacker about the sites' vulnerability and importance for the Defender, which is the type of information she wants to keep secret. This kind of applications, with the corresponding game theoretic analysis, has been considered e.g. by Powell[15], Zhuang et al.[40], and Zhuang and Bier[41,42].

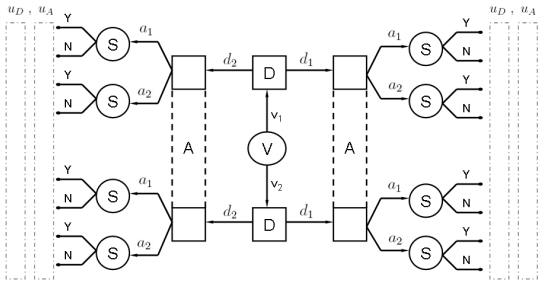
Assume that the Defender and the Attacker have, respectively, sets $\mathcal{D}$ and $\mathcal{A}$ of possible defenses and attacks. We shall also assume that the success level S of an attack is uncertain. The private information (e.g., vulnerabilities) is represented by V, whose value is known by the 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.

Fig. 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. 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 depend also on $V = v$. The arc in the influence diagram from 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. To simplify the figure, we only show two actions per adversary: $\mathcal{D} = \{d_1, d_2\}$ and $\mathcal{A} = \{a_1, a_2\}$; two possible outcomes (failure or success) of an attack: $S \in \{N, Y\}$; and two possible values for $V \in \{v_1, v_2\}$. The game tree reflects the sequential nature of the problem, as well as the asymmetric information. 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[43].



(a) Influence diagram



(b) Game tree

Fig. 7. The sequential Defend-Attack model with Defender's private information

4.1 A Game Theoretic Analysis

We briefly describe how standard game theory solves this model with private and asymmetric information. This is an example of a signaling game[44,45]. The game-theoretic approach requires the probability assessment of 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 $p_A(S|d, a, v)$. 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 the utilities $u_D(d, s, v)$ and $u_A(a, s, v)$, are common knowledge. A solution proceeds, then, as follows.

First, we define strategy functions for each player. As the Defender knows the value of V, her

strategy function is of the form $v \to d(v) \in \mathcal{D}$. As the Attacker makes his decision knowing the Defender's, his strategy function is of the form $d \to a(d) \in \mathcal{A}$. We compute the expected utilities of both players at node (S) of the tree in Fig. 7, for any pair of decisions $(d, a) \in \mathcal{D} \times \mathcal{A}$ and value of private information $V = v$:

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

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

The Attacker's best response against a defense d is

$$a^*(d) = \arg\max_{a \in \mathcal{A}} \int \psi_A(d, a, v)\, \pi_A(V = v \mid d)\, \mathrm{d}v, \quad (9)$$

where $\pi_A(V|d)$ represents the Attacker's updated beliefs about the Defender's private information, after having observed her defense action. We show how to determine $\pi_A(V|d)$ below. For now, we shall assume it is known. Under the assumption that the Defender knows how the Attacker will solve his decision problem for any $d \in \mathcal{D}$, the Defender's maximum expected utility decision, given that she knows the value of $V = v$, would be

$$d^*(v) = \arg\max_{d \in \mathcal{D}} \psi_D(d, a^*(d), v).$$

As commonly accepted in game theory, we allow for randomized strategies. Assuming that $\mathcal{D}$ 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

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

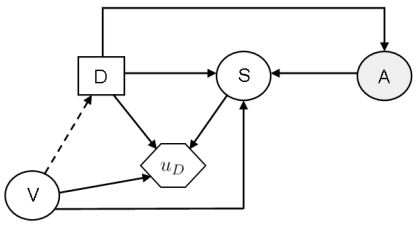
as their associated sets of randomized strategies. Hence, $d^*(v)$ and $a^*(d)$ have associated probability distributions $\pi_{d^*(v)}(d \mid v) \in \Pi_{\mathcal{D}}$ and $\pi_{a^*(d)}(a \mid d) \in \Pi_{\mathcal{A}}$, respectively.

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

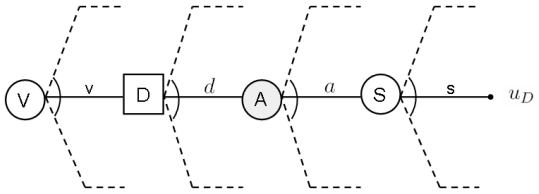
$$\pi_A(V = v \mid d) \propto \pi_A(V = v)\, \pi_{d^*(v)}(D = d \mid v),$$

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

A game theoretic solution can be determined,



(a) Influence diagram



(b) Decision tree

Fig. 8. The Defender's decision problem

then, by finding a pair of strategies $\left(\pi_{d^*(v)}, \pi_{a^*(d)}\right)$ which are a fixed point solution of the equations

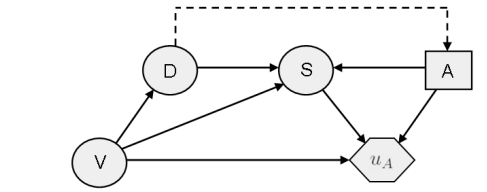
$$\begin{cases} \pi_{d^*(v)} = \arg\max_{\pi \in \Pi_{\mathcal{D}}} \int_{\mathcal{D}} \left[\int_{\mathcal{A}} \psi_D(d, a, v) \right. \\ \qquad \left. \pi_{a^*(d)}(a \mid d)\, \mathrm{d}a \right] \pi(d)\, \mathrm{d}d,\ \forall v \in V \\ \pi_{a^*(d)} = \arg\max_{\pi \in \Pi_{\mathcal{A}}} \int_{\mathcal{A}} \left[\int_{V} \psi_A(d, a, v) \right. \\ \qquad \left. \pi_A(v)\, \pi_{d^*(v)}(d \mid v)\, \mathrm{d}v \right] \pi(a)\, \mathrm{d}a,\ \forall d \in \mathcal{D} \end{cases} \quad (10)$$

Note that a fixed point solution of the equations in (10) is a Nash equilibrium. In addition, we have assumed that the Attacker's learning behavior follows Bayes' rule.

