2-2019

Deception in finitely repeated security games

Thanh H. NGUYEN

Yongzhao WANG

Arunesh SINHA
Singapore Management University, aruneshs@smu.edu.sg

Michael P. WELLMAN

Citation

NGUYEN, Thanh H.; WANG, Yongzhao; SINHA, Arunesh; and WELLMAN, Michael P.. Deception in finitely repeated security games. (2019). *Proceedings of the 33rd AAAI Conference on Artificial Intelligence (AAAI 2019), Honolulu, Hawaii, January 27 - February 1*. 2133-2140.
Available at: https://ink.library.smu.edu.sg/sis_research/4795

Deception in Finitely Repeated Security Games

Thanh H. Nguyen,[1] Yongzhao Wang,[2] Arunesh Sinha,[2] Michael P. Wellman[2]
[1]University of Oregon, thanhhng@cs.uoregon.edu
[2]University of Michigan, {wangyzh,arunesh,wellman}@umich.edu

Abstract

Allocating resources to defend targets from attack is often complicated by uncertainty about the attacker's capabilities, objectives, or other underlying characteristics. In a repeated interaction setting, the defender can collect attack data over time to reduce this uncertainty and learn an effective defense. However, a clever attacker can manipulate the attack data to mislead the defender, influencing the learning process toward its own benefit. We investigate strategic deception on the part of an attacker with private type information, who interacts repeatedly with a defender. We present a detailed computation and analysis of both players' optimal strategies given the attacker may play deceptively. Computational experiments illuminate conditions conducive to strategic deception, and quantify benefits to the attacker. By taking into account the attacker's deception capacity, the defender can significantly mitigate loss from misleading attack actions.

Introduction

Real-world security domains are often characterized by *imperfect information*: uncertainty (particularly on the defender's part) about actions taken or underlying characteristics of the opposing agent. Experience observed through repeated interaction in such domains provides an opportunity for the defender to *learn* about the behaviors and characteristics of attacker(s) (Kar et al., 2017; Gholami et al., 2017; Haghtalab et al., 2016; Nguyen et al., 2016; Xu, Tran-Thanh, and Jennings, 2016; Balcan et al., 2015; Blum, Haghtalab, and Procaccia, 2014; Marecki, Tesauro, and Segal, 2012; Letchford, Conitzer, and Munagala, 2009). For example, in wildlife protection (Fang et al., 2016), repeated interaction with poachers allows the defense authorities to observe poaching signs and patterns over time. From these observations, the defender may infer features of the poacher's capabilities and preferences, and thus design more effective patrolling strategies.

To the extent that the defender relies on data, however, the attacker may choose to modify its behavior to mislead the defender. That is, in a particular interaction the attacker may select an action that does not actually yield the best immediate reward, to avoid revealing sensitive private information. Such *deceptive* behavior could manipulate the outcome of learning to the long-term benefit of the attacker. A savvy defender, therefore, would take into account the attacker's manipulative strategy in designing his own strategy. We label an attacker strategy as *non-manipulative* if it chooses an action without regard to the defender's learning, for example if its behavior in each stage is a myopic best response. Previous work on learning in security games has generally treated the attacker as non-manipulative in this sense (Blum, Haghtalab, and Procaccia, 2014; Marecki, Tesauro, and Segal, 2012).

We study the strategic deployment of attacker deception in finitely repeated security games. We adopt an incomplete-information model, where the defender has underlying uncertainty about the attacker's type. At each time step, the defender updates his belief on the attacker's type based on attack data collected at previous steps. Based on the updated belief, the defender chooses an action to play. The attacker decides its own action, aware that the defender is collecting attack data to infer about the attacker's type. The ultimate goal of both players is to maximize expected utility accumulated over the whole time horizon. A pair of strategies that best-respond to each other, accounting for observations, constitutes a Perfect Bayesian Nash Equilibrium (PBNE).

The paper includes four main contributions. First, we present a non-linear optimization program to find a PBNE of the finitely repeated simultaneous-move game. Second, we present a result that provides an easy technique to find a sequential equilibrium of the game based on a computed PBNE. Third, we show that there exists a PBNE in which players' equilibrium strategies depend only on histories of the attacker's actions. This allows us to represent both players' strategies in a compact form, which helps in significantly speeding up the equilibrium computation of the game. Fourth, we provide a preliminary extension to the Stackelberg game (sequential move) setting.

Finally, we present a detailed experimental analysis of strategic deception, showing how various game factors affect the tendency for the attacker to deviate from myopic best responses to mislead the defender. Our results show that the defender and attacker receive significant loss and benefit respectively if the defender does not address the attacker's deception. By taking into account deceptive attacks, such loss and benefit is reduced drastically.

Related Work

Learning in security games. Most existing work on learning in security games follows a Stackelberg model and assumes the attacker plays myopically at every time step (Kar et al., 2017; Gholami et al., 2017; Haghtalab et al., 2016; Nguyen et al., 2016; Blum, Haghtalab, and Procaccia, 2014; Marecki, Tesauro, and Segal, 2012; Letchford, Conitzer, and Munagala, 2009). Balcan et al. (2015) and Xu, Tran-Thanh, and Jennings (2016) study the problem of learning with no prior knowledge of the attacker's behavior. They take a regret-minimization approach to determine the defender's strategies at each time step.

Secrecy and deception in security games. Previous work studies security scenarios in which information available to the defender and attacker is asymmetric (Guo et al., 2017; Xu et al., 2015; Rabinovich et al., 2015; Hendricks and McAfee, 2006; Brown et al., 2005; Farrell and Rabin, 1996; Zhuang, Bier, and Alagoz, 2010). The defender can exploit that information asymmetricity to strategically reveal or disguise his information to the attacker. This results in responses of the attacker which are in favor of the defender. For example, in the model of Guo et al. (2017), the defender can disguise defense resources to deceive the attacker about the defender's type. We study an opposite scenario in which the attacker acts deceptively to mislead the defender.

Repeated games with incomplete information. Previous work has studied infinitely repeated games with incomplete information (Sorin, 2002; Aumann and Maschler, 1995; Jordan, 1995; Zamir, 1992; Forges, 1988). These studies analyze properties and the convergence of players' strategies in an infinitely repeated game setting. We study the problem of one-sided incomplete information (i.e., uncertainty in the attacker's type) in finitely repeated security games.

Adversarial machine learning. There have been several studies on adversarial machine learning, attempting to investigate different attack scenarios on machine learning algorithms (Brückner, Kanzow, and Scheffer, 2012; Brückner and Scheffer, 2011; Barreno et al., 2010, 2006; Lowd and Meek, 2005). For example, *causative* attacks alters the training process by influencing the training data or *exploratory* attacks attempts to discover information about the learner and its training data. Different machine learning algorithms are then proposed which can resist these sophisticated attacks. Our work focuses on a causative attack scenario in security games. We aim at obtaining effective defense strategies which minimizes the damage of deceptive attacks in security games, given some learning outcome of attack data.

Game Model

In a finitely repeated *simultaneous-move* security game, there is a set of N targets, denoted by $\mathbf{N} = \{1, \ldots, N\}$. A defender attempts to protect these targets by allocating limited security resources over these targets. Conversely, an attacker aims at attacking these targets. We denote by $K < N$ the number of the defender's security resources. At each time step t in a finite time horizon $\mathbf{T} = \{1, \ldots, T\}$, both the defender and the attacker has to decide on which action to take. An action of the defender, $\mathbf{s}$, is an allocation of K resources over $\mathbf{N}$. We denote by $\mathbf{S}$ the set of all feasible actions of the defender. An action of the attacker is a target to attack. There is a set of attacker types $\mathbf{\Lambda} = \{1, \ldots, L\}$. Each type $\lambda \in \mathbf{\Lambda}$ has a prior probability $p^\lambda \in (0, 1)$ such that $\sum_\lambda p^\lambda = 1$. At the beginning, Nature randomly draws a type to play the game according to a prior distribution $\{p^\lambda\}_\lambda$. The attacker knows its type while the defender does not. The defender is aware of $\{p^\lambda\}_\lambda$.

Player payoffs. Each target $i \in \mathbf{N}$ is associated with rewards and penalties of the defender, $\big(R^d(i), P^d(i)\big)$, and the attacker, $\big(R^\lambda(i), P^\lambda(i)\big)$, for every type $\lambda \in \mathbf{\Lambda}$. When the attacker of type $\lambda \in \mathbf{\Lambda}$ attacks i, if the defender is protecting i, the attacker receives a penalty $P^\lambda(i)$ while the defender obtains a reward $R^d(i)$. Conversely, if the defender is not protecting target i, the attacker gets $R^\lambda(i) > P^\lambda(i)$ while the defender receives $P^d(i) < R^d(i)$.

