Learning to Signal in the Goldilocks Zone: Improving Adversary Compliance in Security Games

Sarah Cooney[1], Kai Wang[1], Elizabeth Bondi[1], Thanh Nguyen[2], Phebe Vayanos[1], Hailey Winetrobe[1], Edward A. Cranford[3], Cleotilde Gonzalez[3], Christian Lebiere[3], and Milind Tambe[1]

[1] University of Southern California, Los Angeles, CA 90089
{cooneys,wang319,bondi,phebe.vayanos,hwinetro,tambe}@usc.edu
[2] University of Oregon, Eugene, OR 97403
thanhhng@cs.uoregon.edu
[3] Carnegie Mellon University, Pittsburgh, PA 15289
{cranford,coty,cl}@cmu.edu

Abstract. Many real-world security scenarios can be modeled via a game-theoretic framework known as a security game in which there is a defender trying to protect potential targets from an attacker. Recent work in security games has shown that deceptive signaling by the defender can convince an attacker to withdraw his attack. For instance, a warning message to commuters indicating speed enforcement is in progress ahead might lead to them driving more slowly, even if it turns out no enforcement is in progress. However, the results of this work are limited by the unrealistic assumption that the attackers will behave with perfect rationality, meaning they always choose an action that gives them the best expected reward. We address the problem of training boundedly rational (human) attackers to comply with signals via repeated interaction with signaling without incurring a loss to the defender, and offer the four following contributions: (i) We learn new decision tree and neural network-based models of attacker compliance with signaling. (ii) Based on these machine learning models of a boundedly rational attacker's response to signaling, we develop a theory of signaling in the *Goldilocks zone*, a balance of signaling and deception that increases attacker compliance and improves defender utility. (iii) We present game-theoretic algorithms to solve for signaling schemes based on the learned models of attacker compliance with signaling. (iv) We conduct extensive human subject experiments using an online game. The game simulates the scenario of an inside attacker trying to steal sensitive information from company computers, and results show that our algorithms based on learned models of attacker behavior lead to better attacker compliance and improved defender utility compared to the state-of-the-art algorithm for rational attackers with signaling.

Keywords: Security · Stackelberg Games · Behavioral Modeling and Learning · Bounded Rationality · Signaling · Deception.

1 Introduction

Imagine a highway on which many commuters with a tendency to speed travel each day. Suppose the police have a limited amount of time to patrol this highway, but still want to stop people from speeding. One solution is to use deceptive signals, or warnings. For example, a sign noting that speed enforcement is in progress ahead could be used, even if this is not actually the case. It is easy to imagine that if this sign were displayed very often with no real police patrols, the commuters would quickly realize and continue speeding. However, if the commuters knew there was a good chance of actually being stopped and issued a ticket, they would probably slow down. The question is how often can the police display the sign deceptively (without enforcing speed) and still cause the commuters to slow down?

This is the question answered by Xu et al. with their framework for deceptive signaling in Stackelberg Security Games (SSGs) [30]. SSGs model the interaction between an attacker and a defender (in our example, the commuters and the police), and have successfully helped security agencies worldwide optimize the use of limited security resources to mitigate attacks across domains from protecting ports and flights, to mitigating the poaching of endangered animals [2, 13, 21, 26, 28]. (We use the term attack broadly to refer to any unwanted behavior or illegal activity, such as speeding.) With the addition of signaling, Xu et al.'s framework allows the defender to strategically reveal information about her defensive strategy to the attacker [30]. On seeing a signal (e.g, a warning that speed enforcement is in effect), a compliant attacker will withdraw his attack to the defender's benefit. The main advantage of signaling is the ability to deter attacks using deception, instead of deploying scarce or costly defensive resources.

This signaling framework was shown in simulation to improve the defender utility against a *perfectly rational* attacker (who always takes the action with the *best* utility for him) compared to the traditional SSG model. Unfortunately, real-world attackers are almost always boundedly rational (*not* always selecting the action with the best utility). Therefore, we focus on finding methods to improve the compliance rates of boundedly rational attackers, who may not comply even if it is rational to do so, but instead learn to react via repeated interactions with signals. This framework could be used to deter boundedly rational attackers in a variety of real-world settings where attackers might repeatedly interact with signals. For instance, speeding commuters, fare evaders on public transit [17], opportunistic criminals looking for chances to strike [34], or cyber-attackers repeatedly probing a system [18].

In order to increase the compliance of boundedly rational attackers, we focus on the frequency of signaling, or deciding how often to signal, and use machine learning models and optimization to learn the overall number of warnings to show, as showing too many warnings can cause attackers to simply ignore them. A key result of this paper is the discovery of a *Goldilocks zone* for signaling—a careful balance of signaling and deception that considers underlying characteristics of individual targets—via the use of machine learning models of attacker behavior, which leads to an increase in human attacker compliance and an im-

provement of defender utility. Our main contributions are as follows: (i) We learn new models of attacker compliance with regard to signaling based on decision trees and neural networks. (ii) Utilizing insights from these learned models we propose a theory of signaling in the *Goldilocks zone*, a balance of signaling and deception that increases the compliance of boundedly rational adversaries while mitigating losses to the defender. (iii) We present game-theoretic algorithms to solve for signaling schemes based on the learned models. (iv) Using an online game based on the scenario of an inside attacker, we conduct extensive human subject experiments, which show that against boundedly rational subjects, our new modeling-based signaling algorithms outperform the state-of-the-art algorithm designed for perfectly rational attackers.

