Inducible Equilibrium for Security Games
(Extended Abstract)

Qingyu Guo1, Jiarui Gan2, Fei Fang3, Long Tran-Thanh4, Milind Tambe5, Bo An1
1School of Computer Science and Engineering, Nanyang Technological University, {qguo005, boan}@ntu.edu.sg
2Department of Computer Science, University of Oxford, Jiarui.gan@cs.ox.ac.uk
3School of Computer Science, Carnegie Mellon University, feifang@cmu.edu
4Department of Electronics and Computer Science, University of Southampton, ltt08r@ecs.soton.ac.uk
5Center for Artificial Intelligence in Society, University of Southern California, tambe@usc.edu

ABSTRACT
Strong Stackelberg equilibrium (SSE) is the standard solution concept of Stackelberg security games. The SSE assumes that the follower breaks ties in favor of the leader and this is widely acknowledged and justified by the assertion that the defender can often induce the attacker to choose a preferred action by making an infinitesimal adjustment to her strategy. Unfortunately, in security games with resource assignment constraints, the assertion might not be valid. To overcome this issue, inspired by the notion of inducibility and the pessimistic Stackelberg equilibrium [20, 21], this paper presents the inducible Stackelberg equilibrium (ISE), which is guaranteed to exist and avoids overoptimism as the outcome can always be induced with infinitesimal strategy deviation. Experimental evaluation unveils the significant overoptimism and sub-optimality of SSE and thus, verifies the advantage of the ISE as an alternative solution concept.

KEYWORDS
Security games; inducible Stackelberg equilibrium; utility guarantee

1 INTRODUCTION
The past decade has witnessed the huge success of game theoretic reasoning in complex security domains [1, 2, 4, 12, 15, 17]. Various applications based on the Stackelberg security game (SSG) model have been deployed to protect airports, ports, wildlife and so on [8, 17]. The standard solution concept in Stackelberg games is Stackelberg equilibrium [11], which assumes that both players are rational and have no incentive to deviate in the equilibrium. The strong form of the Stackelberg equilibrium, called Strong Stackelberg Equilibrium (SSE) assumes that the follower will always break ties in favor of the defender and is the most commonly adopted concept in related literature [7, 14, 20] and in most security game applications [17]. In essence, researchers have implicitly or explicitly claimed or assumed that the defender can always induce the favorable strong equilibrium by selecting a strategy arbitrarily close to the equilibrium [3, 6, 10, 19].

However, the assertion that the defender can always induce SSE may break in security domains with resource assignment constraints, e.g., protecting flights with air marshals (FAMS) [10, 18], protection ports [16], protecting targets with externalities [9]. Unfortunately, existing research has failed to realize the potential impossibility to induce SSE in such domains. If the desired SSE cannot be induced, the results claimed in existing works may be overly optimistic. Such overoptimism is highly problematic since these results may be used in making security resource acquisition decisions [13], and the SSE strategy recommended may not be the optimal one, thus failing in the primary mission of security games, which is to optimize the use of limited security resources.

In this paper, we offer remedies for this shortcoming. First, we formalize the notion of overoptimism by defining the utility guarantee of the defender’s strategies, and show with a motivating example that the utility claimed to be guaranteed by the SSE is much higher than the actually guaranteed utility. Inspired by the notion of inducible strategy [20] and the pessimistic Stackelberg equilibrium [21], we propose a new solution concept for security games called inducible Stackelberg equilibrium (ISE) based on a novel tie-breaking rule. ISE possesses nice properties that it is guaranteed to exist and avoids overoptimism as it offers the defender the highest guaranteed utility. Second, we prove that the problem of computing an ISE polynomially reduces to that of computing an SSE and thus, introducing the ISE does not invalidate existing algorithmic results. We also provide algorithmic implementation for computing the ISE and conduct experiments to evaluate our results; our experiments unveil the significant overoptimism and sub-optimality of the SSE, which suggests the practical significance of the ISE solution.

2 MOTIVATING EXAMPLE
This paper focuses on security games with arbitrary schedules (SPARS) model [10]. Consider the SPARS instance shown in the following figure where there are four targets, i.e., $T = \{t_1, t_2, t_3, t_4\}$. The defender has one resource $R = \{r\}$. For a target $t \in T$, the defender’s payoff for an uncovered attack is denoted by $U_d^t(t)$ and for a covered attack $U^c_d(t)$. Similarly, $U_a^t(t)$ and $U_a^c(t)$ are attacker’s payoffs respectively. These payoffs are depicted in the figure. We first consider the scenario without resource assignment constraints, which has a unique SSE with coverage strategy $c = \{c_1, c_2, c_3, c_4\}$, where $c_t$ represents the marginal probability that $t$ is covered by a defender resource. In SSE, all targets in $T$ have the same expected
utility for the attacker and thus form a tie, denoted as $\Gamma(c) = T$. The tie-breaking rule in SSE indicates that the attacker will break the tie $\Gamma(c) = T$ by attacking $t_2$. This can be induced by decreasing the coverage on $t_2$ with infinitesimal amount and increasing the coverage on other targets, making $t_2$ strictly preferred.

