Exploiting Bounded Rationality in Risk-based Cyber Camouflage Games

Omkar Thakoor[1], Shahin Jabbari[2], Palvi Aggarwal[3], Cleotilde Gonzalez[3], Milind Tambe[2], and Phebe Vayanos[1]

[1] University of Southern California, Los Angeles, CA 90007, USA
{othakoor, phebe.vayanos}@usc.edu
[2] Harvard University, Cambridge, MA 02138, USA
{jabbari@seas., milind_tambe@}harvard.edu
[3] Carnegie Mellon University, Pittsburgh, PA 15213, USA
{palvia@andrew., coty@}cmu.edu

Abstract. Recent works have growingly shown that *Cyber deception* can effectively impede the reconnaissance efforts of intelligent cyber attackers. Recently proposed models to optimize a deceptive defense based on camouflaging network and system attributes, have shown effective numerical results on simulated data. However, these models possess a fundamental drawback due to the assumption that an attempted attack is always successful — as a direct consequence of the deceptive strategies being deployed, the attacker runs a significant risk that the attack fails. Further, this risk or uncertainty in the rewards magnifies the boundedly rational behavior in humans which the previous models do not handle. To that end, we present Risk-based Cyber Camouflage Games — a general-sum game model that captures the uncertainty in the attack's success. In case of the rational attackers, we show that optimal defender strategy computation is NP-hard even in the zero-sum case. We provide an MILP formulation for the general problem with constraints on cost and feasibility, along with a pseudo-polynomial time algorithm for the special *unconstrained* setting. Second, for risk-averse attackers, we present a solution based on Prospect theoretic modeling along with a robust variant that minimizes regret. Third, we propose a solution that does not rely on the attacker behavior model or past data, and effective for the broad setting of *strictly competitive games* where previous solutions against bounded rationality prove ineffective. Finally, we provide numerical results that our solutions effectively lower the defender loss.

Keywords: Game Theory · Cyber Deception · Rationality

1 Introduction

Rapidly growing cybercrime [15, 13, 24], has elicited effective defense against adept attackers. Many recent works have proposed *Cyber deception* techniques to thwart the reconnaissance — typically a crucial phase prior to attacking [21, 17]. One deception approach is to camouflage the network by attribute obfuscation [10, 35, 7] to render an attacker's information incomplete or incorrect, creating indecision over their infiltration plan [12, 10, 4, 28]. Optimizing such a deceptive strategy is challenging due to many practical constraints on feasibility and costs of deploying, as well as critically dependent on the attacker's decision-making governed by his behavioral profile, and attacking motives and capabilities. Game theory offers an effective framework for tackling both these aspects and has been successfully adopted in security problems [2, 20, 31, 29].

Attacking a machine amounts to launching an exploit for a particular system configuration — information that is concealed or distorted due to the deceptive defense, thus, an attempted attack may not succeed. Recent game theoretic models for deception via attribute obfuscation [30, 34] have a major shortcoming in ignoring this risk of attack failure as they assume that an attempted attack is guaranteed to provide utility to the attacker. Further, results from recent human subject studies [1] suggest that this risk may unveil risk-aversion in human attackers rather than a perfectly rational behavior of maximizing expected utility that the models assume. Apart from risk-aversion, other behavioral models, e.g., the Quantal response theory [22], also assert that humans exhibit bounded rationality. This can severely affect the performance of a deployed strategy, which has not been considered by the previous works.

As our first main contribution, we present Risk-based Cyber Camouflage Games (RCCG) — a crucial refinement over previous models via redefined strategy space and rewards to explicitly capture the uncertainty in attack success. As foundation, we first consider rational attackers and show analytical results including NP-hardness of optimal strategy computation and its MILP formulation which, while akin to previous models, largely require independent reasoning. Further, we consider risk-averse attackers modeled using Prospect theory [36] and present a solution (PT) that estimates model parameters from data to compute optimal defense. To circumvent the limitations of parametrization and learning errors, we also present a robust solution (MMR) that minimizes worst-case regret for a general prospect theoretic attacker. Finally, we propose a solution ($GEBRA$) free of behavioral modeling assumptions and avoiding reliance on data altogether, that can exploit arbitrary deviations from rationality. Our numerical results show the efficacy of our solutions summarized at the end.