2 Related Work

Two key game-theoretic frameworks, which have been studied and applied extensively, are SSGs, which model interactions between a defender and an attacker [1, 14, 29], and signaling games, which model an interaction between two parties in which one party (the sender) reveals some hidden information to the other (the receiver), with the goal of influencing his behavior [22, 25]. With the growing interest in the use of deception for security, particularly in the cyber realm [7], game theory researchers have also begun incorporating deception into the security and signaling game frameworks [10, 19, 36]. Recent work has combined the security and signaling game frameworks with deception, such that the defender strategically reveals (possibly deceptive) information about her defensive strategy to the attacker in hopes of causing him to withdraw his attack [12, 31].

However, previous work in game-theoretic frameworks with deception has not investigated human behavior in response to deception, but has instead assumed all respondents are perfectly rational. This is a major limitation to translating these frameworks for use in the real world. In contrast, we focus on this combined signaling-security game framework with boundedly rational attackers who need to be trained to comply with signaling, which has not previously been considered.

There is extensive work modeling the behavior of boundedly rational attackers in classic SSGs without signaling. Early models relied on specific assumptions about attacker behavior [24], using functional forms based on these assumptions such as quantal response [20, 32] or prospect theory [33]. More recent work has turned to machine learning models, which use real-world data and do not rely on specific assumptions about attacker behavior [15, 35]. Two methods that have been used to predict the behavior of humans in game-theoretic settings are decision trees, which have been used to predict the actions of poachers in Green Security Games [13], and neural networks, which have recently been used to predict the distribution of actions for a player in normal-form, simultaneous move games [11]. We use both of these models, but in contrast to previous work, we are the first to address human behavior *with regard to signaling in a SSG*.

3 Background

In a SSG, there is a set of targets $T = \{t_1, t_2, \ldots, t_n\}$ which the defender protects by allocating $K < n$ resources over them. A *pure* defense strategy is an allocation of the resources, with a *mixed* strategy being a randomization over these pure strategies. Without scheduling constraints, a *mixed* strategy can be equivalently represented as marginal coverage probabilities over the targets, denoted $\mathbf{z} = \{z_t\}$, with $z_t \in [0, 1]$, and $\sum_t z_t = K$, where z_t is the probability of protecting target t [14]. The attacker is aware of $\mathbf{z}$ (but not the pure strategy) and chooses a target t to attack accordingly. If the defender is protecting t, the attacker incurs a penalty of $U_a^c(t) < 0$ and the defender is rewarded with $U_d^c(t) \geq 0$. If t is unprotected, the attacker gets a reward of $U_a^u(t) > 0$ and the defender gets a penalty of $U_d^u(t) < 0$. Xu et al. [30] introduced a two-stage SSG with a signaling scheme, allowing the defender to influence the attacker's decision making to her benefit by exploiting the fact that the attacker is unaware of the pure strategy at any given time. A round of the two-stage game plays out as follows:

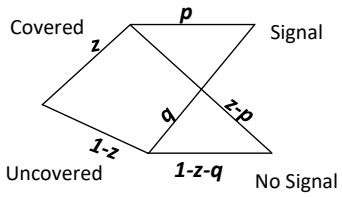

Fig. 1: The signaling scheme for a target t.

1. The defender allocates her resources, covering a random subset of the targets based on her mixed strategy $\mathbf{z}$.
2. Aware of the defender's mixed strategy, the attacker chooses a target, t, to attack accordingly.
3. The defender sends a (possibly deceptive) signal to the attacker regarding the current protection status of t.
4. Based on the information given in the signal, the attacker chooses to either (1) continue attacking or (2) withdraw his attack yielding payoffs of zero for both players.

The first stage (steps 1 & 2) is identical to the classic SSG. The second stage (steps 3 & 4) introduces signaling. We can formalize a signaling scheme as follows:

Definition 1 (Signaling Scheme [30]). *Given (t, z_t) and a signal σ, a signaling scheme regarding t consists of probabilities (p_t, q_t) with $0 \leq p_t \leq z_t$ and $0 \leq q_t \leq 1 - z_t$, such that p_t and q_t are the probabilities of showing σ given that t is currently covered and uncovered, respectively.*

Figure 1 visualizes a signaling scheme for a target t, where z_t is the coverage probability, p_t [q_t] is the probability of signaling given t is covered [uncovered]. A signaling scheme tells the defender how often to warn the attacker broken down into the cases (1) when the warning is true (p_t) and (2) when it is false (q_t). For instance, with probability $(1 - z_t)$, t is not protected by a defensive resource. In this case, we will send a deceptive signal telling the attacker t is protected with probability $\frac{q_t}{1-z_t}$. Intuitively, it is the optimal combination of bluffing and truth telling to ensure the attacker always believes the bluff. The goal is to bluff as much as possible while maintaining this belief.

4 Signaling Schemes for Boundedly Rational Attackers

While previous work focused on a perfectly rational attacker [30], we devise a signaling scheme that increases the compliance of (human) boundedly rational attackers. We let x_t [y_t] be the probability the adversary attacks the chosen target t given a signal is shown [*no* signal is shown]. For a perfectly rational attacker [30], $x_t = 0$ and $y_t = 1$. As we show in Section 7, humans do not behave in such a deterministic manner, so our goal is to find a signaling scheme (p_t and q_t) that provides the most benefit to the defender, despite human behavior. We measure benefit by the expected utility of the defender, defined as follows, where g_t is the probability that the attacker selects t, term (i) is the expected defender utility given no signal is shown, and term (ii) is the expected defender utility otherwise. In each term, we sum up the total expected utility for target t, which has defender reward $U_d^c(t)$ and penalty $U_d^u(t)$:

$$U^d = \sum\nolimits_t g_t[\underbrace{y_t(z_t - p_t)U_d^c(t) + y_t(1 - z_t - q_t)U_d^u(t)}_{(i)} + \underbrace{x_t p_t U_d^c(t) + x_t q_t U_d^u(t)}_{(ii)}]$$

In the signaling scheme proposed by Xu et al. [30] (hereafter referred to as the `peSSE` algorithm), term (ii) is always equal to zero (i.e. $x_t = 0$), which is the optimal solution for a perfectly rational attacker. It is the maximum amount of signaling that can be shown and still cause the attacker to withdraw anytime he sees a signal. Further, under the `peSSE` scheme, the defender only employs deception when a signal is shown. When no signal is shown, it is *always* true that the given target is uncovered (i.e. $z_t - p_t = 0$), and the attacker will succeed. We refer to this type of scheme, as a *1-way* deceptive signaling scheme.