Player observations. At $t + 1 \in \mathbf{T}$, both players observe their actions at previous time steps $\mathbf{h}_t = \{(\mathbf{s}_1, i_1), \ldots, (\mathbf{s}_t, i_t)\}$ where $\mathbf{s}_{t'}$ and $i_{t'}$ are the defender and the attacker actions respectively at time step t'. We denote by $\mathbf{H}_t$ the set of all possible histories of length t and $\mathbf{H} = \{\mathbf{H}_t\}$ (where $t = 0, \ldots, T-1$) the set of all histories. In particular, $\mathbf{H}_0 = \emptyset$. We denote by $\mathbf{h}^a_t = \{i_1, \ldots, i_t\}$ a history of the attacker's actions and $\mathbf{H}^a_t$ the set of all these attack histories.

Behavioral strategies. At each step $t + 1$, given a history $\mathbf{h}_t \in \mathbf{H}_t$, a behavioral strategy of the defender is a probability distribution $\mathbf{x}(\mathbf{h}_t) = \{x(\mathbf{s} \mid \mathbf{h}_t) : \sum_\mathbf{s} x(\mathbf{s} \mid \mathbf{h}_t) = 1, x(\mathbf{s} \mid \mathbf{h}_t) \in [0, 1], \forall \mathbf{s} \in \mathbf{S}\}$ over the defender's action set $\mathbf{S}$. $x(\mathbf{s} \mid \mathbf{h}_t)$ is the probability the defender takes action $\mathbf{s} \in \mathbf{S}$ given the history $\mathbf{h}_t$. Similarly, a behavioral strategy of the attacker of type λ is a probability distribution $\mathbf{y}^\lambda(\mathbf{h}_t) = \{y^\lambda(i \mid \mathbf{h}_t) : \sum_i y^\lambda(i \mid \mathbf{h}_t) = 1, y^\lambda(i \mid \mathbf{h}_t) \in [0, 1], \forall i \in \mathbf{N}\}$ over the attacker's actions $\mathbf{N}$. $y^\lambda(i \mid \mathbf{h}_t)$ is the probability the attacker of type λ attacks target i given $\mathbf{h}_t$. We denote by $\mathbf{x} = \{\mathbf{x}(\mathbf{h}_t)\}$ and $\mathbf{y}^\lambda = \{\mathbf{y}^\lambda(\mathbf{h}_t)\}$ strategies of the defender and attacker of type λ over all $\mathbf{h}_t \in \mathbf{H}$ respectively. Finally, $\mathbf{X}$ and $\mathbf{Y} = \{\mathbf{Y}^\lambda\}$ denote the sets of all strategies $\mathbf{x}$ and $\mathbf{y}^\lambda$ respectively.

Player expected utilities. Let $\mathbf{x}$ and $\mathbf{y} = \{\mathbf{y}^\lambda\}$ be the defender and attacker's behavioral strategies respectively. At each $t + 1$, the defender can update his belief on the attacker types using the Bayes rule, which is formulated as:

$$p(\lambda \mid \mathbf{h}_t) \propto p^\lambda \prod\nolimits_{t'=1}^{t} y^\lambda(i_{t'} \mid \mathbf{h}_{t'-1})$$

where $\mathbf{h}_0 = \emptyset, \mathbf{h}_{t'} = \{\mathbf{h}_{t'-1}, i_{t'}\}$. Let:

$$EU^d_i(\mathbf{x}, \mathbf{h}_t) = \Big[\sum\nolimits_{\mathbf{s}:i\in\mathbf{s}} x(\mathbf{s} \mid \mathbf{h}_t)\Big] \big(R^d(i) - P^d(i)\big) + P^d(i)$$

$$EU^\lambda_i(\mathbf{x}, \mathbf{h}_t) = \Big[\sum\nolimits_{\mathbf{s}:i\in\mathbf{s}} x(\mathbf{s} \mid \mathbf{h}_t)\Big] \big(P^\lambda(i) - R^\lambda(i)\big) + R^\lambda(i)$$

be immediate expected utilities of the defender and the attacker of type λ respectively at target i at step $t+1$ given the defender plays $\mathbf{x}(\mathbf{h}_t)$. Based on the immediate expected utilities at every target, the players' total expected utilities over $\mathbf{T}$ can be computed using backward induction as follows.

At the last time step T, the total expected utilities of the defender and the attacker with respect to history $\mathbf{h}_{T-1} \in$

$\mathbf{H}_{T-1}$ is equal to their immediate expected utilities at $\mathbf{h}_{T-1}$:

$$U_T^d(\mathbf{x}, \mathbf{y} \mid \mathbf{h}_{T-1}) = \sum_{\lambda,i} p(\lambda \mid \mathbf{h}_{T-1}) y^\lambda(i \mid \mathbf{h}_{T-1}) EU_i^d(\mathbf{x}, \mathbf{h}_{T-1})$$

$$U_T^\lambda(\mathbf{x}, \mathbf{y} \mid \mathbf{h}_{T-1}) = \sum\nolimits_i y^\lambda(i \mid \mathbf{h}_{T-1}) EU_i^\lambda(\mathbf{x}, \mathbf{h}_{T-1}), \forall\lambda$$

At time step $t+1 < T$, the total expected utilities of both players with respect to history $\mathbf{h}_t \in \mathbf{H}_t$ consists of (i) the immediate expected utility at $t+1$; and (ii) the future expected utility after $t+1$. These utilities are formulated as:

$$U_{t+1}^d(\mathbf{x}, \mathbf{y} \mid \mathbf{h}_t) = \sum\nolimits_{\lambda,i} p(\lambda \mid \mathbf{h}_t) y^\lambda(i \mid \mathbf{h}_t) EU_i^d(\mathbf{x}, \mathbf{h}_t)$$

$$+ \sum\nolimits_{\mathbf{s},\lambda,i} x(\mathbf{s} \mid \mathbf{h}_t) p(\lambda \mid \mathbf{h}_t) y^\lambda(i \mid \mathbf{h}_t) U_{t+2}^d(\mathbf{x}, \mathbf{y} \mid \mathbf{h}_t, (\mathbf{s}, i))$$

$$U_{t+1}^\lambda(\mathbf{x}, \mathbf{y} \mid \mathbf{h}_t) = \sum\nolimits_i y^\lambda(i \mid \mathbf{h}_t) EU_i^\lambda(\mathbf{x}, \mathbf{h}_t)$$

$$+ \sum\nolimits_{\mathbf{s},i} x(\mathbf{s} \mid \mathbf{h}_t) y^\lambda(i \mid \mathbf{h}_t) U_{t+2}^\lambda(\mathbf{x}, \mathbf{y} \mid \mathbf{h}_t, (\mathbf{s}, i)), \forall\lambda$$

Player's goals. Given any history $\mathbf{h}_t$, both players aim at choosing strategies $\mathbf{x}(\mathbf{h}_t), \{\mathbf{y}^\lambda(\mathbf{h}_t)\}$ that maximize their total expected utility at $\mathbf{h}_t$. In this scenario, the attacker is no longer myopic; it has to reason about all future possibilities to decide on which behavioral strategy to play at each $\mathbf{h}_t$. Such attack strategies (which may not be myopically optimal) are chosen to mislead the defender about the attacker's type, ultimately benefiting the attacker in future steps. These optimal behavioral strategies of players form a PBNE.

Game Equilibria

Definition 1 (PBNE). *Behavioral strategies of the defender $\mathbf{x}^*$ and attacker $\mathbf{y}^*$ form a PBNE of the game if and only if for every $\mathbf{h}_t \in \mathbf{H}_t$ that occurs, we have:*

- *$\mathbf{x}^*$ is the best response of the defender:*

$$U_{t+1}^d(\mathbf{x}^*, \mathbf{y}^* \mid \mathbf{h}_t) \geq U_{t+1}^d(\mathbf{x}, \mathbf{y}^* \mid \mathbf{h}_t), \forall \mathbf{x} \in \mathbf{X}$$

- *$\mathbf{y}^{\lambda,*}$ is the best response of the attacker type λ:*

$$U_{t+1}^\lambda(\mathbf{x}^*, \mathbf{y}^* \mid \mathbf{h}_t) \geq U_{t+1}^\lambda(\mathbf{x}^*, \mathbf{y} \mid \mathbf{h}_t), \forall \mathbf{y} \in \mathbf{Y}$$

Since the action sets of both players are finite, there always exists a PBNE of the game. Our first result extends a given PBNE to a refined sequential equilibrium.