1.1 Related work

Cyber Deception Games [30], and Cyber Camouflage Games (CCG) [34] are game-theoretic models for Cyber deception via attribute obfuscation. In these, the defender can mask the *true configuration* of a machine, creating an uncertainty in the associated reward the attacker receives for attacking the machine. These have a fundamental limitation, namely, the assumption that the attacked machine is guaranteed to provide utility to the attacker. Further, they do not consider that human agents tend to deviate from rationality, particularly when making decisions under risk. Our refined model handles both these crucial issues.

A model using Prospect theory is proposed in [38] for boundedly rational attackers in Stackelberg security games (SSG) [33]. However, it relies on using model parameters from previous literature, discounting the fact that they can largely vary for the specific experimental setups. We provide a solution that learns the parameters from data, as well as a robust solution to deal with uncertainty in the degree of risk-aversion and broadly the parametrization hypothesis. A robust solution for unknown risk-averse attackers has been proposed for SSGs in [27], however, it aims to minimize the worst-case utility, whereas, we take the less conservative approach of minimizing worst-case regret. Previous works on uncertainty in security games consider Bayesian [18], interval-based [19], and regret-based approaches [23], however, these do not directly apply due to fundamental differences between RCCGs and SSGs as explained in [34].

Another approach in [38] is based on the Quantal Response model [22]. However, the attack probabilities therein involve terms that are exponential in rewards, which in turn are non-linear functions of integer variables in our model, leading to an intractable formulation. However, we show effectiveness of our model-free solution for this behavior model as well.

Machine learning models such as Decision Tree and Neural Networks have been used for estimating human behavior [8]. However, the predictive power of such models typically comes with an indispensable complexity (non-linear kernels, functions and deep hidden layers of neural nets, sizeable depth and branching factor of decision trees etc). This does not allow the predicted human response to be written as a simple closed-form expression of the instance features, viz, the strategy decision variables, preventing a concise optimization problem formulation. This is particularly problematic since the alternative of searching for an optimal solution via strategy enumeration is also non-viable — due to the compact input representation via a *polytopal* strategy space [16] in our model.

MATCH [25] and COBRA [26] aim to tackle human attackers in SSGs that avoid the complex task of modeling human decision-making and provide robustness against deviations from rationality. However, their applicability is limited — in *Strictly Competitive* games where deviation from rationality always benefits the defender, they reduce to the standard minimax solution. Our model-free solution GEBRA on the other hand, achieves better computational results than minimax, and MATCH can be seen as its conservative derivative.

2 Risk-based Cyber Camouflage Games (RCCG) model

Here, we describe the components of the RCCG model, explicitly highlighting the key differences with respect to the CCG model [34].

Network Configurations. The network is a set of k machines $\mathcal{K} := \{1, \ldots, k\}$. Each machine has a *true configuration* (TC), which is simply an exhaustive tuple of attributes so that machines having the same TC are identical. $\mathcal{S} := \{1, \ldots, s\}$ is the set of all TCs. The *true state of the network* (TSN) is a vector $\boldsymbol{n} = (n_i)_{i \in \mathcal{S}}$ with n_i denoting the number of machines with TC i. Note that $\sum_{i \in \mathcal{S}} n_i = k$.

The defender can disguise the TCs using deception techniques. Each machine is "masked" with an *observed configuration* (OC). The set of OCs is denoted by $\mathcal{T}$. Similar to a TC, an OC corresponds to an attribute tuple that fully comprises the attacker view, so that machines with the same OC are indistinguishable.

Deception Strategies. We represent the defender strategy as an integer matrix Φ, where Φ_{ij} is the no. of machines with TC i, masked with OC j. The *observed* *state of the network* (OSN) is a function of $\varPhi$, denoted as $\boldsymbol{m}(\varPhi) := (m_j(\varPhi))_{j\in\mathcal{T}}$, where $m_j(\varPhi) = \sum_i \varPhi_{ij}$ denotes the no. of machines under OC j for strategy $\varPhi$.