As we show in Section 7, in the presence of a boundedly rational attacker, using a signaling scheme, even one designed for perfectly rational attackers, improves defender utility when compared to the traditional SSG framework. We also show that boundedly rational attackers display a training effect via experience with signals. As they experience signals and the consequences of attacking, they become more compliant—attacking less frequently as time goes on. However, under the `peSSE` scheme, this decrease in attack probability is both gradual and small in magnitude. Therefore, we seek a way to both increase the overall rate of compliance and to speed up the training process without incurring additional loss to the defender. The natural starting point based on insights from literature on using warnings to deter risky cyber behavior [16], is to adjust the false positive (deception) rate. We used a regression tree to learn the probability of attack given a signal (x_t), based on features of each target, including the rate of deception. However, in order to handle instances in which there is no signal—a sure loss to the defender—the optimization process suggested more signaling. We will show this led to the defender being worse off than under the `peSSE` scheme.

Given these results, we hypothesized that the overall frequency of signaling, not just the deception rate, also has an impact on attacker behavior. In particular, that a high frequency of signaling was causing the attacker to become desensitized and less compliant. Therefore, we propose a new scheme which we

call a *2-way* deceptive signaling scheme, which lowers the overall frequency of showing a signal without changing the deception rate, and introduces uncertainty for the attacker when no signal is shown. As shown in Section 7, 2-way signaling schemes result in faster training of the attacker, an overall increase in compliance, and better expected utility for the defender against boundedly rational attackers. In a 2-way signaling scheme, we decrease p_t and q_t proportionally to reduce the frequency of signaling, while adding uncertainty about the protection status of t when no signal is shown. We formally define the new scheme as follows:

Definition 2 (2-Way Signaling Scheme). *Let* $\mathbf{f}$ *be a vector such that* $\mathbf{f} \in \mathbb{R}^{|T|}$ *and* $f_t \in [0, 1]$ *for all* $t \in T$*. Then,*

$$\textit{(i)}\ p_t = f_t z_t \qquad\qquad \textit{(ii)}\ q_t = -p_t U_a^c(t)/U_a^u(t)$$

In equation (ii), we ensure that the expected value when a signal is shown is equal to zero for all targets. This keeps the deception rate consistent with the peSSE scheme, allowing us to focus on the effect of signaling frequency without confounding the effect of changes in deception rate. Intuitively, f_t is the proportion of signals shown compared to the peSSE strategy. For example, if $f_t = 0.5$, we show half as many signals as the peSSE strategy.

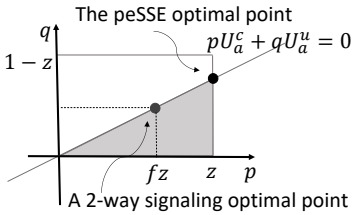

Fig. 2: The (p, q)-feasible regions for the peSSE and 2-way signaling schemes.

We can visualize 2-way signaling in relation to the peSSE scheme by looking at the feasible region of (p, q) in the optimization used to solve for the peSSE scheme (Figure 2). Figure 2 gives the intuition for part two of the following theorem (its proof is in the appendix)[4]:

Theorem 1. *Given a 2-way deception scheme with* $f_t \in (0, 1)\ \forall\ t$*, if the attacker is perfectly rational, then:*

(i) The attacker's expected utility per target will be equal to his expected utility under the peSSE *signaling scheme.*
(ii) The defender's expected utility per target will be worse than hers under the peSSE *signaling scheme.*

Two-way signaling makes the signaling scheme sub-optimal for the defender against a perfectly rational attacker, but as we show in Section 7, it improves her utility against boundedly rational attackers. The question is *how to choose the correct value of* f_t?

As a baseline, we uniformly reduce the signaling frequency on all targets ($f_t = 0.75, \forall t$), and show that this leads to faster training of subjects, an improvement in the end compliance rate, and an improvement in expected utility

[4] Link:https://www.dropbox.com/s/uum5tpnb4h1gmym/ECML$_S$*upplement.pdf?dl* = 0

for the defender. However, we hypothesize that we can do better by exploiting the boundedly rational attackers' differing preferences over the targets [20, 23]. We consider learned models of attacker behavior with regard to signaling to determine optimal frequencies of signaling across targets, leading us to find the *Goldilocks zone* for signaling for each target (Section 6.3), which outperforms the baseline's uniform reduction of signaling.

5 Learning Models of Attacker Compliance

Recent work has shown machine learning models of human behavior to outperform classic statistically-based behavioral models such as `SUQR` [9]. Therefore, to model the attacker's response to signaling, we chose two machine learning methods: (i) a decision tree (`DT`), which has shown recent success in applications to patrol planning to stop poachers [8]; and (ii) a neural network (`NN`), which is generally considered the state-of-the-art in predictive modeling.