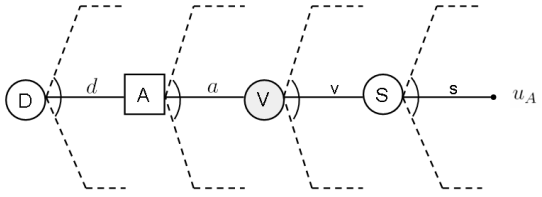
4.2 The ARA Approach

For a more realistic approach, we weaken the common knowledge assumption. We consider the Defender's decision problem as a standard decision analysis problem, illustrated in Fig. 8, with the Attacker's decision node perceived now as a random variable. Similarly, her decision tree shows uncertainty about the Attacker's decision by replacing [A] with (A).

Assume, 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



(a) Influence diagram



(b) Decision tree

Fig. 9. The Defender's analysis of the Attacker's decision

that she has chosen defense d. Obtaining this will require that the Defender analyzes the problem from the Attacker's perspective. Assume for a moment that she has assessed $p_D(A|d)$. Then, the Defender can obtain her maximum expected utility defense by solving the tree in Fig. 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

$$(d, v) \to \psi_D(d, v) = \int \psi_D(d, a, v)\; p_D(A = a \mid d)\; \mathrm{d}a$$

- At decision node D, solve

$$v \to d^*(v) = \arg\max_{d \in \mathcal{D}} \psi_D(d, v).$$

To assess $p_D(A|d)$, the Defender must place herself on the Attacker's shoes and solve his decision problem from her perspective. Fig. 9 represents the Attacker's decision problem, as seen by the Defender. Note that the Defender's decision is represented as a random variable in the Attacker's analysis, as it is not under his control. The arrow from (D) to [A] in the influence diagram indicates that the Defender's decision will be known to him at the time he has to decide. As the Attacker does not know the Defender's private information v, his uncertainty is represented through a probability distribution $p_A(V)$, describing the Attacker's (prior) beliefs about the Defender's private information. We assume that the Defender analyzes the Attacker's decision considering that he is an expected utility maximizer and that he uses Bayes rule to learn about the Defender's private information from the observation of her defense decision. Thus, the arrow in the influence diagram from (V) to (D), which represents probabilistic dependence, can be inverted to obtain the Attacker's (posterior) beliefs about v: $p_A(V|D = d)$. However, to obtain this 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 $a^*(d)$ for any $d \in \mathcal{D}$ by solving backwards the tree in Fig. 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)

$$\psi_A(d, a) = \int \psi_A(d, a, v)\; p_A(V = v \mid d)\; \mathrm{d}v. \tag{11}$$

- At decision node A, solve

$$d \to a^*(d) = \arg\max_{a \in \mathcal{A}} \psi_A(d, a).$$

However, the Defender does not know (p_A, u_A), but she has beliefs about them, say $(P_A, U_A) \sim F$, which will be relevant in her analysis of the Attacker's decision problem. 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 (11), through, 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, the Defender's predictive distribution about the Attacker's response to any of her defense choices d is defined through

$$p_D(A = a|d) = \mathbb{P}_F\Big[a = \arg\max_{x \in \mathcal{A}} \Psi_A(d, x)\Big],\ \forall a \in \mathcal{A}.$$

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

$$\hat{p}_D(A = a|d) = \#\{1 \le i \le n :\ a_i^*(d) = a\}/n,\ \forall a \in \mathcal{A},$$

when $A \mid d$ is discrete, or

$$\hat{p}_D(A \le a|d) = \#\{1 \le i \le n :\ a_i^*(d) \le a\}/n,\ \forall a \in \mathcal{A},$$

when $A \mid d$ is absolutely continuous.

To sum up, the elicitation of $F = (P_A(S|d,a,v), P_A(V|d), U_A(a,s,v))$ allows the Defender to solve her problem of assessing $p_D(A|d)$. The Defender may have enough information and judgment available to directly assess $P_A(S|d,a,v)$ and $U_A(a,s,v)$. However, the assessment of $P_A(V|d)$ requires a deeper analysis, as it has a strategic component.

Specifically, assuming that the Attacker has prior knowledge over V modeled through $p_A(V)$, his posterior beliefs about V, after he observes $D = d$, become:

$$p_A(V = v|d) \propto p_A(V = v)\; p_A(D = d|v), \qquad (12)$$

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 $p_A(D|v)$ requires an analysis of how the Attacker analyzes the Defender's decision. Assuming he thinks that she is an expected utility maximizer, and that the decision problem she tries to solve is as in Fig. 8, with A^1 representing the Attacker's decision within this level of recursive modeling, the Defender's elicitation of a probability distribution $G = (\, U_D(d,s,v),\; P_D(S|d,a,v),\; P_D(A^1|d)\,)$ representing the Attacker's probabilistic assessments of her utilities and probabilities, allows her to solve her problem of assessing $p_A(D|v)$ by evaluating a tree like the one in Fig. 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^1, compute for each (d,v)

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

- At decision node D, solve for each v

  $$v \rightarrow p_A(D = d|v) = \mathbb{P}_G\Big[d = \arg\max_{x \in \mathcal{D}}\; \Psi_D(x,v)\Big],\; \forall d \in \mathcal{D}. \qquad (13)$$

As the Attacker's beliefs represented within G are assessed from the Defender's perspective, her uncertainty about these beliefs when acknowledged within G will produce the distribution $P_A(D|v)$ in (13), representing what the Defender beliefs to be $p_A(D|v)$. Note also that $p_A(V)$ in (12) represents the Attacker's prior knowledge about the Defender's private information. As the Defender does not have access to this distribution, 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. Then, from the Defender's perspective, the Attacker's learning about V modeled in (12) becomes

$$P_A(V = v|d) \propto P_A(V = v)\; P_A(D = d|v). \qquad (14)$$

The only difficulty for the Defender at this step, in order to obtain $P_A(D|v)$, 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. We may go further 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, similarly as described in §2.2. However, if no information is available at this level, we can 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 reaching a robust enough recommendation. In such case, it would be desirable to collect more data and/or judgement through intelligence.