Deception feasibility and costs. Achieving deception is often costly and not arbitrarily feasible. We have *feasibility* constraints given by a (0,1)-matrix $\varPi$, where $\varPi_{ij} = 1$ if a machine with TC i can be masked with OC j. Next, we assume that masking a TC i with an OC j (if so feasible), has a cost of c_{ij} incurred by the defender, denoting the aggregated cost from deployment, maintenance, degraded functionality, etc. We assume the total cost is to be bounded by a *budget* B.

These translate to linear constraints to define the valid defender strategy set:

$$\mathcal{F} = \left\{ \varPhi \,\middle|\, \begin{matrix} \varPhi_{ij} \in \mathbb{Z}_{\geq 0}, \quad \varPhi_{ij} \leq \varPi_{ij} n_i \ \forall (i,j) \in \mathcal{S} \times \mathcal{T}, \\ \sum\limits_{j\in\mathcal{T}} \varPhi_{ij} = n_i \quad \forall i \in \mathcal{S}, \quad \sum\limits_{i\in\mathcal{S}} \sum\limits_{j\in\mathcal{T}} \varPhi_{ij}\ c_{ij} \leq B \end{matrix} \right\}$$

The first and the third constraints follow from the definitions of $\varPhi$ and $\boldsymbol{n}$. The second imposes the feasibility constraints, and the fourth, the budget constraint.

Remark 1. The budget constraint can encode feasibility constraints as a special case by setting a cost higher than the budget for an infeasible masking. The latter are still stated explicitly for the useful interpretation and practical significance.

Defender and Attacker Valuations. A machine with TC i gets successfully attacked if the attacker uncovers the disguised OC and uses the correct exploit corresponding to TC i. In such a case, the attacker gets a utility v_i — his *valuation* of TC i. Collectively, these are represented as a vector $\boldsymbol{v}$. Analogously, we define valuations $\boldsymbol{u}$ representing the defender's loss.

Remark 2. For ease of interpretation, we assign a 0 utility to the players when the attack is unsuccessful, which sets a constant reference point. Hence, unlike CCGs, valuations cannot be freely shifted. Further, a successful attack typically is undesirable for the defender (except, e.g., honeypots), and to let the valuations be typically positive values, they represent the defender's loss; its minimization is the defender objective unlike maximization in CCGs.

Attacker Strategies. As the attacker cannot distinguish between machines with the same OC, he chooses an OC from which to attack a random machine. Attacking a machine requires choosing an exploit to launch for a particular TC. Thus, the attack can be described as a pair of decisions $(i,j) \in \mathcal{S} \times \mathcal{T}$. This significant difference in attack strategy space definition and the imminent player utility definitions as a consequence, cause the fundamental distinction in the practical scope as well as the technical solutions of the RCCG model.

We model the interaction as a Stackelberg game to capture the order of player decisions. The defender is the leader who knows the TSN $\boldsymbol{n}$ and can deploy a deception strategy $\varPhi$, and the attacker chooses a pair $(i,j) \in \mathcal{S} \times \mathcal{T}$. The defender can only play a pure strategy since it is typically not possible to change the network frequently, making the attacker's view of the network static. As in Schlenker et al. [30], Thakoor et al. [34], we assume the attacker can use the defender's strategy $\varPhi$ to perfectly compute the utilities from different attacks, which is justified via insider information leakage or other means of surveillance.

Suppose the defender plays a strategy $\varPhi$, and the attacker attacks using an exploit for TC i on a machine masked with OC j. Among $m_j(\varPhi)$ machines masked by OC j, $\varPhi_{ij}$ are of TC i. Hence, the attack is successful with a probability $\frac{\varPhi_{ij}}{m_j(\varPhi)}$. Consequently, the player utilities are given by

$$U^{\mathrm{a}}(\varPhi, i, j) = \frac{\varPhi_{ij}}{m_j(\varPhi)} v_i \quad , \quad U^{\mathrm{d}}(\varPhi, i, j) = \frac{\varPhi_{ij}}{m_j(\varPhi)} u_i. \tag{1}$$

Note that these expressions imply that if the player valuations ($\boldsymbol{v}$ or $\boldsymbol{u}$) are simultaneously scaled by a positive constant (for normalization etc.), it preserves the relative order of player utilities, and in particular, the best responses to any strategies, thus keeping the problem equivalent.