Round	Model	Accuracy	Precision	Recall
Round 1:	DT	0.711	0.714	0.986
	NN	0.783	0.783	1.0
Round 2:	DT	0.725	0.727	0.995
	NN	0.720	0.731	0.973
Round 3:	DT	0.690	0.705	0.944
	NN	0.683	0.744	0.822
Round 4:	DT	0.654	0.660	0.935
	NN	0.623	0.680	0.786

Table 1: Accuracy of Attacker Models

We compiled a data set of 17,786 instances on which subjects saw a signal, from three different experiments—`peSSE`, deception-based, and 2-way signaling baseline (see Section 7.1). The features of each data point were the attacker reward and penalty ($U_a^u(t)$ & $U_a^c(t)$), the coverage probability (z_t), and the signaling frequency ($p_t + q_t$), for the attacker's target selection t. We predicted the subject's action (1=Attack, 0 = Withdraw). In order to account for the fact that the level of experience with signals so far has an impact on subject behavior, we separated the data by round, resulting in four data sets with 4448, 4475, 4229, and 4634 instances, respectively. We trained `DT` and `NN` models on each round separately. The `DT` model was trained in R using the `rpart` library, which utilizes the CART algorithm to create classification trees [27]. The complexity parameter (CP) was set to 0.003 to avoid over-fitting. The `NN` was built in Python using the Keras[5] library. The network was composed of two hidden layers with 50 and 100 nodes, respectively. For training, a weighted categorical crossentropy loss function was used, where the "attack" class (1) was weighted by 0.4 and the "withdraw" class (0) was weighted by 0.6 due to class imbalance. The Nesterov Adam optimizer was used with Glorot normal initialization. The number of nodes, optimizer, and initialization were determined using randomized search hyperparameter optimization from scikit-learn[6]. This was repeated for multiple

[5] https://keras.io
[6] https://scikit-learn.org/stable/

weights, and the best combination on the validation set was used. Table 1 shows the precision, recall, and mean accuracy on 100 random 80/20 splits of the data.

Despite similar accuracy, the models have different strengths. The `DT` model can give more insight into the features that are most important for increasing compliance. In fact, `rpart` gives an importance value to each variable, and consistent with our hypothesis, frequency is the most important feature. However, the `DT` model has a more coarse-grained set of predicted attack probabilities. The `NN` model is a black-box when it comes to explaining the importance of different features, but gives more fine-grained predictions of attack probability. In the following, we propose new game-theoretic algorithms to find the corresponding optimal signaling scheme for the defender based on both models. As we show in Section 7, both methods outperform the `peSSE` algorithm, with the `NN` method only slightly outperforming the `DT` scheme. Thus, practitioners can choose a method based on the trade-off between performance and explainability best suited to their application.

6 Using Learned Models of Behavior to Compute a Signaling Scheme

Using the `DT` and `NN` models of attacker compliance, our goal is to compute a signaling scheme that maximizes defender expected utility as expressed in Section 4. Evidence from initial experiments show that when there is no signal participants always attack, so we simplify the computation, letting $y_t = 1.0$, and encode the probability of attack given a signal (x_t) as a function of the models' predictions. We focus only on finding the signaling probabilities (p_t, q_t), setting the coverage (z_t) using the algorithm given in [30] and using experimental data to set the selection probabilities (g_t).

6.1 Decision Tree Based Signaling Scheme

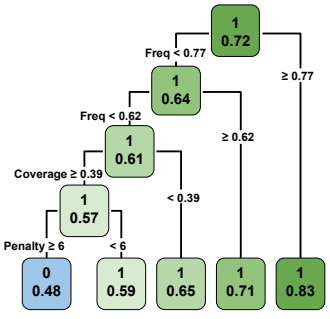

Fig. 3: The `DT` modeling the probability of attack given a signal for Round 2 of the Insider Attack Game.

Our goal is to determine what signaling frequency to set in order to maximize the defender's expected utility, where the attacker's response to signaling is given by a `DT`. For example, Figure 3, where each node lists the predicted action (0= No Attack, 1= Attack) and the percentage of attacks ($1's$) at the node (x_t). We will use a mixed-integer linear program (MILP) to find the optimal frequencies. Here we introduce general techniques for building a MILP from the `DT` model. (The full MILP based on Figure 3 is in the appendix.)

We begin by linearizing the expression of defender utility introduced in Section 4, which requires introducing two additional variables, $m_t =$

$$x_t q_t \text{ and } n_t = x_t p_t\text{:}$$

$$U^d = \sum\nolimits_t g_t[(z_t - p_t)U_d^c(t) + (1 - z_t - q_t)U_d^u(t) + n_t U_d^c(t) + m_t U_d^u(t)]$$

Each branch splits the data on one of four features—attacker reward (U_a^u), attacker penalty (U_a^c), and coverage probability (z_t), and signaling frequency ($p_t + q_t$). We define binary variables to represent the frequency branches. For each branch on frequency of α, we define a binary variable b_t such that $b_t = 1$ if $p_t + q_t \geq \alpha$ and $b_t = 0$ otherwise. This is enforced by the following constraints, where $M, \epsilon > 0$ are a large and small constants, respectively:

$$\alpha - (1 - b_t)M \leq p_t + q_t \leq \alpha + b_t M - \epsilon$$

For each leaf, we define constraints that enforce that the correct predicted value is substituted for (x_t), constraining the values of m_t and n_t. For example, the constraints on m_t associated with the fourth leaf in Figure 3 are as follows:

$$0.71q_t - b_t M - (1-c_t)M \leq m_t \leq 0.71q_t + b_t M + (1-c_t)M$$

where b_t and c_t are the binary variables associated with branching on frequency $= 0.77$ and $= 0.62$, respectively. These constraints enforce that $x_t = 0.71$, meaning $m_t = 0.71q_t$, when $b_t = 0$ and $c_t = 1$, which is equivalent to frequency $\in [0.62, 0.77)$.