We illustrate the ARA approach to the sequential Defend-Attack model with Defender's private information with a simple numerical example.

Example. Consider a Defender who needs to protect two sites against a potential terrorist attack. The Defender has a limited amount of defensive resources distributed between both sites. The Attacker knows the total amount of defensive resources. This is common knowledge as it was publicized by the Defender. However, the actual distribution of defensive resources between both sites (v_1 for Site 1, v_2 for Site 2 for each unit of defensive resources, $v_1 + v_2 = 1$) is only known by the Defender.

The Attacker has the capacity of launching an attack against either one of the two sites, but not both, and he has announced that he will attack one of the sites with all his available resources. Therefore, the available actions for the Attacker are $a \in \mathcal{A} = \{a_1, a_2\}$ with a_i representing attacking Site i, $i =$

$1, 2$. The Defender has the option of re-distributing her allocation of defensive resources by moving d of them from one site to another. This move will be observed by the Attacker before deciding which site to attack. The set of alternatives for the Defender is then $\mathcal{D}(v_1) = \{d : v_1 - 1 \leq d \leq v_1\}$, where, for example, $d = 0.5$ means that she will move half of her total defensive resources from Site 1 to Site 2, and $d = -0.5$ that half of her defensive resources will be moved from Site 2 to Site 1. If $v_1 \in [0, 1]$ is the initial distribution of defensive resources between sites, the distribution after her move d will be: $v_1 - d$ for Site 1 and $v_2 + d$ for Site 2.

The probability that an attacked site is destroyed (successful attack) depends on the amount of resources committed to that site by the Defender. The more defensive resources allocated to the site, the lower the probability that the attack will be successful. Specifically, they both share the (commonly known) beliefs that

$$p(S_1 = 1|d, a, v) = \begin{cases} 1 - (v_1 - d), \text{ if } a = a_1 \\ 0, \text{ if } a = a_2 \end{cases} \quad (15)$$

$$p(S_2 = 1|d, a, v) = \begin{cases} 0, \text{ if } a = a_1 \\ 1 - (v_2 + d), \text{ if } a = a_2 \end{cases} \quad (16)$$

Note that as deploying more defensive resources at a site reduces the probability that an attack on that site will be successful, $v = (v_1, v_2)$ can be interpreted as a measure of the sites' vulnerability before the Defender's move: the lower the amount of initial defensive resources allocated to a site, the higher the success probability of an attack on the site, and the higher the need for the Defender to move resources to that site in order to reduce that risk. The Defender keeps the information about the sites' vulnerability v as a secret, and, thus, the Attacker does not know v. However, the Defender thinks the Attacker is capable of learning about her private information v by observing her move d using Bayes' rule as in (12).

The Attacker's and Defender's objectives are commonly known to be maximizing the probability of succeeding in his attack (for the Attacker), and minimizing this probability (for the Defender). Thus, the utilities for the Defender and the Attacker associated with each outcome are known to be, respectively,

$$u_D(s_1, s_2) = \begin{cases} 1, \text{ iff } S_1 = 0 \text{ and } S_2 = 0 \\ 0, \text{ otherwise} \end{cases} \quad (17)$$

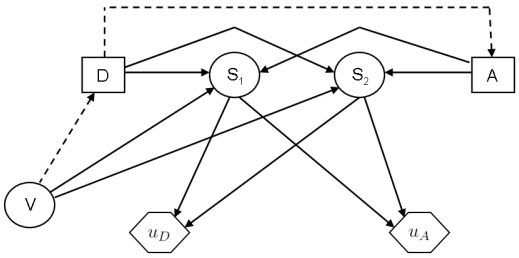


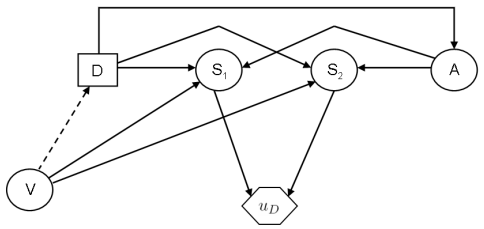
Fig. 10. A sequential Defend-Attack resource allocation problem between two sites with Defender's private information about the sites' vulnerabilities

$$u_A(s_1, s_2) = \begin{cases} 1, \text{ iff } S_1 = 1 \text{ or } S_2 = 1 \\ 0, \text{ otherwise} \end{cases} \quad (18)$$

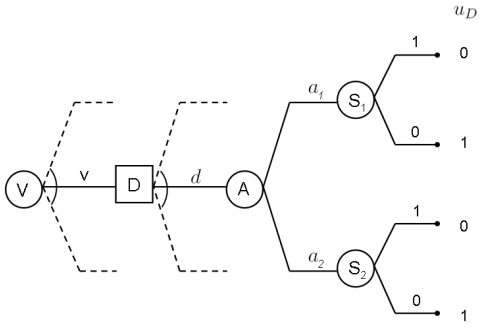
Fig. 10 shows coupled influence diagrams representing this decision situation. Node S_i represents the uncertainty associated with the success of an attack carried out against Site i, with $i = 1, 2$. These uncertainties depend on the actions taken by both the Attacker and the Defender, as well as on the initial distribution v of defensive resources secretly allocated by the Defender between both sites.

To simplify, we have assumed that the Defender's and the Attacker's preferences for the different outcomes are commonly known to be described respectively by (17) and (18), and that both the Defender and the Attacker share the same commonly known beliefs about $S_1 \mid d, a, v$ and $S_2 \mid d, a, v$, described by (15) and (16) respectively. If v were also common knowledge, then the optimal decision for the Defender would be $d = (v_1 - v_2)/2$, thus leaving the same amount of defensive resources in each of both sites after her re-distributing move, with the Attacker indifferent between striking any of the sites.