Theorem 1. *For each PBNE, there is a sequential equilibrium in which players' strategies are identical to the ones in the PBNE at histories that occur with a positive probability.*

Proof. We denote by $(\mathbf{x}, \mathbf{y})$ a PBNE of the game. We follow the trembling-hand approach to find a corresponding sequential equilibrium of $(\mathbf{x}, \mathbf{y})$. Let $r^d(\mathbf{h}_t)$ and $r^\lambda(\mathbf{h}_t)$ be the ratios of the number of zero probabilities to the number of non-zero probabilities in $\mathbf{x}(\mathbf{h}_t)$ and $\mathbf{y}^\lambda(\mathbf{h}_t)$ respectively. For each $\epsilon > 0$, we construct a new fully mixed behavioral strategy of the defender and the attacker, $(\mathbf{x}_\epsilon, \mathbf{y}_\epsilon)$, as follows:

$$x_\epsilon(\mathbf{s} \mid \mathbf{h}_t) = \epsilon, \text{ if } x(\mathbf{s} \mid \mathbf{h}_t) = 0$$

$$x_\epsilon(\mathbf{s} \mid \mathbf{h}_t) = x(\mathbf{s} \mid \mathbf{h}_t) - \epsilon * r^d(\mathbf{h}_t), \text{ if } x(\mathbf{s} \mid \mathbf{h}_t) > 0$$

$$y_\epsilon^\lambda(i \mid \mathbf{h}_t) = \epsilon, \text{ if } y^\lambda(i \mid \mathbf{h}_t) = 0$$

$$y_\epsilon^\lambda(i \mid \mathbf{h}_t) = y^\lambda(i \mid \mathbf{h}_t) - \epsilon * r^\lambda(\mathbf{h}_t), \text{ if } y^\lambda(i \mid \mathbf{h}_t) > 0$$

ϵ is chosen to be small enough such that all resulting probabilities are positive. We use the breath-first search (according to the time horizon $\mathbf{T}$) to examine the whole history set. When encountering a history with a zero probability of occurrence $p(\mathbf{h}_t) = 0$ according to $(\mathbf{x}, \mathbf{y})$, we construct a new belief of the defender over attacker types at $\mathbf{h}_t$ as follows:

$$p'(\lambda \mid \mathbf{h}_t) = \lim_{\epsilon \to 0} p_\epsilon(\lambda \mid \mathbf{h}_t) = \lim_{\epsilon \to 0} \frac{p^\lambda \prod_{t'} y_\epsilon^\lambda(i_{t'} \mid \mathbf{h}_{t'-1})}{\sum_{\lambda'} p^{\lambda'} \prod_{t'} y_\epsilon^{\lambda'}(i_{t'} \mid \mathbf{h}_{t'-1})}$$

We find a PBNE of the corresponding sub-game starting from this history $\mathbf{h}_t$ with this new belief. We then replace the strategies of the sub-game in $(\mathbf{x}, \mathbf{y})$ with these new equilibrium strategies. We also update $(\mathbf{x}_\epsilon, \mathbf{y}_\epsilon)$ accordingly with the updated strategies of the sub-game. This process will continue until all histories are examined. The resulting strategies $(\mathbf{x}', \mathbf{y}')$ with belief $\{p'(\lambda \mid \mathbf{h}_t)\}$ belong to a sequential equilibrium of the game. Indeed, it is straightforward to prove:

$$x'(\mathbf{s} \mid \mathbf{h}_t) = \lim_{\epsilon \to 0} x_\epsilon(\mathbf{s} \mid \mathbf{h}_t)$$

$$y^{\lambda,'}(i \mid \mathbf{h}_t) = \lim_{\epsilon \to 0} y_\epsilon^\lambda(i \mid \mathbf{h}_t)$$

$$p'(\lambda \mid \mathbf{h}_t) = \lim_{\epsilon \to 0} p_\epsilon(\lambda \mid \mathbf{h}_t) \text{ (by definition)}$$

Furthermore, the updating process only replaces strategies in $(\mathbf{x}, \mathbf{y})$ at histories $\mathbf{h}_t$ with a zero-probability of occurrence by a PBNE of the sub-game at $\mathbf{h}_t$ respective to the belief $p'(\lambda \mid \mathbf{h}_t)$. Therefore, $(\mathbf{x}', \mathbf{y}')$ is the best response of the players at every $\mathbf{h}_t$ according to $p'(\lambda \mid \mathbf{h}_t)$. □

Next, we present a result that enables a compact representation of the game. We denote by $\mathbf{X}^a$ a subset of behavioral strategies of the defender in which all the strategies are independent of histories of the defender's actions. In other words, for all $\mathbf{x} \in \mathbf{X}^a$, $x(\mathbf{s} \mid \mathbf{h}_t) = x(\mathbf{s} \mid \mathbf{h}_t^a)$ for every history $\mathbf{h}_t$ where $\mathbf{h}_t^a$ is the corresponding history of attacker actions. Similarly, $\mathbf{Y}^a$ is a subset of behavioral strategies of the attacker.

Theorem 2. *There exists a PBNE of the game in which the equilibrium strategies of the players only depend on the histories of the attacker's actions.*

Proof. We use Brouwer's fixed-point theorem and a backward induction method. We are going to show that there exists a PBNE $\mathbf{x}^* \in \mathbf{X}^a$ and $\mathbf{y}^* \in \mathbf{Y}^a$ such that: $\forall \mathbf{h}_t$

$$U_{t+1}^d(\mathbf{x}^*, \mathbf{y}^* \mid \mathbf{h}_t) \geq U_{t+1}^d(\mathbf{x}, \mathbf{y}^* \mid \mathbf{h}_t), \forall \mathbf{x} \in \mathbf{X} \quad (1)$$

$$U_{t+1}^\lambda(\mathbf{x}^*, \mathbf{y}^* \mid \mathbf{h}_t) \geq U_{t+1}^\lambda(\mathbf{x}^*, \mathbf{y} \mid \mathbf{h}_t), \forall \mathbf{y} \in \mathbf{Y} \quad (2)$$

We denote by $\mathbf{x}\{\mathbf{x}(\mathbf{h}_t) \leftarrow \mathbf{s}\}$ the defender strategy obtained by replacing $\mathbf{x}(\mathbf{h}_t)$ in $\mathbf{x}$ by a defense action $\mathbf{s}$. Similarly, $\mathbf{y}^\lambda\{\mathbf{y}^\lambda(\mathbf{h}_t) \leftarrow i\}$ is the attacker strategy of type λ obtained by replacing $\mathbf{y}^\lambda(\mathbf{h}_t)$ in $\mathbf{y}^\lambda$ by an attack action i. We define:

$$\phi_\mathbf{s}^d(\mathbf{x}, \mathbf{y} \mid \mathbf{h}_t) = \max\{0, U_{t+1}^d(\mathbf{x}\{\mathbf{x}(\mathbf{h}_t) \leftarrow \mathbf{s}\}, \mathbf{y} \mid \mathbf{h}_t) - U_{t+1}^d(\mathbf{x}, \mathbf{y} \mid \mathbf{h}_t)\}$$

$$\phi_i^\lambda(\mathbf{x}, \mathbf{y} \mid \mathbf{h}_t) = \max\{0, U_{t+1}^\lambda(\mathbf{x}, \mathbf{y}\{\mathbf{y}^\lambda(\mathbf{h}_t) \leftarrow i\}) \mid \mathbf{h}_t) - U_{t+1}^\lambda(\mathbf{x}, \mathbf{y} \mid \mathbf{h}_t)\}$$

which are non-negative continuous functions in $(\mathbf{x}, \mathbf{y})$. We define a function $F : (\mathbf{X}^a, \mathbf{Y}^a) \to (\mathbf{X}^a, \mathbf{Y}^a)$ as follows:

$F(\mathbf{x},\mathbf{y}) = (\mathbf{x}',\mathbf{y}')$ where

$$x'(\mathbf{s} \mid \mathbf{h}_t) = \frac{x(\mathbf{s} \mid \mathbf{h}_t) + \phi^d_{\mathbf{s}}(\mathbf{x},\mathbf{y} \mid \mathbf{h}_t)}{\sum_{\mathbf{s}'} x(\mathbf{s}' \mid \mathbf{h}_t) + \phi^d_{\mathbf{s}'}(\mathbf{x},\mathbf{y} \mid \mathbf{h}_t)}$$

$$y^{\lambda,'}(i \mid \mathbf{h}_t) = \frac{y^\lambda(i \mid \mathbf{h}_t) + \phi^\lambda_i(\mathbf{x},\mathbf{y} \mid \mathbf{h}_t)}{\sum_j y^\lambda(j \mid \mathbf{h}_t) + \phi^\lambda_j(\mathbf{x},\mathbf{y} \mid \mathbf{h}_t)}$$

Since F is continuous over a convex and compact set $(\mathbf{X}^a, \mathbf{Y}^a)$, there exists $(\mathbf{x}^*, \mathbf{y}^*)$ such that $F(\mathbf{x}^*, \mathbf{y}^*) = (\mathbf{x}^*, \mathbf{y}^*)$ according to the Brouwer's fixed point theorem. On the other hand, according to the linearity of expectation, there must be an action $\mathbf{s}$ such that $U^d_{t+1}(\mathbf{x}^*\{\mathbf{x}^*(\mathbf{h}_t) \leftarrow \mathbf{s}\}, \mathbf{y}^* \,|\, \mathbf{h}_t) - U^d_{t+1}(\mathbf{x}^*, \mathbf{y}^* \,|\, \mathbf{h}_t) \leq 0$, meaning that $\phi^d_{\mathbf{s}}(\mathbf{x}, \mathbf{y} \mid \mathbf{h}_t) = 0$. Therefore, we have:

$$x^*(\mathbf{s} \mid \mathbf{h}_t) = \frac{x^*(\mathbf{s} \mid \mathbf{h}_t)}{1 + \sum_{\mathbf{s}'} \phi^d_{\mathbf{s}'}(\mathbf{x}^*, \mathbf{y}^* \mid \mathbf{h}_t)}$$

This implies $\phi^d_{\mathbf{s}'}(\mathbf{x}^*, \mathbf{y}^* \mid \mathbf{h}_t) = 0, \forall \mathbf{s}'$. Similarly, $\phi^\lambda_i(\mathbf{x}^*, \mathbf{y}^* \mid \mathbf{h}_t) = 0, \forall i$. As a result, we obtain:

$$U^d_{t+1}(\mathbf{x}^*\{\mathbf{x}^*(\mathbf{h}_t) \leftarrow \mathbf{s}\}, \mathbf{y}^* \,|\, \mathbf{h}_t) \leq U^d_{t+1}(\mathbf{x}^*, \mathbf{y}^* \,|\, \mathbf{h}_t) \quad (3)$$

$$U^\lambda_{t+1}(\mathbf{x}^*, \mathbf{y}^*\{\mathbf{y}^{\lambda,*}(\mathbf{h}_t) \leftarrow i\}) \,|\, \mathbf{h}_t) \leq U^\lambda_{t+1}(\mathbf{x}^*, \mathbf{y}^* \,|\, \mathbf{h}_t) \quad (4)$$

for all $\mathbf{s}$ and i and $\mathbf{h}_t$. While the above result holds for any convex and compact $(\mathbf{X}^a, \mathbf{Y}^a)$, showing that the fixed point profile provides more utility than any deviation to $\mathbf{X}$ or $\mathbf{Y}$ requires $(\mathbf{X}^a, \mathbf{Y}^a)$ to depend on the attacker's past actions. This dependence is due to the dependence of the posterior belief on the attacker's past actions, as can be seen in the next steps of the proof. Based on the above inequalities, we show that $(\mathbf{x}^*, \mathbf{y}^*)$ satisfy (1–2) using backward induction.

At last time step T, for every $\mathbf{x} \in \mathbf{X}$, we have:

$$\begin{aligned}
&U^d_T(\mathbf{x}^*, \mathbf{y}^* \mid \mathbf{h}_{T-1}) \\
&\geq \sum\nolimits_{\mathbf{s}} x(\mathbf{s} \,|\, \mathbf{h}_{T-1}) U^d_T(\mathbf{x}^*\{\mathbf{x}^*(\mathbf{h}_{T-1}) \leftarrow \mathbf{s}\}, \mathbf{y}^* \,|\, \mathbf{h}_{T-1}) \\
&= U^d_T(\mathbf{x}, \mathbf{y}^* \mid \mathbf{h}_{T-1})
\end{aligned}$$

Therefore, $\mathbf{x}^*$ is the defender best response against the attacker's strategy $\mathbf{y}^*$ at time step T.

At time step $t + 1 < T$, suppose that (1–2) hold true for all $t' > t + 1$, then for every $\mathbf{x} \in \mathbf{X}$, we have:

$$\begin{aligned}
&U^d_{t+1}(\mathbf{x}^*, \mathbf{y}^* \mid \mathbf{h}_t) \\
&\geq \sum\nolimits_{\mathbf{s}} x(\mathbf{s} \mid \mathbf{h}_t) U^d_{t+1}(\mathbf{x}^*\{\mathbf{x}^*(\mathbf{h}_t) \leftarrow \mathbf{s}\}, \mathbf{y}^* \mid \mathbf{h}_t) \\
&= \sum\nolimits_{\lambda,i} p(\lambda \,|\, \mathbf{h}_t) y^{\lambda,*}(i \,|\, \mathbf{h}_t) EU^d_i(\mathbf{x}, \mathbf{h}_t) + \sum\nolimits_{\mathbf{s}} x(\mathbf{s} \mid \mathbf{h}_t) \\
&\qquad \times \sum\nolimits_{i,\lambda} p(\lambda \mid \mathbf{h}_t) y^{\lambda,*}(i \mid \mathbf{h}_t) U^d_{t+2}(\mathbf{x}^*, \mathbf{y}^* \mid \mathbf{h}_t, (\mathbf{s}, i)) \\
&\geq \sum\nolimits_{\lambda,i} p(\lambda \,|\, \mathbf{h}_t) y^{\lambda,*}(i \,|\, \mathbf{h}_t) EU^d_i(\mathbf{x}, \mathbf{h}_t) + \sum\nolimits_{\mathbf{s}} x(\mathbf{s} \mid \mathbf{h}_t) \\
&\qquad \times \sum\nolimits_{i,\lambda} p(\lambda \mid \mathbf{h}_t) y^{\lambda,*}(i \mid \mathbf{h}_t) U^d_{t+2}(\mathbf{x}, \mathbf{y}^* \mid \mathbf{h}_t, (\mathbf{s}, i)) \\
&= U^d_{t+1}(\mathbf{x}, \mathbf{y}^* \mid \mathbf{h}_t)
\end{aligned}$$

Therefore, $\mathbf{x}^*$ is the defender best response against the attacker's strategy $\mathbf{y}^*$ at $t + 1$. Similarly, $\mathbf{y}^*$ is the attacker's best response against $\mathbf{x}^*$ at all time steps. □

Equilibrium Computation

Based on Theorem 2, in computing a PBNE, we only need to search over the strategy sets $(\mathbf{X}^a, \mathbf{Y}^a)$. We also only need to consider attack histories $\{\mathbf{h}^a_t\}$. We can now represent the defender behavioral strategies as compact marginal coverage probabilities over targets. We overload the notation $\mathbf{x}(\mathbf{h}^a_t) = \{x(i \mid \mathbf{h}^a_t)\}$ where $x(i \mid \mathbf{h}^a_t)$ is the defender's coverage probability at target i at history $\mathbf{h}^a_t$ such that $\sum_i x(i \mid \mathbf{h}^a_t) \leq K$ and $x(i \mid \mathbf{h}^a_t) \in [0, 1]$ for all $i \in \mathbf{N}$. In particular, $x(i \mid \mathbf{h}^a_t) = \sum_{\mathbf{s}: i \in \mathbf{s}} x(\mathbf{s} \mid \mathbf{h}^a_t)$. The players' immediate and total expected utilities can be reformulated accordingly as follows:

$$EU^d_i(\mathbf{x}, \mathbf{h}^a_t) = x(i \mid \mathbf{h}^a_t)(R^d(i) - P^d(i)) + P^d(i) \quad (5)$$

$$EU^\lambda_i(\mathbf{x}, \mathbf{h}^a_t) = x(i \mid \mathbf{h}^a_t)(P^\lambda(i) - R^\lambda(i)) + R^\lambda(i) \quad (6)$$