6.2 Neural Network Based Signaling Scheme

To optimize over the black-box `NN` model, we optimize over a piece-wise linear (PWL) approximation of the predictions using the technique described in [8]. We let f_t define frequency $(p_t + q_t)$ according to definition 2, and introduce the constraint, $p_t = f_t z_t \; \forall \; t \in T$

Then, we let $\chi_t(f_t)$ be the black-box function predicting attack probability given a signal (x_t), according to the static features z_t, $U_a^c(t)$, and $U_a^u(t)$, taking f_t as an argument. We build a data set (D_χ) of sample predictions at m levels of f_t for each of the T targets, defined by z_t, U_a^u, and U_a^c. Using D_χ, we construct the PWL approximation, representing any value $f_t \in [0, 1]$ and it's prediction $\chi(f_t)$, as a convex combination of its nearest neighbors in the data set for t. Let $B \in D_\chi$ be the break points of the PWL function. We define sets of weights $\lambda_{t,i}$ such that they belong to a *Specially Ordered Set of Type 2*—a set of variables in which at most two can be non-zero, and the non-zero variables must be consecutive. We can then approximate $\chi_t(f_t)$ as a convex combination of (X_t, λ_t) as $\bar{\chi}_t(f_t) = \sum_i \lambda_{t,i}\chi_t(B_{t,i})$. We replace x_t with this expression to formulate defender utility:

$$\begin{aligned} U^d = &\sum\nolimits_t g_t[y_t(z_t - p_t)U_d^c(t) + y_t(1 - z_t - q_t)U_d^u(t) \\ &+ (\sum\nolimits_i \lambda_{t,i}\bar{\chi}_t(B_{t,i}))p_t U_d^c(t) + (\sum\nolimits_i \lambda_{t,i}\bar{\chi}_t(B_{t,i}))q_t U_d^u(t)] \end{aligned} \tag{1}$$

6.3 Signaling in the Goldilocks Zone

We now show empirically that using a learned model of attacker behavior *should* improve on a scheme that uniformly reduces signaling frequency on all targets. First, we show there is an expression which can be used to compute the optimal value of f_t for each target individually.

Theorem 2. *Finding an optimal NN-based signaling scheme is equivalent to minimizing* $f_t(y_t - \bar{\chi}(f_t))(U_d^c(t)U_a^u(t) - U_d^u(t)U_a^c(t))$ *for each* $t \in T$ *individually, where* $\bar{\chi}(f_t)$ *is the piecewise-linear version of* $\chi(f_t)$*. Specifically, if* $U_d^c(t)U_a^u(t) - U_d^u(t)U_a^c(t) < 0[> 0]$*, it is equivalent to maximizing [minimizing]* $f_t(y_t - \bar{\chi}(f_t))$ *for all* $t \in T$*.*

In our experimental setting, the utilities satisfy $U_d^c(t)U_a^u(t) - U_d^u(t)U_a^c(t) < 0$, and since the attacker empirically always attacks when no signal is presented ($y_t = 1$), we have the following simplified corollary following from Theorem 2:

Corollary 1. *The optimal* `NN`*-based solution coincides with* `peSSE` *when* $f_t(1 - \chi(f_t))$ *is monotonically increasing for all* t*.*

We refer to the value of f_t given by this computation as the *Goldilocks zone* for each target. To give a better intuition about finding the *Goldilocks zone*, we visualize the trend of the function described in Theorem 2, $f_t(1 - \chi(f_t))$, using the `DT`'s and `NN`'s predictions of $\chi(f_t)$ (dropping the bar over $\chi(f_t)$ for simplicity and setting $y_t = 1$, per our setup). The graphs in Figure 4 show a plot of $f_t(1 - \chi(f_t))$ on the y-axis at 20 levels of f_t (x-axis) for two of the targets from round 2 of our experiment. Observe that the relationship between f_t and $f_t(y_t - \chi(f_t))$ is different for the two targets. Notice that for Target 1 (left), the baseline value of $\mathbf{f} = 0.75$ (yellow dot) is sub-optimal in that it signals too little compared to the optimal `NN` scheme. However, for Target 4 (right), the baseline signals too often compared to the optimal `NN` solution. Notice that this is also true for the `DT` scheme. By optimizing over our learned models, we can find the *Goldilocks zone* for signaling for each target. As we show in Section 7, the learning-based signaling schemes actually outperform the baseline in practice.

In general, we find that for more conservative, and thus typically less desirable targets, like Target 1 (reward 5/ penalty 3), the optimal signaling rate is higher, with f_t tending toward 1. With more risky, but more appealing, targets such as Target 4 (reward 8/ penalty 9), the *Goldilocks zone* is lower, with f_t tending toward 0.5. A table of the values of f_t for all of the targets under the evaluated signaling schemes can be found in the appendix.

7 Experiments and Results

To evaluate the signaling schemes, we recruited human subjects from Amazon Mechanical Turk to play an online game based on the inside attacker scenario described in [4]. Before starting the game, subjects were given instructions about how it worked, took a short quiz on the instructions, and played a practice round

of 5 trials, allowing them to get a sense for the game. Subjects played four rounds of 25 trials each. To study how the subjects' behavior changed with repeated exposure to signaling, the four rounds were played in a fixed order. Each round had six targets (computers) with a different coverage and payoff structure (see Table 1 in [4], as well as the online appendix). After selecting a target t, with probability $(p_t + q_t)$, the subject is shown a warning message. Given a warning and the probability it is false, the subjects then decided whether or not to attack. For consistency, the subjects were also given the choice to attack or withdraw even when no signal was shown. Screenshots of the game interface and details about the participant pool and payment structure can be found in the appendix.

7.1 Evaluated Algorithms

We compare the solution quality of the signaling schemes given by the following algorithms: (i) *no-signaling algorithm*—the defender plays according to the SSE (equivalently, $f_t = 0 \ \forall \ t$); (ii) `peSSE`—the optimal signaling scheme for a perfectly rational attacker [30] (equivalently, $f_t = 1 \ \forall \ t$); (iii) *2-way signaling baseline*—we set $f_t = 0.75 \ \forall \ t$; (iv) `DT` *based algorithm*; (v) `NN` *based algorithm*; and (vi) *deception-based algorithm*.