When v is privately known by the Defender only, the game-theoretic approach based on Bayes-Nash equilibrium assumes that the Attacker's beliefs over v are common knowledge. We deem unrealistic this assumption that the Attacker will reveal his beliefs about v, and solve the problem for the Defender weakening it. Fig. 11 shows the influence diagram and the decision tree representing the Defender's decision problem, in which the Attacker's decision is perceived by her as an uncertainty and the value of v is observed by her before making her decision. The analysis of her uncertainty about the Attacker's decision requires that she thinks about the decision problem faced by the Attacker, which is shown in Fig. 12, where now her decision is modeled as an

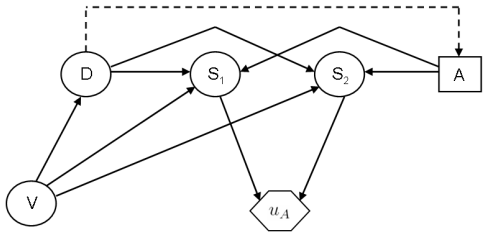


(a) Influence diagram

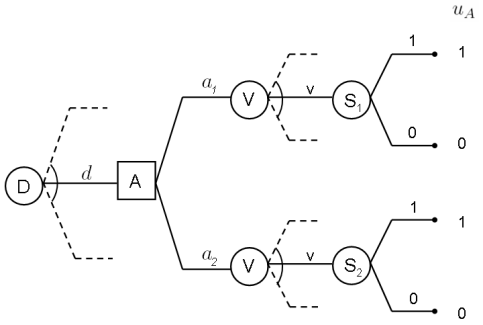


(b) Decision tree

Fig. 11. The Defender's resource re-allocation decision problem



(a) Influence diagram



(b) Decision tree

Fig. 12. The Defender's analysis of the Attacker's decision on what site to attack

uncertainty from the Attacker's perspective, and the value of v is unknown to him at the time he makes his decision.

Assume that we are able to obtain from the Defender the following assessments:

- $v = (v_1 = 0.5, v_2 = 1 - v_1 = 0.5)$: The initial allocation of defensive resources between both sites.
- $V = (V_1, V_2 = 1 - V_1)$: The Attacker's assessment of the Defender's private information, as elicited by the Defender. The Defender knows v, but the Attacker does not. The Attacker's (prior) beliefs over the possible values of $V_1 = v_1 \in [0, 1]$ will be represented by $p_A(V_1)$ and will be elicited from the Defender's perspective through $P_A(V_1)$. Based on the information available to her, she believes that $p_A(V_1)$ is a beta distribution $\mathcal{B}e(\alpha, \beta)$ with mean $\mu = \alpha/(\alpha + \beta)$ and precision $\nu = \alpha + \beta$ within the ranges $\mu \in [0.5, 0.8]$ and $\nu \in [10, 30]$, respectively. Based on this, we model $P_A(V_1)$ as a hierarchical beta distribution $\mathcal{B}e(\mu\,\nu, (1 - \mu)\,\nu)$ with $\mu \sim \mathcal{U}(0.5, 0.8)$ and $\nu \sim \mathcal{U}(10, 30)$.
- A heuristic for assessing $p_D(A^1|d)$, the Defender's beliefs about which site the Attacker will attack after observing her move, within the model used by the Defender to represent how the Attacker thinks she will solve her decision problem. To solve the Defender's problem from the Attacker's perspective, he would need to assess $p_D(A^1|d)$. The heuristic model for the Attacker's choice at this level of analysis assumes that the Attacker will choose the site with less defensive resources and that he will not revise his estimate $\hat{v}$ of v after observing her move d. Specifically,

$$A^1 = a_1, \text{ if } \hat{v}_1 - d < 0.5,$$
$$A^1 = a_2, \text{ if } \hat{v}_1 - d > 0.5,$$

  where $\hat{v}_1$ represents the Attacker's estimate of v_1, the initial defensive resources in Site 1. Thus, the Defender ends the hierarchy of recursive decision analysis at this point, disregarding the modeling of further and more complex levels of analysis.
- $\hat{V} = (\hat{V}_1, \hat{V}_2 = 1 - \hat{V}_1)$: The Defender's beliefs of the Attacker's estimate $\hat{v}$ of her private information v, as would be assessed by the Attacker. The heuristic above has reduced the assessment of $p_D(A^1|d)$ to the assessment of $\hat{V}$. We use a a triangular distribution on $[0, 1]$ with mode ϕ to model the Defender's beliefs

about $\hat{v}_1$. Thus, $\hat{V}_1|\phi \sim Tri(min = 0,\ mode = \phi,\ max = 1)$. The Defender also thinks that the Attacker believes that she overestimates what he thinks to be v_1. Specifically, $\phi \mid \theta, \sigma \sim \mathcal{N}(V_1 + \theta, \sigma)$ truncated on $[0, 1]$, with the Defender assessing that $\theta \sim \mathcal{U}(0.1, 0.2)$ and $\sigma \sim \mathcal{U}(0.1, 0.3)$.

We find a solution to the Defender's decision problem as follows.