$$\begin{aligned}
U^d_{t+1}(\mathbf{x}, \mathbf{y} \mid \mathbf{h}^a_t) &= \sum\nolimits_{\lambda,i} p(\lambda \mid \mathbf{h}^a_t) y^\lambda(i \mid \mathbf{h}^a_t) EU^d_i(\mathbf{x}, \mathbf{h}^a_t) \\
&\quad + \sum\nolimits_{i,\lambda} p(\lambda \mid \mathbf{h}^a_t) y^\lambda(i \mid \mathbf{h}^a_t) U^d_{t+2}(\mathbf{x}, \mathbf{y} \mid \mathbf{h}^a_t, i) \quad (7)
\end{aligned}$$

$$\begin{aligned}
U^\lambda_{t+1}(\mathbf{x}, \mathbf{y} \mid \mathbf{h}^a_t) &= \sum\nolimits_i y^\lambda(i \mid \mathbf{h}^a_t) EU^\lambda_i(\mathbf{x}, \mathbf{h}^a_t) \\
&\quad + \sum\nolimits_i y^\lambda(i \mid \mathbf{h}^a_t) U^\lambda_{t+2}(\mathbf{x}, \mathbf{y} \mid \mathbf{h}^a_t, i) \quad (8)
\end{aligned}$$

Note that the total expected utilities of players at last time step $U^d_T(\mathbf{x}, \mathbf{y} \mid \mathbf{h}^a_{T-1})$ and $U^\lambda_T(\mathbf{x}, \mathbf{y} \mid \mathbf{h}^a_{T-1})$ do not have the second term (which represents the future expected utility) as in Equations (7–8). In the following, we present a backward-induction based method to find the attacker (defender) best response against a fixed behavioral strategy of the defender (attacker). We then introduce a program to compute a PBNE based on these best-response solutions.

Attacker best response

Given a defender's strategy $\mathbf{x}$, we can compute a best response of the attacker of type λ using backward induction:

At last time step T. Given a history $\mathbf{h}^a_{T-1}$, finding a best response of the attacker type λ against $\mathbf{x}$ is formulated as the following linear program:

$$\max_{\mathbf{y}^\lambda(\mathbf{h}^a_{T-1})} \sum\nolimits_i y^\lambda(i \mid \mathbf{h}^a_{T-1}) EU^\lambda_i(\mathbf{x}, \mathbf{h}^a_{T-1}) \quad (9)$$

$$\text{s.t. } \sum\nolimits_i y^\lambda(i \,|\, \mathbf{h}^a_{T-1}) = 1, y^\lambda(i \,|\, \mathbf{h}^a_{T-1}) \geq 0, \forall i \quad (10)$$

which maximizes the attacker's total expected utility at $\mathbf{h}^a_{T-1}$. Its corresponding dual program:

$$\min_{v^\lambda(\mathbf{h}^a_{T-1})} \; v^\lambda(\mathbf{h}^a_{T-1}) \quad (11)$$

$$\text{s.t. } v^\lambda(\mathbf{h}^a_{T-1}) \geq EU^\lambda_i(\mathbf{x}, \mathbf{h}^a_{T-1}), \forall i \in \mathbf{N}. \quad (12)$$

According to complementary slackness, any optimal primal and dual solutions $(\mathbf{y}^{\lambda,*}(\mathbf{h}^a_{T-1}), v^\lambda_*(\mathbf{h}^a_{T-1}))$ satisfies: $\forall i$

$$y^{\lambda,*}(i \mid \mathbf{h}^a_{T-1}) \left[v^\lambda_*(\mathbf{h}^a_{T-1}) - EU^\lambda_i(\mathbf{x}, \mathbf{h}^a_{T-1})\right] = 0 \quad (13)$$

At time step $t + 1 < T$. Given a history $\mathbf{h}^a_t$, finding a best response of the attacker type λ is formulated as:

$$\max_{\mathbf{y}^\lambda(\mathbf{h}^a_t)} \sum\nolimits_i y^\lambda(i \,|\, \mathbf{h}^a_t) \left[EU^\lambda_i(\mathbf{x}, \mathbf{h}^a_t) + v^\lambda_*(\mathbf{h}^a_t, i)\right] \quad (14)$$

$$\text{s.t. } \sum\nolimits_i y^\lambda(i \,|\, \mathbf{h}^a_t) = 1, y^\lambda(i \,|\, \mathbf{h}^a_t) \geq 0, \forall i \in \mathbf{N}. \quad (15)$$

which maximizes the attacker's total expected utility at $\mathbf{h}^a_t$ where $v^\lambda_*(\mathbf{h}^a_t, i)$ is the attacker's optimal total expected utility at $(\mathbf{h}^a_t, i)$. Its corresponding dual program:

$$\min_{v^\lambda(\mathbf{h}^a_t)} v^\lambda(\mathbf{h}^a_t) \tag{16}$$

$$\text{s.t. } v^\lambda(\mathbf{h}^a_t) \geq EU^\lambda_i(\mathbf{x}, \mathbf{h}^a_t) + v^\lambda_*(\mathbf{h}^a_t, i), \forall i \tag{17}$$

According to complementary slackness: any optimal solution $(\mathbf{y}^{\lambda,*}(\mathbf{h}^a_t), v^\lambda_*(\mathbf{h}^a_t))$ must satisfy: $\forall i$:

$$y^{\lambda,*}(i \,|\, \mathbf{h}_t)\left[v^\lambda_*(\mathbf{h}_t) - EU^\lambda_i(\mathbf{x}, \mathbf{h}_t) - v^\lambda_*(\mathbf{h}_t, i)\right] = 0 \tag{18}$$

Defender best response

Given an attack strategy $\mathbf{y}$, we denote by:

$$\bar{p}(\lambda \mid \mathbf{h}^a_t) = p^\lambda \prod_{t'=1}^{t} y^\lambda(i_{t'} \mid \mathbf{h}^a_{t'-1})$$

Then the defender's belief on attacker types at each history $\mathbf{h}^a_t$ can be computed as follows:

$$p(\lambda \mid \mathbf{h}^a_t) = \frac{\bar{p}(\lambda \mid \mathbf{h}^a_t)}{\sum_{\lambda'} \bar{p}(\lambda' \mid \mathbf{h}^a_t)} \tag{19}$$

Similar to the computation of a best response of attacker, we can compute a best response of the defender against an attack strategy $\mathbf{y}$ using backward induction as follows:

At last time step T. Given a history $\mathbf{h}^a_{T-1}$, finding a best response $\mathbf{x}(\mathbf{h}^a_{T-1})$ can be formulated as:

$$\max \sum_\lambda \bar{p}(\lambda \,|\, \mathbf{h}^a_{T\text{-}1}) \sum_i y^\lambda(i \,|\, \mathbf{h}_{T\text{-}1}) EU^d_i(\mathbf{x}, \mathbf{h}^a_{T\text{-}1}) \tag{20}$$

$$\text{s.t. } \sum_i x(i \,|\, \mathbf{h}_{T\text{-}1}) \leq K, x(i \,|\, \mathbf{h}_{T\text{-}1}) \in [0,1], \forall i. \tag{21}$$

Proposition 1. *For every attack history $\mathbf{h}^a_{T-1}$, we denote by $v^d_*(\mathbf{h}^a_{T-1})$ the defender's optimal total expected utility against the attacker's strategy $\mathbf{y}$ at $\mathbf{h}^a_{T-1}$. Then:*

$$v^d_*(\mathbf{h}^a_{T-1}) = \frac{\bar{v}^d_*(\mathbf{h}^a_{T-1})}{\sum_{\lambda'} \bar{p}(\lambda' \mid \mathbf{h}^a_{T-1})}$$

The proof of Proposition 1 is in the appendix.[1] Here, $\bar{v}^d_*(\mathbf{h}^a_{T-1})$ the optimal objective of (20–21). By removing the constant $P^d(i)$ in $EU^d_i(\mathbf{x}, \mathbf{h}^a_{T-1})$ (Equation 5) and taking the dualty, we obtain the corresponding dual program:

$$\min\ K\bar{v}^d(0 \mid \mathbf{h}^a_{T-1}) + \sum_i \bar{v}^d(i \mid \mathbf{h}^a_{T-1}) \tag{22}$$

$$\text{s.t. } \bar{v}^d(0 \mid \mathbf{h}^a_{T-1}) \geq 0, \bar{v}^d(i \mid \mathbf{h}^a_{T-1}) \geq 0, \forall i \tag{23}$$

$$\bar{v}^d(0 \mid \mathbf{h}^a_{T-1}) + \bar{v}^d(i \mid \mathbf{h}^a_{T-1}) \geq \tag{24}$$

$$\sum_\lambda \bar{p}(\lambda \mid \mathbf{h}^a_{T\text{-}1}) y^\lambda(i \mid \mathbf{h}^a_{T\text{-}1}) \left(R^d(i) - P^d(i)\right), \forall i.$$