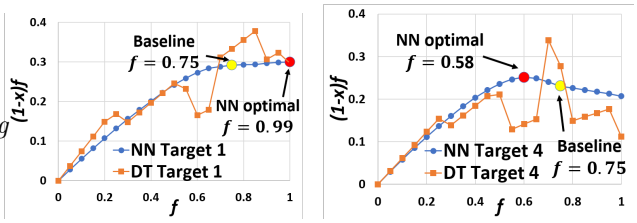

Fig. 4: The relationship between f_t & $(1 - x_t)f_t$ given by the `NN` and `DT` for targets 1 (left) and 4 (right) in round 2 of the insider attack game. For some targets, the baseline signaling frequency is too low [high].

Evaluation Criteria. We evaluate the algorithms with regard to the average defender expected utility, which is defined for each trial as follows:

$$\frac{1}{N}\sum_{i=1}^{m} A_i[(-1)(1 - z_t)]$$

where A_i is the action take by the attacker at round i ($A_i = 1$ being attack and $A_i = 0$ being withdraw), N is the number of participants, and m is the number of trials. We report $p-$values for a 2-tailed t-test comparing mean expected defender utility per trial. The net score was computed across rounds (e.g, earning 20 points in round 1 and -10 points in round 2 would result in a net score of 10), so we report statistics at both the round and aggregate levels.

7.2 Human Subject Results

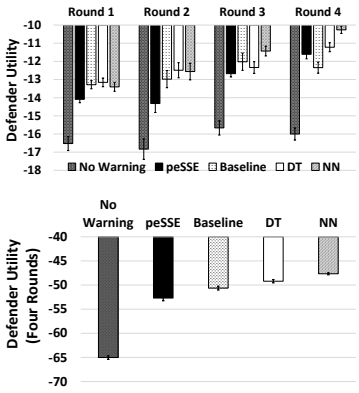

Fig. 5: Comparison of average expected defender utility at the round (top) and aggregate (bottom) levels

Signaling Works. Figure 5 (top) shows the average defender expected utility (y-axis) for each round of the insider attack game. It shows that there is significant benefit ($p < 0.01$) to the defender when using signaling against boundedly rational attackers compared to using no signaling, even when using the `peSSE` algorithm, designed for perfectly rational attackers. This is also true at the aggregate level ($p < 0.01$) (Figure 5 (bottom)).

Signaling Frequency Matters. At the aggregate level, all three 2-way signaling schemes outperformed the `peSSE` algorithm at $p < 0.01$ (Figure 5 (bottom)). As we hypothesized, reducing the frequency of signaling improves performance against boundedly rational attackers.

Learning-Based Schemes Perform Best. As can be seen in Figure 5, the signaling algorithms based on learned models of attacker behavior performed the best, outperforming both the `peSSE` and 2-way baseline schemes. The `DT` scheme outperformed the `peSSE` in rounds 1 ($p < 0.01$), 2 ($p < 0.01$), and 4 ($p < 0.08$) with no significant difference in round 3. It outperformed the baseline in rounds 2 ($p < 0.03$) and 4 ($p < 0.01$), with no significant difference in utility in rounds 1 and 3. The `NN`-based algorithm outperformed `peSSE` in all rounds ($p < 0.01$). It also outperformed the baseline in rounds 2 ($p < 0.08$), 3 and 4 (both $p < 0.01$), with no significant difference in round 1.

The Goldilocks Zone for Signaling. A key finding of of our experiments is that using learned models of subject behavior to find the proper signaling frequency (the *Goldilocks zone*) increases its impact, which aligns with our theoretical results (Section 6.3). Figure 6 (left), shows the average percent of trials in each round on which subjects saw a signal across the four signaling algorithms. The baseline algorithm signals the least and also achieves almost the best compliance (Figure 6 (right), the average rate of attack given a signal). The `DT` and `NN` based algorithms have middling signaling frequencies on average, and also middling levels of compliance, raising the question: *How do they outperform the baseline scheme?*

Although the baseline achieves high rates of compliance in the signaling case, we did not achieve compliance in the no-signaling case with any of the algorithms. (The average attack rate on instances of no-signal was upwards of 96% across all conditions.) As Figure 6 (middle) shows, the baseline has a much higher rate of no-signal instances, which are almost always attacked, resulting in high losses for the defender. The `DT` and `NN` schemes give up some compliance in the case of a signal by signaling in a more middling range, but make up for this loss by

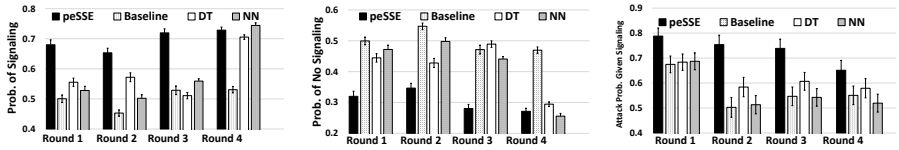

Fig. 6: The balance between lowering signaling probability, to increase compliance, and allowing many instances with no signal. (left) Probability of Showing a Signal. (middle) Probability of Showing No Signal. (right) Probability of Attack Given a Signal.

having less no signal instances. In general, lowering the signaling frequency can increase compliance with regard to signals, but must be carefully balanced so that instances in which no signal is shown do not offset the gain to the defender.

The learning based approaches do not just find a uniform frequency of signaling somewhere between $f_t = 0.75$ and $f_t = 1$. As mentioned in Section 6.3, the `NN`-based algorithm tends to increase the rate of signaling on less desirable targets, while decreasing it on more popular targets. This varied signaling frequency tuned to the features of each target is what causes the middling range of frequency on average, and also what allows the model-based algorithms to outperform the baseline, by performing better at the target level. Additional discussion of the performance of individual targets can be found in the appendix.