1. Compute the Defender's beliefs over A^1 based on the proposed heuristic:

$$p_D(A^1 = a_1|d) = \Pr(\hat{V}_1 - d < 0.5|\phi),$$

where $\hat{V}_1|\phi \sim Tri(0,\ mode = \phi,\ 1)$. Thus, given ϕ,

$$p_D(A^1 = a_1|d) = \tag{19}$$
$$= \begin{cases} 0, & 0.5 + d \le 0 \\ (0.5 + d)^2/\phi, & 0 \le 0.5 + d \le \phi \\ 1 - (0.5 - d)^2/(1 - \phi), & \phi \le 0.5 + d \le 1 \\ 1, & 1 \le 0.5 + d \end{cases}$$

2. Given $p_D(A^1|d)$, the Defender can solve her decision problem by working backwards the tree in Fig. 11 as follows:

- `At chance nodes` S_1 `and` S_2`, compute for each` (v, d, a)`,`

$$\psi_D(v, d, a) = \sum_{s_1 \in \{0,1\}} \sum_{s_2 \in \{0,1\}} \Big[u_D(s_1, s_2)\ p(S_1 = s_1, S_2 = s_2 \mid d, a, v)\Big]$$
$$= p(S_1 = 0 \mid d, a, v)\ p(S_2 = 0 \mid d, a, v)$$
$$= \begin{cases} v_1 - d, & a = a_1 \\ v_2 + d, & a = a_2 \end{cases}$$

- `At chance node` A^1`, compute for each` (v, d)`,`

$$\psi_D(v, d) = \psi_D(v, d, a_1)\ p_D(A^1 = a_1|d) + \psi_D(v, d, a_2)\ (1 - p_D(A^1 = a_1|d))$$

`Thus,` $\psi_D(v, d)$ `is`

$$\begin{array}{ll} v_2 + d, & \\ \text{if } d \le -0.5, & \\ (v_1 - d)\ (0.5 + d)^2/\phi\ + & \\ (v_2 + d)\ (1 - (0.5 + d)^2/\phi), & \\ \text{if } -0.5 \le d \le -0.5 + \phi, & \\ (v_1 - d)\ (1 - (0.5 - d)^2/(1 - \phi))\ + & \\ (v_2 + d)\ (0.5 - d)^2/(1 - \phi), & \\ \text{if } -0.5 + \phi \le d \le 0.5, & \\ v_1 - d, & \\ \text{if } 0.5 \le d. & \end{array}$$

- `At decision node` D`, solve for each` v`,`

$$d^*(v) = \underset{d \in [v_1 - 1, v_1]}{\arg\max}\ \psi_D(v, d)$$

`Thus, for each` $v_1 \in [0, 1]$`, the optimal decision for the Defender,` $d^*(v_1)$`, is`

$$-\frac{1}{2} + \frac{2\,v_1 + \sqrt{4\,v_1^2 + 6\,\phi}}{6}, \tag{20}$$

`if` $v_1 \le (6\phi - 1)/4$`, and`

$$\frac{1}{2} + \frac{2\,(v_1 - 1) - \sqrt{4\,(v_1 - 1)^2 + 6\,(1 - \phi)}}{6}, \tag{21}$$

`if` $v_1 \ge (6\phi - 1)/4$.

From the perspective of the Attacker, $p_D(A^1|d)$ is not known since he does not have access to the value of ϕ used by the Defender in (19). Should the Attacker know ϕ, he would be able to anticipate the Defender's optimal move, $d^*(v)$, for each possible initial allocation v. At this level of the recursive analysis, the Attacker' beliefs about ϕ are propagated to define $p_A(D|v_1)$ from $d^*(v_1)$ in (20)–(21), when $\phi \mid v_1, \theta, \sigma \sim \mathcal{N}(v_1 + \theta, \sigma)$ truncated on $[0, 1]$.

3. Given $p_A(D|v)$ and $p_A(V)$, the Attacker can learn about V from his observation of $D = d$ by computing $p_A(V|d)$ using (12), and solve his decision problem by working backwards the tree in Fig. 12 as follows:

- `At chance nodes` S_1 `and` S_2`, compute for each` (d, a, v)`,`

$$\psi_A(d, a, v) = \sum_{s_1 \in \{0,1\}} \sum_{s_2 \in \{0,1\}} \Big[u_A(s_1, s_2)\ p(S_1 = s_1, S_2 = s_2 \mid d, a, v)\Big]$$
$$= 1 - p(S_1 = 0 \mid d, a, v)\ p(S_2 = 0 \mid d, a, v)$$
$$= \begin{cases} p(S_1 = 1 \mid d, a_1, v), & a = a_1 \\ p(S_2 = 1 \mid d, a_2, v), & a = a_2 \end{cases}$$

- `At chance node` V`, compute for each` (d, a)`,`

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

- `At decision node` A`, solve for each` d`,`

$$a^*(d) = \underset{a \in \{a_1, a_2\}}{\arg\max}\ \psi_A(d, a),$$

`obtaining that, for each` $d \in [-1, 1]$`,`

$$\begin{array}{ll} a^*(d) = a_1, & \psi_A(d, a_1) > \psi_A(d, a_2) \\ a^*(d) = a_2, & \psi_A(d, a_1) < \psi_A(d, a_2) \end{array}$$

`where`

$$\psi_A(d, a_1) > \psi_A(d, a_2) \iff \int (1 - 2\,v_1 + 2\,d)\ p_A(V_1 = v_1|d)\ \mathrm{d}v_1 > 0 \iff E_{p_A}(V_1 \mid d) - d < 1/2. \tag{22}$$

`Thus, the optimal decision for the`

`Attacker is`

$$a^*(d) = a_1, \text{ if } E_{p_A}(V_1 \mid d) - d < 1/2,$$
$$a^*(d) = a_2, \text{ if } E_{p_A}(V_1 \mid d) - d > 1/2.$$

The Attacker then chooses a_1 (attack Site 1) when, after observing d, he expects less defensive resources in Site 1 than in Site 2, where the Attacker's expectation is computed with respect to his updated beliefs about V: $p_A(V|d)$. Should the Defender know $p_A(V|d)$, she would be able to anticipate the Attacker's choice by computing (22). However, she does not know it as she is uncertain about the Attacker's probabilities $p_A(D|v)$ and $p_A(V)$, which are necessary to compute his $p_A(V|d)$. But we can assess her beliefs $P_A(V) = p_A(V|\mu, \nu)$, $\mu \sim \mathcal{U}(0.5, 0.8)$, $\nu \sim \mathcal{U}(10, 30)$, and $P_A(D|v) = p_A(D|v, \theta, \sigma)$, $\theta \sim \mathcal{U}(0.1, 0.2)$, $\sigma \sim \mathcal{U}(0.1, 0.3)$, to obtain $P_A(V|d)$ as in (14), and, in turn, compute her predictive probability of the Attacker's decision

$$p_D(A = a_1|d) = \mathbb{P}_{\mu,\nu,\theta,\sigma}[E_{p_A}(V_1 \mid d) - d < 1/2].$$