Next, we show analytical results on optimal strategy computation for a rational attacker, which lay the foundation for further tackling bounded rationality.

3 Rational attackers

The attacker having to choose a TC-OC pair as an attack here rather than just an OC as in the CCG model [34], requires entirely new techniques for our analytical results, despite close resemblance in the optimization problem as below.

Optimization problem Previous work on general-sum Stackelberg games has typically used *Strong Stackelberg equilibria* (SSE). This assumes that in case of multiple best responses, the follower breaks ties in favor of the leader (i.e., minimizing defender loss). The leader can *induce* this with mixed strategies, which is not possible in RCCGs as the defender is restricted to pure strategies [14].

Hence, we consider the worst-case assumption that the attacker breaks ties against the defender, leading to *Weak Stackelberg Equilibria* (WSE) [6]. WSE may not always exist [37], but it does when the leader can only play a finite set of pure strategies as in CCG. Hence, we assume that the attacker chooses a best response to the defender strategy $\varPhi$, maximizing the defender loss in case of a tie. This defender utility is denoted as $U^{\mathrm{wse}}(\varPhi)$, defined as the optimal value of the inner Optimization Problem (OP) in the following, while the defender aims to compute a strategy to minimize $U^{\mathrm{wse}}(\varPhi)$ as given by the outer objective.

$$\begin{aligned} \underset{\varPhi}{\operatorname{argmin}} \quad & \max_{i,j} U^{\mathrm{d}}(\varPhi, i, j) \\ & \text{s.t. } U^{\mathrm{a}}(\varPhi, i, j) \geq U^{\mathrm{a}}(\varPhi, i', j') \;\; \forall i' \in \mathcal{S}, \; \forall j' \in \mathcal{T}. \end{aligned} \tag{2}$$

Next, we show results on optimal strategy computation shown for the important special cases — the zero-sum and *unconstrained* settings. While similar results have been shown for CCG, independent proof techniques are needed herein due to a distinctive model structure (see Appendix for omitted proofs).

3.1 Zero-sum RCCG

In the zero-sum setting, the defender loss equals the attacker reward, i.e. $\boldsymbol{v} = \boldsymbol{u}$.

Theorem 1. *Zero-sum RCCG is NP-hard.*

Proof Sketch. We reduce from the problem "Exact Cover by 3-Sets" ($ExC3$ for brevity) which is known to be NP-complete. Given an instance of $ExC3$, we construct an instance of RCCG for which the minimum defender loss is precisely equal to a certain value if and only if the given $ExC3$ instance is YES. □

For the special unconstrained setting[4](i.e. with no feasibility or budget constraints), we show the following.

Proposition 1. *Unconstrained zero-sum RCCG always has an optimal strategy that uses just one OC, thus computable in* $O(1)$ *time.*

Thus, both these results hold for RCCG, same as for CCG (albeit, they do not follow from the latter, requiring independent derivation).

3.2 Unconstrained General-sum RCCG

Proposition 2. *Unconstrained RCCG always has an optimal strategy that uses just two OCs.*

This result is crucial for an efficient algorithm to compute an optimal strategy (Algorithm 1), named Strategy Optimization by Best Response Enumeration (SOBRE). SOBRE constructs an optimal strategy with two OCs, due to Proposition 2, with attacker best response being (say) OC 1 (Note: this is without loss of generality in the unconstrained setting). It classifies the candidate strategies by triplets (i, n, m) (Line 2) where the attacker best response is $(i, 1)$, and OC 1 masks n machines of TC i, and m machines in total. It uses a subroutine DPBRF (Dynamic Programming for Best Response Feasibility) to construct a strategy yieldsing the desired best response (Line 6) if it exists, and then compares the defender utility from all such feasible candidates, to compute the optimal (Lines 7,8). For details on DPBRF and runtime analysis, refer to the Appendix.