However, with resource assignment constraints, the defender will not be able to decrease the coverage on one target arbitrarily while simultaneously not decreasing coverage on all other targets. Suppose $f = \{s_1, s_2, s_3, s_4\}$ as shown in the figure. (There is only one resource.) The game still has a unique SSE where the defender plays $x = \left(\frac{1}{2}, \frac{1}{3}, \frac{1}{10}, 0\right)$ and the attacker is assumed to attack $t_3$, bringing the defender an expected utility of $-\frac{1}{3}$. Such outcome is explicitly or implicitly considered with previous mentioned infinitesimal strategy deviation in security game literature [10]. Unfortunately, there exists no strategy arbitrarily close to $x$ which makes $t_2$ be strictly preferred by the attacker. If $x_1$ is decreased, the attacker will prefer $t_1$ over $t_2$; otherwise $t_3$ or $t_4$ will be attacked. That is to say, any infinitesimal strategy deviation will cause the attacker to attack $t_1$, $t_3$, or $t_4$. The best induced outcome for the defender is only arbitrarily close to $-\frac{1}{3}$, achieved by decreasing $x_1$ with infinitesimal amount and the attacker is induced to attack $t_1$.

3 INDUCTIBLE STACKELBERG EQUILIBRIUM

The above example reveals a failure of the attempt to induce the desired SSE outcome by playing a strategy arbitrarily close to the SSE strategy. To formalise this situation, we propose the notion called utility guarantee. Let $X$ be the strategy space of the defender, and consider only the pure strategy for the attacker. Let $U_d : X \times T \to \mathbb{R}$ be the expected utility function for player $d$. Let $\Gamma(x) = \arg\max_{t \in T} U_d(x, t)$ denote the attack set w.r.t. $x$.

**Definition 1 (Utility Guarantee).** We say an expected utility $v$ can be guaranteed by defender’s mixed strategy $x$ iff: $\forall \epsilon > 0, \forall \delta > 0, \exists x' \in X$ such that $\|x - x'\| \leq \delta$ and $U_d(x', f^W(x')) \geq v - \epsilon$, where $f^W : X \to T$ satisfies $f^W(x') \in \arg\min_{t \in T}(x') U_d(x', t)$ for all $x' \in X$. Let $U^g(x) \subseteq \mathbb{R}$ be the set of utilities guaranteed by $x$. $

In other words, if a strategy $x$ guarantees a utility value $v$, the defender can obtain an expected utility at least arbitrarily close to $v$ by playing a strategy arbitrarily close to $x$, regardless of how the attacker actually breaks the tie (the spirit of “guaranteed”). As shown in the motivating example, the utility of an SSE strategy might fail to be guaranteed. This results in overoptimism and a suboptimal solution. To remedy the overoptimism, we propose a new solution concept called Inducible Stackelberg Equilibrium (ISE), based on the notion of inducibility [20].

**Definition 2 (Inducible Target).** A target $t$ is inducible iff there exists at least one defender’s mixed strategy $x \in X$ such that $t$ is the unique best response target against $x$.

We denote by $T^I = \{t \in T | \exists x \in X : \Gamma(x) = \{t\}\}$ the set of inducible targets. ISE is a profile $(x, f^I(x))$ where $f^I : X \to T$ satisfies that $f^I(x) \in \arg\max_{t \in \Gamma(x) \cap T} \max_{x' \in X} U_d(x', f^I(x'))$.

Comparing ISE with SSE, we notice that in ISE, the attacker also breaks the ties in favor of the defender, and the only difference is that the attacker is restricted to attack the inducible targets in ties. It turns out that the inducible targets characterize the highest utility guarantee achieved by any mixed strategy of defender.

One may notice that the SSE strategy coincides with the so-called pessimistic leader-follower equilibrium [5, 21]. In fact, ISE is a generalization of the pessimistic Stackelberg equilibrium in the context of security games where the notion of utility guarantee is proposed to formalise the indcibility issue we observed, and the tie-breaking rule $f^I$ is provided to make the solution consistent with SSEs in security games literature. Following the similar analysis in [5, 21], we prove several nice properties of ISE, namely, existence and the optimality w.r.t. the utility guarantee, making ISE an appealing alternative of SSE to overcome the potential overoptimism.

4 EXPERIMENTAL EVALUATION

We conduct experiments to evaluate the overoptimism and the sub-optimality of SSE. The rewards and penalties are all integers randomly drawn from $[0, 5]$ and $[-5, 0]$ respectively. The game has 200 targets, 1 resource, number of schedules $S$ and length per schedule $l$ in $\{16, 18, 20\}$. As shown in Figure 1, SSE suffers from overoptimism and sub-optimality since a significant proportion of random instances are spotted with overoptimistic and/or sub-optimal SSEs. Moreover, once the overoptimism and sub-optimality occur on an SSE, the actual utility guarantee (“SSE-g”) is significantly less than the provided amount by SSE (“SSE-u”), so as the guaranteed utility of SSE and the optimal utility in ISE (“ISE-g”).

**ACKNOWLEDGMENTS**

This research was supported by MURI grant W911NF-17-1-0370 and the National Research Foundation, Prime Minister’s Office, Singapore under its IDM Futures Funding Initiative. Long Tran-Thanh was supported by the EPSRC funded project STRICT (EP/N02026X/1).
REFERENCES