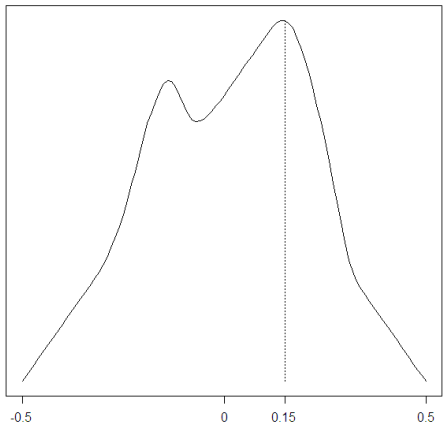
4. We use Monte Carlo simulation to estimate $p_D(A = a_1|d)$

`For` $k = 1, \ldots, n_k$
  `Simulate` $p^k(V_1) \sim P_A(V_1)$
    $\mu^k \sim \mathcal{U}(0.5, 0.8)$
    $\nu^k \sim \mathcal{U}(10, 30)$
    `Set` $p^k(V_1) = \mathcal{B}e(\mu^k\ \nu^k, (1 - \mu^k)\ \nu^k)$
  `Simulate` $p^k(D|v_1) \sim P_A(D|v_1)$
    $\theta^k \sim \mathcal{U}(0.1, 0.2)$
    $\sigma^k \sim \mathcal{U}(0.1, 0.3)$
    `Set` $\phi_k|v_1 = \mathcal{N}(v_1 + \theta^k, \sigma^k)$ `truncated on` $[0, 1]$
      `Thus,` $p^k(D|v_1) = \mathbb{P}_{\phi_k|v_1}(D = d^*(v_1))$
  `Simulate` (v_1^i, d^i) `from` $p^k(V_1, D) \sim P_A(V_1, D)$
    `For` $i = 1, \ldots, n_i$
      $v_1^i \sim p^k(V_1)$
      $\phi_k^i \sim \phi_k|V_1 = v_1^i$
      $d^i = d^*(v_1^i)$, `with` $\phi = \phi_k^i$ `in` (20)-(21)
        `Thus,` $d^i \sim p^k(D|V_1 = v_1^i)$
  `For every` $-1 \leq d \leq 1$
    $\{ v_j^d = v_1^i : (v_1^i, d^i = d) \}_{j=1}^{n_d} \sim p^k(V_1|d)$
    `Approximate` $E_{p^k}(V_1|d)$ `by` $\sum_{j=1}^{n_d} v_j^d / n_d$
`Approximate` $p_D(A = a_1|d)$ `by`
  $\#\{1 \leq k \leq n_k : E_{p^k}(V_1|d) - d < 1/2\}/n_k$.

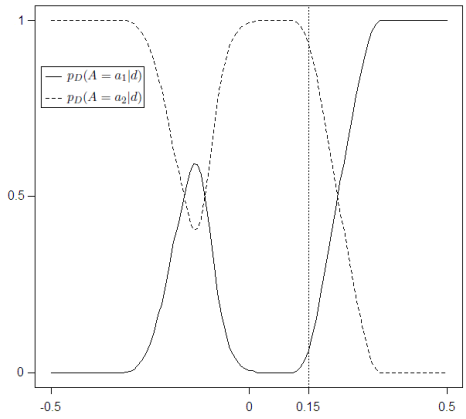
5. Once $p_D(A|d)$ has been approximated, the Defender can find her (Monte Carlo estimated) maximum expected utility decision $d^*(v)$ by solving the decision tree in Fig. 11 using backwards induction:

$$d^* = \underset{d \in [-v_2\ ,\ v_1]}{\arg\max}\ (v_1 - d)\ p_D(A = a_1 \mid d) + (v_2 + d)\ p_D(A = a_2 \mid d),$$

with $v = (v_1 = 0.5, v_2 = 0.5)$



$\psi_D(d),\ -0.5 \leq d \leq 0.5$



$p_D(A|d),\ -0.5 \leq d \leq 0.5$

Fig. 13. Defender's expected utilities and predictive probabilities of site strike

We used $n_k = n_i = 5,000$ to run the Monte Carlo simulation and obtained that the Defender's optimal move is $d^* = 0.15$. Thus, given the assessments from the Defender, her maximum expected utility action is to re-allocate 15% of her defensive resources by moving them from Site 1 to Site 2, decreasing her resources in Site 1 from 50% ($v_1 = 0.5$) to 35% ($v_1 - d^*$) and increasing them in Site 2 from 50% ($v_2 = 0.5$) to 65% ($v_2 + d^*$). Fig. 13 shows the Defender's (Monte Carlo estimated) expected utilities $\psi_D(d)$ of her feasible moves $-0.5 \leq d \leq 0.5$ as well as her (Monte Carlo estimated) predictive probabilities $p_D(A|d)$ of each site being attacked for each of her feasible d. For $d^* = 0.15$, we have that $p_D(A = a_1|d^*) = 0.06$ and $p_D(A = a_2|d^*) = 0.94$.

We may see how the solution proposed by the ARA approach is consistent with the Defender's beliefs on the Attacker overestimating her resources

initially allocated to Site 1. The move of some defensive resources from Site 1 to Site 2 will reinforce the Attacker's perceived beliefs. Thus, the Defender's expected utility increases by moving resources to Site 2 until her predictive probability of an attack to Site 1 starts increasing sharply, see Fig. 13. This allows for the increase of the relative strength of Site 2, the site that a priori is more likely to be attacked, until more resources sent to Site 2 start signalling that Site 1 will end up with less resources. Sending too much defensive resources to Site 2 would allow the Attacker to change his beliefs about which site has less resources and make the Defender vulnerable to an attack on Site 1.