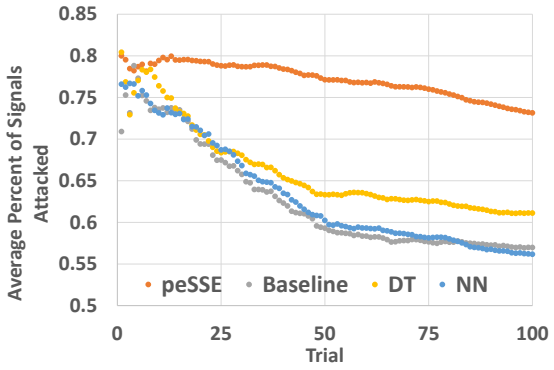

Fig. 7: The average percent of signals attacked (y-axis) up to the current trial (x-axis).

Exploiting the Training Effect. Boundedly rational subjects adjust their response to signaling given repeated exposure to signals and the consequences of attacking. The y-axis of Figure 7 shows the average percent of signals attacked up to the

current trial, which is given on the x-axis. It shows that initially subjects behave in a very exploratory manner, attacking frequently. However, as time passes they become more compliant. In rounds 1-3, the average rate of signaling of the `NN` signaling scheme falls between the `peSSE` and baseline algorithms, but in round 4 the `NN` signaling scheme is equal to the `peSSE` scheme (see Figure 6 (right)). Yet, the the defender's expected utility is significantly better than in the `peSSE` experiment. As expected, using 2-way signaling in rounds 1-3 leads to an increased rate of compliance by the final round, as well as a sharp and early drop in attack probability over the course of the first two rounds, compared to the `peSSE` scheme. Boosting the level of signaling in the fourth round exploits this improved compliance rate, taking advantage of the benefit of signaling to increase the defender's expected utility. We see a similar effect with the `DT` algorithm. However, this effect is not exploited by the baseline algorithm, which uniformly reduces signaling in all four rounds, and actually performs significantly worse than the `peSSE` scheme in round 4 (see Figure 5), even though the level of compliance with signaling is much lower.

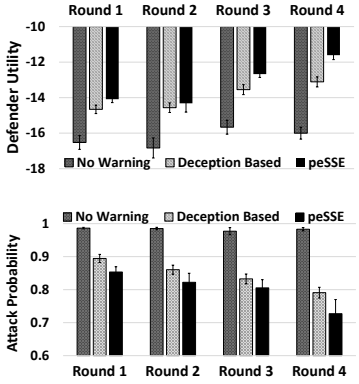

Fig. 8: Average expected defender utility (top) and attack probability (bottom) between the deception-based, `peSSE`, and no-signaling schemes.

Only Learning the Deception Rate Does Not Work. As discussed in Section 4, we compared `peSSE` with a regression tree-based algorithm that learned the optimal deception rate, but *ignored signaling frequency*. This method led to significantly lower defender expected utility ($p < 0.01$) (Figure 8) and significantly higher attack probability ($p < 0.06$) (Figure 8), as by not accounting for the frequency of signaling, it signals too much, causing subjects to become desensitized and non-compliant.

8 Conclusions and Future Work

We have shown that using machine learning to model an attacker's response to deceptive signaling leads to an optimal signaling scheme to deter boundedly rational attackers. We present decision tree- and neural network-based signaling schemes to find the *Goldilocks zone* for signaling. We show via human subject experiments that learning-based signaling schemes improve defender performance, and that these schemes lead humans to become more compliant over repeated interaction. Whereas our results are based on the Mechanical Turk population and game setting, further testing should use realistic simulation with expert participants [6] or even occur "in the wild" [5]. Personalized signaling schemes [3] and defending against adversary manipulation of the system should also be studied.

Acknowledgments. This research was sponsored by the Army Research Office and accomplished under MURI Grant Number W911NF-17-1-0370.