According to complementary slackness, any optimal solutions $(\mathbf{x}^*(\mathbf{h}^a_{T-1}), \{\bar{v}^d_*(i \,|\, \mathbf{h}^a_{T-1})\}, \bar{v}^d_*(0 \,|\, \mathbf{h}^a_{T-1}))$ satisfies: $\forall i$

$$x^*(i \mid \mathbf{h}^a_{T-1})[\bar{v}^d_*(0 \mid \mathbf{h}^a_{T-1}) + \bar{v}^d_*(i \mid \mathbf{h}^a_{T-1}) \tag{25}$$

$$- \sum_\lambda \bar{p}(\lambda \,|\, \mathbf{h}^a_{T-1}) y^\lambda(i \,|\, \mathbf{h}^a_{T-1}) \left(R^d(i) - P^d(i)\right)] = 0$$

$$\bar{v}^d_*(i \mid \mathbf{h}^a_{T-1}) \left[x^*(i \mid \mathbf{h}^a_{T-1}) - 1\right] = 0 \tag{26}$$

$$\bar{v}^d_*(0 \mid \mathbf{h}^a_{T-1}) \left[\sum_j x^*(j \mid \mathbf{h}^a_{T-1}) - K\right] = 0 \tag{27}$$

At time step $t + 1 < T$. Given a history $\mathbf{h}^a_t$, finding an optimal behavioral strategy $\mathbf{x}(\mathbf{h}^a_t)$ can be formulated as the following program:

$$\max_{\mathbf{x}(\mathbf{h}^a_t)} \sum_{\lambda, i} \bar{p}(\lambda \mid \mathbf{h}^a_t) y^\lambda(i \mid \mathbf{h}^a_t) EU^d_i(\mathbf{x}, \mathbf{h}^a_t) \tag{28}$$

$$+ \sum_i \bar{v}^d_*(\mathbf{h}^a_t, i)$$

$$\text{s.t. } \sum_i x(i \mid \mathbf{h}^a_t) \leq K, x(i \mid \mathbf{h}^a_t) \in [0,1], \forall i \tag{29}$$

We denote by $\bar{v}^d_*(\mathbf{h}^a_t)$ the optimal objective of (28–29) at $\mathbf{h}_t$. In (28), $\bar{v}^d_*(\mathbf{h}^a_t, i)$ is the optimal objective of this primal program (28–29) but with respect to the history $(\mathbf{h}^a_t, i)$.

Proposition 2. *For every attack history $\mathbf{h}^a_t$, we denote by $v^d_*(\mathbf{h}^a_t)$ the defender's optimal total expected utility against the attacker's strategy $\mathbf{y}$ at $\mathbf{h}^a_t$. Then:*

$$v^d_*(\mathbf{h}^a_t) = \frac{\bar{v}^d_*(\mathbf{h}^a_t)}{\sum_{\lambda'} \bar{p}(\lambda' \mid \mathbf{h}^a_t)}$$

The proof of Proposition 2 is in the appendix. In (28), the term $\sum_i \bar{v}^d_*(\mathbf{h}^a_t, i)$ and the term $P^d(i)$ in $EU^d_i(\mathbf{x}, \mathbf{h}^a_t)$ are constant. By removing these constants and taking the dual, we obtain the corresponding dual program:

$$\min\ K\bar{v}^d(0 \mid \mathbf{h}^a_t) + \sum_i \bar{v}^d(i \mid \mathbf{h}^a_t) \tag{30}$$

$$\text{s.t. } \bar{v}^d(0 \mid \mathbf{h}^a_t) \geq 0, \bar{v}^d(i \mid \mathbf{h}^a_t) \geq 0, \forall i \in \mathbf{N} \tag{31}$$

$$\bar{v}^d(0 \mid \mathbf{h}^a_t) + \bar{v}^d(i \mid \mathbf{h}^a_t) \geq \tag{32}$$

$$\sum_\lambda \bar{p}(\lambda \mid \mathbf{h}_t) y^\lambda(i \mid \mathbf{h}^a_t) \left(R^d(i) - P^d(i)\right), \forall i.$$

According to complementary slackness, any optimal solution $(\mathbf{x}^*(\mathbf{h}^a_t), \{\bar{v}^d_*(i \mid \mathbf{h}^a_t)\}, \bar{v}^d_*(0 \mid \mathbf{h}^a_t))$ satisfies: $\forall i$:

$$x^*(i \mid \mathbf{h}^a_t)\left[\bar{v}^d_*(0 \mid \mathbf{h}^a_t) + \bar{v}^d_*(i \mid \mathbf{h}^a_t) - \right. \tag{33}$$

$$\left. \sum_\lambda \bar{p}(\lambda \mid \mathbf{h}^a_t) y^\lambda(i \mid \mathbf{h}^a_t) \left(R^d(i) - P^d(i)\right)\right] = 0$$

$$\bar{v}^d_*(i \mid \mathbf{h}^a_t) \left[x^*(i \mid \mathbf{h}^a_t) - 1\right] = 0 \tag{34}$$

$$\bar{v}^d_*(0 \mid \mathbf{h}^a_t) \left[\sum_i x^*(i \mid \mathbf{h}^a_t) - K\right] = 0 \tag{35}$$

Equilibrium computation program

Based on the computation of players' best responses, a pair of behavioral strategies $(\mathbf{x}, \mathbf{y})$ forms a PBNE if and only if these strategies satisfy (i) the feasibility constraints (21,23,24,29,31,32) and (10,12,15,17); and (ii) the complementary slackness constraints (25–27, 33–35) and (13, 18). Since finding strategies which satisfy these slackness constraints is not straightforward, we convert the problem of finding a PBNE into the following program:

$$\min \delta \quad \text{such that } \forall i, \mathbf{h}^a_t: \tag{36}$$

$$\delta \geq y^\lambda(i \mid \mathbf{h}^a_t) \left[v^\lambda(\mathbf{h}^a_t) - EU^\lambda_i(\mathbf{x}, \mathbf{h}^a_t) - v^\lambda(\mathbf{h}^a_t, i)\right] \tag{37}$$

$$\delta \geq x(i \mid \mathbf{h}^a_t)\left[\bar{v}^d(0 \mid \mathbf{h}^a_t) + \bar{v}^d(i \mid \mathbf{h}^a_t)\right. \tag{38}$$

$$\left. - \sum_\lambda \bar{p}(\lambda \mid \mathbf{h}^a_t) y^\lambda(i \mid \mathbf{h}^a_t) \left(R^d(i) - P^d(i)\right)\right]$$

$$\delta \geq \bar{v}^d(i \mid \mathbf{h}^a_t) \left[x(i \mid \mathbf{h}^a_t) - 1\right] \tag{39}$$

$$\delta \geq \bar{v}^d(0 \mid \mathbf{h}^a_t) \left[\sum_i x(i \mid \mathbf{h}^a_t) - K\right] \tag{40}$$

$$\text{Constraints } (21,23,24,29,31,32), (10,12,15,17) \tag{41}$$

[1] Link: https://ix.cs.uoregon.edu/~thanhhng/publications/ConfPaper/AAAI19_Appendix.pdf

where $v^{\lambda}(\mathbf{h}_t^a, i) = 0$ if $t = T - 1$. Constraints (37) and (38–40) correspond to the complementary slackness constraints of the attacker and defender respectively. Note that any equilibrium of the game is a feasible solution of the program (36–41) which returns an objective value of $\delta = 0$. On the other hand, the right-hand side of constraints (37–40) is always non-negative due to constraint (41). Thus, $\delta \geq 0$ for all feasible solutions of the program (36–41). It means that any equilibrium of the game is an optimal solution of (36–41). In addition, since the optimal objective value $\delta = 0$, any optimal solution of (36–41) returns a value of zero for all the right-hand sides of (37–40). Therefore, any optimal solution of this program is a PBNE.