Finally, we also note that the Defender's expected utility function has another local (but not global) maximum which corresponds to $d = -0.15$. This move would allow the Defender to increase the relative strength of Site 1 while at the same time making the Attacker believe that she had less resources in that site. Thus, moving some defensive resources from Site 2 to Site 1 would make the Attacker revise down his (prior) beliefs on the initial amount of defensive resources in Site 1, increasing the Defender's predictive probability of an attack to Site 1 to a point in which for $d = -0.15$, $p_D(A = a_1|d = -0.15) = 0.59$. At this point, the move of more resources to Site 1 would decrease her perceived likelihood of an attack to Site 1, making, in turn, her expected utility to decrease again, see Fig. 13. $\triangle$

5. DISCUSSION

We have provided an account of how the ARA framework can support a Defender against an intelligent adversary, the Attacker, whereby the Defender assesses the probabilities of the adversaries' actions before computing her maximum expected utility defense. 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. Instead of using point estimates for the Attacker's probabilities and utilities, which would lead to a point estimate of his maximum expected utility decision, we build a distribution over them, acknowledging the uncertainty on the estimates. Part of these uncertainties can be directly elicited from the Defender, but other parts may require strategic thinking as illustrated above. This leads to a hierarchy of recursive decision analysis, in which the Defender accommodates as much information as she can, until she may not provide additional information, step at which we use a noninformative distribution to close the hierarchy of analysis. A clear advantage of structuring the Attacker's problem is that the Defender can isolate different uncertainty and value components of her problem and accommodate different expertise, facilitating the assessment of the adversary's priorities and beliefs, as noted in Merrick and Parnell[46].

We have used a subjective expected utility model to predict the terrorist's decision behavior. One could question the hypothesis of the adversary being rational, but the recent terrorist behaviour literature supports such hypothesis[47,1], meaning that terrorists tend to use their limited offensive resources to cause significant damage where there can be a high probability of success. Note that we could assume other optimizing models to describe terrorists' decision behavior, but our arguments could be easily translated without changing our methodology.

We have focused on how the ARA framework could be applied to basic counterterrorism models in order to illustrate the key methodological steps of the analytic framework. The models we have discussed are relatively simple, but they retain the essence of counterterrorism decision making. Real problems are much more complex, with 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. In those cases, we would expect to deploy more complex coupled influence diagrams for the Defender and the Attacker, partitioned according to time and information, possibly as in Koller and Milch[48], in sequences of defend-attack-defend moves, simultaneous defend-attack moves and sequential defend-attack moves with private information. Thus, we view the three models treated here as basic model building blocks to deal with more complex problems.

Note that we have paid little attention to the numerical intricacies associated with the need to optimize resources at the decision nodes. Of course, when it comes to the application to real problems, the structuring of very complex decision dynamics with multiple uncertainties and types of adversaries, and the corresponding elicitation processes and calculations necessary to find the optimal decisions would become intractable. MCMC methods, specially of the augmented simulation

type, see Bielza et al.[49], might be very relevant. However, setting up a realistic conceptual framework of analysis is a necessary first step to consider before we may approximate and simplify in order to find satisfactory and meaningful recommendations for the Defender in real applications.

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 cooperation among the defenders to share resources and reduce the risks posed by the adversaries, possibly as described in Rios and Rios Insua[50,51]. The ARA approach would provide internal advice to the group of defenders using external predictive models of the attackers' decision behavior, which would also include the possibility of cooperation among various attackers.

Finally, note that the ARA framework might find applications in other contexts. Areas such as marketing and cybersecurity seem relevant. In these cases we might be facing a large and uncertain number of adversaries. Rios Insua et al.[35] referred to some simple auctions. This is in line with the recent debate between using decision analysis or game theory models for the analysis of competing situations, well reflected in papers such as Rothkopf[52] or van Bingsbergen and Marx[53].

ACKNOWLEDGMENT

We are grateful to SAMSI-NSF for providing the initial support for this research and the COST-ESF ALGODEC program for its later stages. DRI is grateful to the MICINN e-COLABORA and ARA projects and the RIESGOS-CM and FP7 SECONOMICS programs. JR is grateful to the US/UK ITA research program. The views and conclusions contained in this paper are those of the authors and should not be interpreted as representing the opinions and policies of any of the supporting agencies. Discussions with David Banks are gratefully acknowledged. We are also grateful to many suggestions by the referees.