Algorithm 1: SOBRE

```
1 Initialize $minUtil \leftarrow \infty$
2 for $i = 1, \ldots, s; n = 0, \ldots n_i; m = n, \ldots, k$ do
3     if $(n/m < (n_i - n)/(|K| - m))$ continue
4     $util \leftarrow (n/m)u_i$
5     if $(util \geq minUtil)$ continue
6     if $DPBRF(i, n, m)$
7         Update $minUtil \leftarrow util$
8 Return $minUtil$
```

Theorem 2. *The optimal strategy in an unconstrained RCCG can be computed in time* $O(k)^4$.

Remark 3. Note that the input can be expressed in $O(st)$ bits, which makes this algorithm pseudo-polynomial. However, it becomes a poly-time algorithm under the practical assumption of constant-bounded no. of machines per TC, (so that,

[4] The *unconstrained* setting accents the inherent challenge of strategic deception even when sophisticated techniques can arbitrarily mask TCs with any OCs at low cost.

$k = O(s)$, or more generally, if k in terms of s is polynomially bounded). In contrast, unconstrained CCG is NP-hard even under this restriction. This distinction arises since in RCCG, the best response utility given the attack strategy and the no. of machines masked by the corresponding OC, depends on only the count of attacked TC as opposed to all the TCs in CCG.

3.3 Constrained General-sum RCCG

For this general setting of RCCG, $U^{\mathrm{wse}}(\Phi)$ is given by OP (2), and thus, computing its minimum is a bilevel OP. Reducing to a single-level Mixed Integer Linear Program (MILP) is typically hard [32]. (in particular, computing an SSE allows such a reduction due to attacker's tiebreaking favoring the defender's objective therein, however, the worst-case tiebreaking of WSE does not). Notwithstanding the redefined attack strategies, a single-level OP can be formulated analogous to CCGs by assuming an ϵ-rational attacker instead of fully rational (as it can be shown that for sufficiently small ϵ, it gives the optimal solution for rationality):

$$\begin{aligned}
&\min_{\Phi,\boldsymbol{q},\gamma,\alpha} \ \gamma && && (3)\\
&\text{s.t.} \quad \alpha,\gamma \in \mathbb{R},\ \Phi \in \mathcal{F},\ \boldsymbol{q} \in \{0,1\}^{|\mathcal{I}|\times|\mathcal{J}|} && && \\
&\qquad q_{11} + \ldots + q_{st} \geq 1 && && (3a)\\
&\qquad \epsilon(1-q_{ij}) \leq \alpha - U^{\mathrm{a}}(\Phi,j,i) && \forall i \in \mathcal{S}\ \forall j \in \mathcal{T} && (3b)\\
&\qquad M(1-q_{ij}) \geq \alpha - U^{\mathrm{a}}(\Phi,j,i) && \forall i \in \mathcal{S}\ \forall j \in \mathcal{T} && (3c)\\
&\qquad U^{\mathrm{d}}(\Phi,j,i) \leq \gamma + M(1-q_{ij}) && \forall i \in \mathcal{S}\ \forall j \in \mathcal{T} && (3d)\\
&\qquad q_{ij} \leq \Phi_{ij} && \forall i \in \mathcal{S}\ \forall j \in \mathcal{T}. && (3e)
\end{aligned}$$

The defender aims to minimize the objective γ which captures the defender's optimal utility. The binary variables q_{ij} indicate if attacking (i, j) is an optimal attacker strategy, and as specified by (3a), there must be at least one. As per (3b) and (3c), α is the optimal attacker utility, and this enforces $q_{ij} = 1$ for all the ϵ-optimal attacker strategies (using a big-M constant). (3e) ensures that only the OCs which actually mask a machine are considered as valid attacker responses. Finally, (3d) captures the worst-case tie-breaking by requiring that γ is the highest defender loss from a possible ϵ-optimal attacker response. Using an alternate strategy representation with binary decision variables enables linearization to an MILP, that can be sped up with *symmetry-breaking* cuts [34].

Next, we consider human attackers who typically exhibit bounded rationality.