References

1. An, B., Tambe, M., Ordonez, F., Shieh, E., Kiekintveld, C.: Refinement of strong stackelberg equilibria in security games. In: Twenty-Fifth AAAI (2011)
2. Basilico, N., Gatti, N.: Strategic guard placement for optimal response toalarms in security games. In: Proceedings of the 2014 AAMAS. pp. 1481–1482 (2014)
3. Cranford, E.A., Gonzalez, C., Aggarwal, P., Cooney, S., Tambe, M., Lebiere, C.: Stowards personalized deceptive signaling for cyber defense using cognitive models. In: Proceedings of the Proceedings of the 17th ICCM. p. (in press) (2019)
4. Cranford, E.A., Lebiere, C., Gonzalez, C., Cooney, S., Vayanos, P., Tambe, M.: Learning about cyber deception through simulations: Predictions of human decision making with deceptive signals in stackelberg security games. In: CogSci 2018. pp. 25–28 (2018)
5. Delle Fave, F.M., Brown, M., Zhang, C., Shieh, E., Jiang, A.X., Rosoff, H., Tambe, M., Sullivan, J.: Security games in the field: an initial study on a transit system. In: Proceedings of the 2014 AAMAS. pp. 1363–1364 (2014)
6. Ferguson-Walter, K., Shade, T., Rogers, A., Niedbala, E., Trumbo, M., Nauer, K., Divis, K., Jones, A., Combs, A., Abbott, R.: The tularosa study: An experimental design and implementation to quantify the effectiveness of cyber deception. In: Proceedings of the 52nd Hawaii International Conference on System Sciences (2019)
7. Fraunholz, D., Anton, S.D., Lipps, C., Reti, D., Krohmer, D., Pohl, F., Tammen, M., Schotten, H.D.: Demystifying deception technology: A survey. arXiv preprint arXiv:1804.06196 (2018)
8. Gholami, S., Mc Carthy, S., Dilkina, B., Plumptre, A., Tambe, M., Driciru, M., Wanyama, F., Rwetsiba, A., Nsubaga, M., Mabonga, J., et al.: Adversary models account for imperfect crime data: Forecasting and planning against real-world poachers (2018)
9. Gholami, S., Yadav, A., Tran-Thanh, L., Dilkina, B., Tambe, M.: Don't put all your strategies in one basket: Playing green security games with imperfect prior knowledge. In: Proceedings of the 18th AAMAS. pp. 395–403 (2019)
10. Guo, Q., An, B., Bosanskỳ, B., Kiekintveld, C.: Comparing strategic secrecy and stackelberg commitment in security games. In: IJCAI. pp. 3691–3699 (2017)
11. Hartford, J.S., Wright, J.R., Leyton-Brown, K.: Deep learning for predicting human strategic behavior. In: NIPS. pp. 2424–2432 (2016)
12. He, X., Islam, M.M., Jin, R., Dai, H.: Foresighted deception in dynamic security games. In: 2017 IEEE ICC. pp. 1–6 (2017)
13. Kar, D., Ford, B., Gholami, S., Fang, F., Plumptre, A., Tambe, M., Driciru, M., Wanyama, F., Rwetsiba, A., Nsubaga, M., et al.: Cloudy with a chance of poaching: adversary behavior modeling and forecasting with real-world poaching data. In: Proceedings of the 16th AAMAS. pp. 159–167 (2017)
14. Korzhyk, D., Conitzer, V., Parr, R.: Complexity of computing optimal stackelberg strategies in security resource allocation games. In: Twenty-Fourth AAAI (2010)
15. Kraus, S.: Predicting human decision-making: From prediction to action. In: Proceedings of the 6th HAI. pp. 1–1. ACM (2018)
16. Krol, K., Moroz, M., Sasse, M.A.: Don't work. can't work? why it's time to rethink security warnings. In: risk and security of internet and systems (CRiSIS), 2012. pp. 1–8. IEEE (2012)
17. Luber, S., Yin, Z., Delle Fave, F.M., Jiang, A.X., Tambe, M., Sullivan, J.P.: Game-theoretic patrol strategies for transit systems: the trusts system and its mobile app. In: AAMAS. pp. 1377–1378. Citeseer (2013)

18. Maimon, D., Alper, M., Sobesto, B., Cukier, M.: Restrictive deterrent effects of a warning banner in an attacked computer system. Criminology **52**(1), 33–59 (2014)
19. Nguyen, T.H., Wang, Y., Sinha, A., Wellman, M.P.: Deception in finitely repeated security games. In: 33th AAAI (2019)
20. Nguyen, T.H., Yang, R., Azaria, A., Kraus, S., Tambe, M.: Analyzing the effectiveness of adversary modeling in security games. In: AAAI (2013)
21. Okamoto, S., Hazon, N., Sycara, K.: Solving non-zero sum multiagent network flow security games with attack costs. In: Proceedings of the 11th AAMAS-Volume 2. pp. 879–888 (2012)
22. Pawlick, J., Zhu, Q.: Deception by design: evidence-based signaling games for network defense. arXiv preprint arXiv:1503.05458 (2015)
23. Pita, J., Jain, M., Ordóñez, F., Tambe, M., Kraus, S., Magori-Cohen, R.: Effective solutions for real-world stackelberg games: When agents must deal with human uncertainties. In: Proceedings of The 8th AAMAS-Volume 1. pp. 369–376 (2009)
24. Smith, S.W.: Security and cognitive bias: exploring the role of the mind. IEEE Security & Privacy **10**(5), 75–78 (2012)
25. Sobel, J.: Signaling games. Encyclopedia of Complexity and Systems Science **19**, 8125–8139 (2009)
26. Tambe, M.: Security and game theory: algorithms, deployed systems, lessons learned. Cambridge University Press (2011)
27. Timofeev, R.: Classification and regression trees (cart) theory and applications. Humboldt University, Berlin (2004)
28. Wang, B., Zhang, Y., Zhou, Z.H., Zhong, S.: On repeated stackelberg security game with the cooperative human behavior model for wildlife protection. Applied Intelligence **49**(3), 1002–1015 (2019)
29. Wilczyński, A., Jakóbik, A., Kołodziej, J.: Stackelberg security games: Models, applications and computational aspects. Journal of Telecommunications and Information Technology (2016)
30. Xu, H., Rabinovich, Z., Dughmi, S., Tambe, M.: Exploring information asymmetry in two-stage security games. In: AAAI. pp. 1057–1063 (2015)
31. Xu, H., Wang, K., Vayanos, P., Tambe, M.: Strategic coordination of human patrollers and mobile sensors with signaling for security games. In: Thirty-Second AAAI (2018)
32. Yang, R., Kiekintveld, C., Ordonez, F., Tambe, M., John, R.: Improving resource allocation strategy against human adversaries in security games. In: Twenty-Second IJCAI (2011)
33. Yang, R., Kiekintveld, C., OrdóñEz, F., Tambe, M., John, R.: Improving resource allocation strategies against human adversaries in security games: An extended study. Artificial Intelligence **195**, 440–469 (2013)
34. Zhang, C., Jiang, A.X., Short, M.B., Brantingham, P.J., Tambe, M.: Defending against opportunistic criminals: New game-theoretic frameworks and algorithms. In: GameSec. pp. 3–22. Springer (2014)
35. Zhang, C., Sinha, A., Tambe, M.: Keeping pace with criminals: Designing patrol allocation against adaptive opportunistic criminals. In: Proceedings of the 2015 AAMAS. pp. 1351–1359 (2015)
36. Zhuang, J., Bier, V.M., Alagoz, O.: Modeling secrecy and deception in a multiple-period attacker–defender signaling game. European Journal of Operational Research **203**(2), 409–418 (2010)