Extension to Stackelberg Setting

In the Stackelberg game model, a mixed strategy of the defender is defined as a probability distribution $\mathbf{m} = \{\mathbf{m}(\mathbf{s}) : \sum_{\mathbf{s}} \mathbf{m}(\mathbf{s}) = 1, \mathbf{m}(\mathbf{s}) \in [0, 1]\}$ over the action set $\mathbf{S}$. We denote by $\mathbf{M}$ the set of all mixed strategies of the defender. At each time step, the defender commits to a mixed strategy. The attacker is aware of that mixed strategy and then decides which target to attack. Therefore, in finitely repeated Stackelberg games, at each time step $t + 1$, an observation of the defender is a history $\mathbf{h}_t = \{(\mathbf{m}_1, i_1), \ldots, (\mathbf{m}_t, i_t)\}$ while an observation of the attacker is a history $(\mathbf{h}_t, \mathbf{m}_{t+1})$.

The behavioral strategy of the defender at $\mathbf{h}_t$ is a probability distribution $\mathbf{x}(\mathbf{h}_t) = \{x(\mathbf{m} \mid \mathbf{h}_t) : \sum_{\mathbf{m}} x(\mathbf{m} \mid \mathbf{h}_t) = 1, x(\mathbf{m} \mid \mathbf{h}_t) \in [0, 1]\}$ over the set of mixed strategies of the defender. On the other hand, a behavioral strategy of the attacker of type λ at $(\mathbf{h}_t, \mathbf{m}_{t+1})$ is a probability distribution $\mathbf{y}^{\lambda}(\mathbf{h}_t, \mathbf{m}_{t+1}) = \{y^{\lambda}(i \mid \mathbf{h}_t, \mathbf{m}_{t+1}) : \sum_i y^{\lambda}(i \mid \mathbf{h}_t, \mathbf{m}_{t+1}) = 1, y^{\lambda}(i \mid \mathbf{h}_t, \mathbf{m}_{t+1}) \in [0, 1]\}$. A PBNE of Stackelberg security games is then defined similarly as simultaneous-move games. Since the set of mixed strategies of the defender is infinite, the existence of a PBNE in Stackelberg security games is an open research question. Nevertheless, we can compute an ϵ-PBNE by discretizing this set of defense mixed strategies and applying the same backward induction method as in the simultaneous case.

We specifically analyze the deception of the attacker in finitely repeated Stackelberg security games with $|\mathbf{N}| = 2$, $|\mathbf{\Lambda}| = 2$, and $K = 1$. We adopt the tradition in Stackelberg security game that rewards and penalties are strictly positive and negative respectively for both players. We consider a game scenario in which the defender only plays a *pure* behavioral strategy in $\mathbf{X}^{\text{pure}} = \{\mathbf{x} : x(\mathbf{m} \mid \mathbf{h}_t) = 1, \text{ for some } \mathbf{m} \in \mathbf{M}, \forall \mathbf{h}_t\}$.

Theorem 3. *In a finitely repeated Stackelberg security game with $|\mathbf{N}| = 2$, $|\mathbf{\Lambda}| = 2$, and $K = 1$, if the defender only plays a pure behavioral strategy in $\mathbf{X}^{pure}$ and the rewards and penalties are strictly positive and negative respectively for both the players, there exists a PBNE of the game in which the attacker plays a myopic best response at every history ($\mathbf{h}_t, \mathbf{m}_{t+1}$).*

One significance of this preliminary result is that the assumption about a myopic attacker in previous work on finitely repeated Stackelberg security games is justified (at least in the simple setting of this result) even when the attackers care about future expected utility. In future research, we aim to generalize this special case and explore the deception patterns for multiple targets and multiple types in the Stackelberg setting.

Experiments

We focus on the attacker's strategic use of deception. In our experiments, the players' rewards and penalties are generated uniformly at random in the range $[1, 10]$ and $[-10, -1]$ respectively.

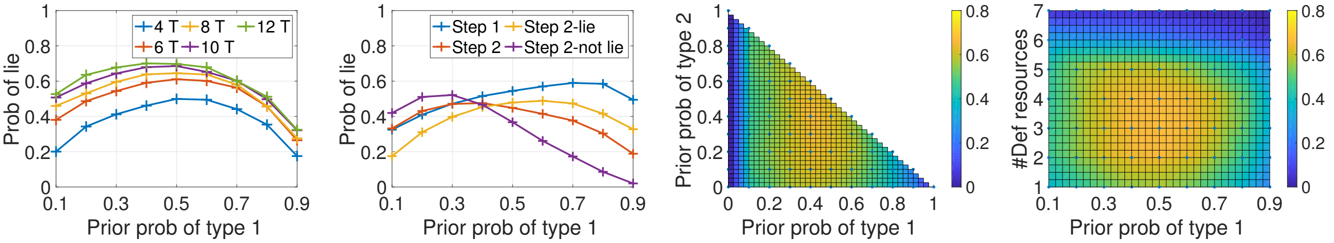
Analysis of attacker deception

The purpose of deception is to shift the defender's belief away from the attacker's true type. Any action on part of the attacker toward this purpose must take into account similar reasoning by other attacker types. Further, shaping the belief of the defender is beneficial only if it results in a later gain for the attacker. In the following, we present our results with respect to an attacker of type 1. The behavior for other attacker types is symmetric.

In our first experiment, we analyze games with number of attacker types: $|\mathbf{\Lambda}| = 2$, number of targets: $|\mathbf{N}| \in \{4, 6, 8, 10, 12\}$, and number of time steps $|\mathbf{T}| \in \{2, 3\}$. Results are shown in Figure 1(a)(b). The x-axis is the prior probability of attacker type 1. The y-axis is the probability a type-1 attacker attacks a myopically non-optimal target (i.e., probability of deceptive action, or "lie" for short) at time step $t = 1$ or $t = 2$ (for 3-step games). Each data point is averaged over 220 game instances. Figure 1(a) shows results for 2-step games; each curve corresponds to a number of targets. Overall, the attacker's tendency to deceive is roughly concave in the prior probability of its true type. This makes sense, as deception has relatively less power to change the beliefs of a defender when they start near extremes. We also see an increase in deception with the number of targets. This reflects the growth in options for deception, as well as increased potential benefit for misleading the defender.

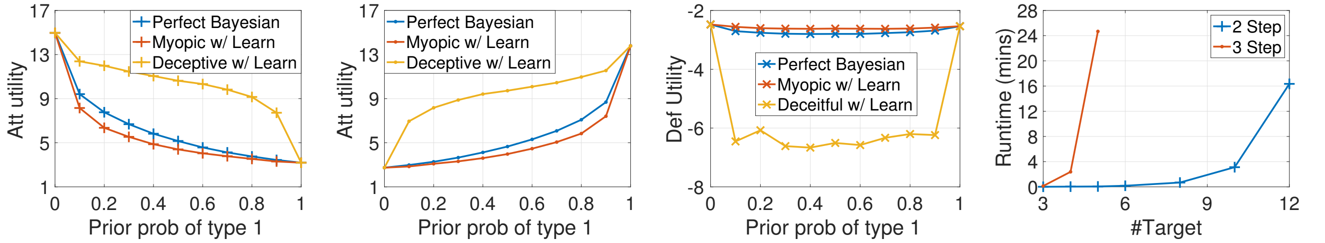
Results for 3-step games are shown in Figure 1(b). We present deception probabilities for the attacker of type 1 at: (1) Step 1; (ii) Step 2; (iii) Step 2-lie (step 2 conditioned that the attacker lied at step 1); and (iv) Step 2-not lie (step 2 conditioned that the attacker did not lie at step 1). In this figure, $|\mathbf{N}| = 4$. As for the 2-step game, the probability of deception in each case is roughly concave in the prior. The probability of deception at step 1 is somewhat elevated in the 3-step game, since the attacker accrues longer-term benefit from misleading the defender. Moreover, the peak is shifted to the right, reflecting increased chance for successful deception given its opportunity to repeat the lie over two periods.

Indeed, given that the attacker lies at step 1, the attacker lies with roughly proportional probability at step 2 (blue curve versus yellow curve). On the other hand, when the attacker does not lie at step 1, its pattern of deception at step 2 (purple curve) is qualitatively different. Switching to be deceptive at step 2 is more promising at low priors (where the act has some chance of misleading), and very unlikely at high priors where there is little chance to mislead the defender if it had not already started in step 1.



(a) 2-type, 2-step games (b) 2-type, 3-step games (c) 3-type, 2-step games (d) 2-type, 2-step games

Figure 1: Attacker deception analysis, attacker type 1.