REFERENCES

1. English R. Terrorism: How to Respond. Oxford University Press, 2009.
2. Lomborg B. Solutions for the World's Biggest Problems. Cambridge University Press, 2008.
3. Ezell BC, Bennett SP, von Winterfeldt D, Sokolowski J, Collins AJ. Probabilistic risk analysis and terrorism. Risk Analysis, 2010; 30(4):575-589.
4. Gutfraind A. Terrorism as a mathematical problem. SIAM News, October 2009.
5. Wein L. Homeland security: from mathematical models to policy implementation. Operations Research, 2009; 57:801-811.
6. Parnell G, Banks D, Borio L, Brown G, Cox LA, Gannon J, Harvill E, Kunreuther H, Morse S, Pappaioanou M, Pollack S, Singpurwalla N, Wilson A. Report on Methodological Improvements to the Department of Homeland Security's Biological Agent Risk Analysis. National Academies Press, 2008.
7. Enders W, Sandler T. The Political Economy of Terrorism. Cambridge University Press, 2009.
8. Deisler Jr, PF. A perspective: Risk analysis as a tool for reducing the risks of terrorism. Risk Analysis, 2002; 22(3):405-413.
9. Garrick BJ. Perspectives on the use of risk assessment to address terrorism. Risk Analysis, 2002; 22(3):421-423.
10. Dillon RL, Liebe RM, Bestafka T. Risk-based decision making for terrorism applications. Risk Analysis 2009; 29(3):321-335.
11. Cox Jr, LA. Improving risk-based decision making for terrorism applications. Risk Analysis, 2009; 29(3):336-341.
12. Brown G, Cox Jr, LA. How probabilistic risk assessment can mislead terrorism risk analysts Risk Analysis, 2011; 31(2):196-204.
13. Siqueira K, Sandler T. Terrorists vs the government: Strategic intervention, support and sponsorship. Journal of Conflict Resolution, 2006; 50:878-898.
14. Arce D, Sandler T. Terrorist signalling and the value of intelligence. British Journal Political Science, 2007; 37:573-586.
15. Powell R. Allocating defensive resources with private information about vulnerability. American Political Science Review, 2007; 101:799-809.
16. Zhuang J, Bier V. Balancing terrorism and natural disasters – Defensive strategy with endogenous attack effort. Operations Research, 2007; 55(5):976-991.
17. Brown G, Carlyle M, Salmeron J, Wood K. Analyzing the vulnerability of critical infrastructure to attack and planning defenses. Tutorials in Operations Research. INFORMS, 2005. p. 102-123.
18. Brown G, Carlyle M, Salmeron J, Wood K. Defending critical infrastructure. Interfaces, 2006; 36(6):530-544.
19. Brown G, Carlyle M, Wood R. Optimizing Department of Homeland Security Defense Investments: Applying Defender-Attacker(-Defender) Optimization to Terror Risk Assessment and Mitigation. National Academies Press, 2008. Appendix E.
20. Kardes E. Robust stochastic games and applications to counter-terrorism strategies. CREATE report, 2005.
21. Bier V, Azaiez N. Game Theoretic Risk Analysis of Security Threats. Springer, 2009.
22. Hausken K. Probabilistic risk analysis and game theory. Risk Analysis, 2002; 22(1):17-27.
23. Cox Jr, LA. Game theory and risk analysis. Risk Analysis, 2009; 29(8):1062-1068.
24. Raiffa H, Richardson J Metcalfe D. Negotiation Analysis. Harvard University Press, 2002.
25. Pinker EJ. An analysis of short-term responses to threats of terrorism. Management Science, 2007; 53(6):865-880.
26. Merrick JRW, McLay LA. Is screening cargo containers for smuggled nuclear threats worthwhile?. Decision Analysis, 2010; 7(2):155-171.
27. Parnell G, Smith C, Moxley D. Intelligent adversary risk analysis: a bioterrorism risk management model. Risk

Analysis, 2010; 30(1):32-48.
28. Harsanyi J. Subjective probability and the theory of games: Comments on Kadane and Larkey's paper. Management Science, 1982; 28(2):120-124.
29. Kadane JB, Larkey PD. Subjective probability and the theory of games. Management Science, 1982; 28(2):113-120. Reply: 124.
30. Raiffa H. The Art and Science of Negotiation. Harvard University Press, 1982.
31. Banks D, Anderson S. Game theory and risk analysis in the context of the smallpox threat. In: Wilson A, Wilson G, Olwell D, editors. Statistical Methods in Counterterrorism, 2006. p. 9-22.
32. Pate-Cornell E, Guikema S. Probabilistic modeling or terrorist threats: a systematic analysis approach to setting priorities among countermeasures. Military Operations Research, 2002; 7:5-23.
33. Rios Insua D, Ruggeri F. Robust Bayesian Analysis. Springer, 2000.
34. von Winterfeldt D, O'Sullivan TM. Should we protect commercial airplanes against surface-to-air missile attacks by terrorists?. Decision Analysis, 2006; 3(2):63-75.
35. Rios Insua D, Rios J, Banks D. Adversarial risk analysis. Journal of the American Statistical Association, 2009; 104(486):841-854.
36. Gibbons R. A Primer in Game Theory. Pearson Education Ltd., 1992.
37. Harsanyi J. Games with incomplete information played by Bayesian players, I-III. Part I. The basic model. Management Science, 1967; 14(3):159-182.
38. French S, Rios Insua D. Statistical Decision Theory. Arnold, 2000.
39. Sevillano JC, Rios J, Rios Insua D. Adversarial risk analysis: The Somali pirates case. Decision Analysis, forthcoming.
40. Zhuang J, Bier V, Alagoz O. Modeling secrecy and deception in a multiple-period attacker-defender signaling game. European Journal of Operational Research, 2010; 203(2):409-418.
41. Zhuang J, Bier V. Reasons for secrecy and deception in homeland-security resource allocation. Risk Analysis, 2010; 30(12):1737-1743.
42. Zhuang J, Bier V. Secrecy and deception at equilibrium, with applications to anti-terrorism resource allocation. Defence and Peace Economics, 2011; 22(1):43-61.
43. Harrington JE. Games, Strategies and Decision Making. Worth, 2009.
44. Aliprantis C, Chakrabarti S. Games and Decision Making. Oxford University Press, 2000.
45. Cobb B, Basuchoudhary A. A decision analysis approach to solving the signaling game. Decision Analysis, 2009; 6:239-255.
46. Merrick JRW, Parnell G. A comparative analysis of PRA and intelligent adversary methods for counterterrorism management. Risk Analysis, 2011; forthcoming.
47. Schaefer A. Inside the Terrorist Mind. Mind, 2006; 18(6):72-79.
48. Koller D, Milch B. Multi-agent influence diagrams for representing and solving games. Games and Economic Behavior, 2003; 45:181-221.
49. Bielza C, Muller P, Rios Insua D. Decision analysis by augmented probability simulation. Management Science, 1999; 45:995-1008.
50. Rios J, Rios Insua D. Supporting negotiations over influence diagrams. Decision Analysis, 2009; 6(3):153-171.
51. Rios J, Rios Insua D. Balanced increment and concession methods for negotiation support. RACSAM, 2010; 104:41-56.
52. Rothkopf M. Decision Analysis: The right tool for auctions. Decision Analysis, 2007; 4:167-172.
53. van Bingsbergen JH, Marx LM. Exploring relations between decision analysis and game theory. Decision Analysis 2007; 4:32-40.