(a) Attacker type 1 (b) Attacker type 2 (c) Defender (d) Runtime

Figure 2: Solution comparison and runtime performance.

In our second experiment, we analyze deceptive strategies of the attacker type 1 in 2-step games with the number of attacker types is $|\mathbf{\Lambda}| = 3$. In these games, $|\mathbf{N}| = 8$. The result is shown in Figure 1(c). The x-axis and y-axis represent prior probabilities of types 1 and 2 respectively. Figure 1(c) shows that the attacker deception tendency is unimodal with respect the prior of its type 1, and less sensitive ot the distribution across other types.

In our third experiment, we vary the number of defender resources in 2-step games with 2 attacker types. The result is shown in Figure 1(d). When the number of defender resources is high (close to the number of the targets), the defender can provide a high coverage probability at all targets. Specially, when $K = |\mathbf{N}|$, the defender protects all targets all the time. As a result, the attacker may not achieve any benefit by lying. Therefore, the attacker lies less when K gets closer to $|\mathbf{N}|$.

Solution quality and runtime performance

In our last experiment, we compare the players' utilities for playing strategies computed in three scenarios:

1. Perfect Bayesian. The attacker is rationally deceptive and the defender takes into account the potential deceit.
2. Myopic w/ Learn. The attacker is myopic and the defender also assumes so.
3. Deceptive w/ Learn. The attacker is rationally deceptive while the defender assumes the attacker is myopic.

The defender performs a Bayesian update on his belief about the attacker's type in all three cases. Results are shown in Figures 2(a)(b)(c), averaging over 220 3-step game instances with two attacker types and five targets. The x-axis is the prior probability of type 1. The y-axis is the attacker utility of each type or the defender utility on average. Figures 2(a)(b)(c) show if the defender does not account for deception, the rationally deceptive attacker achieves a significant gain while the defender suffers a significant loss (yellow versus red curves). When the defender accounts for the prospect of deception of the attacker, such gains and losses are drastically reduced (blue versus red).

Finally, we display in Figure 2(d) the runtime performance of our equilibrium-finding algorithm. The x-axis is the number of targets and the y-axis is runtime in minutes. For 2-step games, the runtime remains modest for up to ten targets. For $|\mathbf{T}| = 3$, the runtime grows quickly and exceeds 24 minutes when the number of targets is five.

Summary

We study the problem of deception in finitely repeated security games. In these games, the defender collects attack data over time to learn about the attacker type while the attacker plays deceptively to mislead the defender. We present a detailed analysis and computation of finding optimal strategies of players in the games. We then show through computational experiments that the attacker (defender) receives a great benefit (loss) when the defender does not take into account deceptive attacks. Conversely, such benefit (loss) is reduced significantly when the defender addresses the attacker's deception.

Acknowledgment This work was supported in part by MURI grant W911NF-13-1-0421 from the US Army Research Office.

References

Aumann, R. J., and Maschler, M. 1995. *Repeated Games with Incomplete Information*. MIT Press.

Balcan, M.-F.; Blum, A.; Haghtalab, N.; and Procaccia, A. D. 2015. Commitment without regrets: Online learning in Stackelberg security games. In *16th ACM Conference on Economics and Computation*, 61–78.

Barreno, M.; Nelson, B.; Sears, R.; Joseph, A. D.; and Tygar, J. D. 2006. Can machine learning be secure? In *ACM Symposium on Information, Computer and Communications Security*, 16–25.

Barreno, M.; Nelson, B.; Joseph, A. D.; and Tygar, J. D. 2010. The security of machine learning. *Machine Learning* 81(2):121–148.

Blum, A.; Haghtalab, N.; and Procaccia, A. D. 2014. Learning optimal commitment to overcome insecurity. In *Advances in Neural Information Processing Systems*, 1826–1834.

Brown, G.; Carlyle, M.; Diehl, D.; Kline, J.; and Wood, K. 2005. A two-sided optimization for theater ballistic missile defense. *Operations Research* 53(5):745–763.

Brückner, M., and Scheffer, T. 2011. Stackelberg games for adversarial prediction problems. In *ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 547–555.

Brückner, M.; Kanzow, C.; and Scheffer, T. 2012. Static prediction games for adversarial learning problems. *Journal of Machine Learning Research* 13:2617–2654.

Fang, F.; Nguyen, T. H.; Pickles, R.; Lam, W. Y.; Clements, G. R.; An, B.; Singh, A.; Tambe, M.; and Lemieux, A. 2016. Deploying PAWS: Field optimization of the protection assistant for wildlife security. In *30th AAAI Conference on Artificial Intelligence*, 3966–3973.

Farrell, J., and Rabin, M. 1996. Cheap talk. *Journal of Economic Perspectives* 10(3):103–118.

Forges, F. 1988. Repeated games of incomplete information: non-zero-sum. Technical report, Université Catholique de Louvain, Center for Operations Research and Econometrics (CORE).

Gholami, S.; Ford, B.; Fang, F.; Plumptre, A.; Tambe, M.; Driciru, M.; Wanyama, F.; Rwetsiba, A.; Nsubaga, M.; and Mabonga, J. 2017. Taking it for a test drive: a hybrid spatio-temporal model for wildlife poaching prediction evaluated through a controlled field test. In *European Conference on Machine Learning & Principles and Practice of Knowledge Discovery in Databases*.

Guo, Q.; An, B.; Bosansky, B.; and Kiekintveld, C. 2017. Comparing strategic secrecy and Stackelberg commitment in security games. In *26th International Joint Conference on Artificial Intelligence*.

Haghtalab, N.; Fang, F.; Nguyen, T. H.; Sinha, A.; Procaccia, A. D.; and Tambe, M. 2016. Three strategies to success: Learning adversary models in security games. In *25th International Joint Conference on Artificial Intelligence*, 308–314.

Hendricks, K., and McAfee, R. P. 2006. Feints. *Journal of Economics & Management Strategy* 15(2):431–456.

Jordan, J. S. 1995. Bayesian learning in repeated games. *Games and Economic Behavior* 9(1):8–20.

Kar, D.; Ford, B.; Gholami, S.; Fang, F.; Plumptre, A.; Tambe, M.; Driciru, M.; Wanyama, F.; Rwetsiba, A.; Nsubaga, M.; et al. 2017. Cloudy with a chance of poaching: Adversary behavior modeling and forecasting with real-world poaching data. In *16th International Conference on Autonomous Agents and Multi-Agent Systems*, 159–167.

Letchford, J.; Conitzer, V.; and Munagala, K. 2009. Learning and approximating the optimal strategy to commit to. In *International Symposium on Algorithmic Game Theory*, 250–262.

Lowd, D., and Meek, C. 2005. Adversarial learning. In *ACM SIGKDD International Conference on Knowledge Discovery in Data Mining*, 641–647.

Marecki, J.; Tesauro, G.; and Segal, R. 2012. Playing repeated Stackelberg games with unknown opponents. In *11th International Conference on Autonomous Agents and Multiagent Systems*, 821–828.

Nguyen, T. H.; Sinha, A.; Gholami, S.; Plumptre, A.; Joppa, L.; Tambe, M.; Driciru, M.; Wanyama, F.; Rwetsiba, A.; Critchlow, R.; et al. 2016. Capture: A new predictive anti-poaching tool for wildlife protection. In *15th International Conference on Autonomous Agents and Multi-Agent Systems*, 767–775.

Rabinovich, Z.; Jiang, A. X.; Jain, M.; and Xu, H. 2015. Information disclosure as a means to security. In *14th International Conference on Autonomous Agents and Multi-Agent Systems*, 645–653.

Sorin, S. 2002. *A first course on zero-sum repeated games*, volume 37. Springer Science & Business Media.

Xu, H.; Rabinovich, Z.; Dughmi, S.; and Tambe, M. 2015. Exploring information asymmetry in two-stage security games. In *29th AAAI Conference on Artificial Intelligence*, 1057–1063.

Xu, H.; Tran-Thanh, L.; and Jennings, N. R. 2016. Playing repeated security games with no prior knowledge. In *15th International Conference on Autonomous Agents and Multi-Agent Systems*, 104–112.

Zamir, S. 1992. Repeated games of incomplete information: Zero-sum. *Handbook of Game Theory with Economic Applications* 1:109–154.

Zhuang, J.; Bier, V. M.; and Alagoz, O. 2010. Modeling secrecy and deception in a multi-period attacker-defender signaling game. *European Journal of Operational Research* 203:409–418.