4 A Model-driven Approach with Prospect Theory

A well-studied model for the risk-behavior of humans is Prospect theory [36]. As per this, humans under risk make decisions to maximize the *prospect*, which differs from the utilitarian approach in that the reward value and the probability of any event are transformed as follows. We have a value transformation function R that is monotone increasing and concave, s.t., the outcome reward v (value of the machine attacked), gets perceived as $R(v)$ by the attacker. A parameterization of the form $R_\lambda(v) = c(v/c)^\lambda$ is commonly considered in the literature, with $\lambda < 1$ capturing the risk-aversion of the attacker[5], and we use $c = \max_i v_i$ so that the perceived values are normalized to the same range as true values. Prospect theory also proposes a probability weighting function Π, such that the probability p of an event is perceived as $\Pi(p)$. A function of the form $\Pi_\delta(p) = p^\delta/(p^\delta + (1-p)^\delta)^{1/\delta}$ has been previously proposed in literature, parametrized by δ. In our problem, the attack success probability p is a non-linear non-convex function of the decision variables Φ_{ij} and applying a function as above loses tractability. For simplicity, we omit the probability weighting from our solution which shows effective results regardless. Future work could explore the benefits of incorporating this additional complexity.

Thus, each of the attacker's strategies (i, j) has a prospect

$$f_\lambda(\Phi, i, j) = \frac{\Phi_{ij}}{m_\Phi(j)} R_\lambda(v_i) \tag{4}$$

as a function of the player strategies, parametrized by λ. This value transformation makes the problem inherently harder (even in the simpler zero-sum setting).

Learning the parameter λ is a key challenge. Once λ is estimated, the defender computes an optimal strategy for the prospect theoretic attacker, by simply modifying (3), replacing the valuations v_i with the transformed values $R_\lambda(v_i)$. More generally, with this replacement, all results from Section 3 for rational attackers apply here too.

4.1 Learning model parameters from data

Suppose we have data consisting of a set of instances $\mathcal{N}$ from a study such as [1]. A particular instance $n \in \mathcal{N}$ corresponds to a particular human subject that plays against a particular defense strategy Φ_n, and decides to attack (i_n, j_n) having the maximum prospect. The instances come from different subjects who may have a different parameter λ, however, at the time of deployment, the defender cannot estimate the risk-averseness of an individual in advance and play a different strategy accordingly. Hence, we aim to compute a strategy against a particular λ that works well for the whole population[6]. Due to different subjects, different instances may have different attack responses for the same defender strategy, and requiring a strict prospect-maximization may not yield any feasible λ. Hence, we define the likelihood of an instance, by considering a soft-max function instead, so that the probability of attacking (i_n, j_n) is[7]

$$P_n(\lambda) = \frac{\exp(f_\lambda(\Phi_n, i_n, j_n))}{\sum_{i,j} \exp(f_\lambda(\Phi_n, i, j))}.$$

Using the Maximum Likelihood Estimation approach, we choose λ which maximizes the likelihood $\prod_n P_n(\lambda)$, or, log likelihood $\sum_n \log P_n(\lambda)$. (Note: Manually

[5] The conventional usage of the symbol λ in prospect theoretic models is different.

[6] This avoids learning a complex distribution of λ from limited data, and the subsequent need for a Bayesian game formulation with attackers coming from a continuous distribution which is not expressible as an MILP

[7] When considering a continuous range of λ for payoff transformations, the degenerate cases of tie-breaking between strategies are zero-probability events and thus ignored.

eliminating anomalous instances from data which indicate complete irrationality can help avoid over-fitting). Finding such a solution via the standard approach of *Gradient Descent* does not have the convergence guarantee due to the likelihood being non-convex and we resort to *Grid Search* instead.

4.2 Robust solution with Prospect Theory

The learning error can be sizeable if the subject population has a high variance of λ or if limited data is available (for sensitivity analysis, see Appendix). Further, the parameterization hypothesis may not fit well, degrading solution quality. To circumvent both these issues, we propose a solution offering robustness when the attacker behavior cannot be predicted with certainty. We assume a prospect-theoretic attacker, but with no assumption of a parametrized model or data availability. Thus, the defender knowledge of value transformations has uncertainty, which we handle with the minimax regret framework [9, 5], seen to be less conservative in contrast with a purely maximin approach that focuses on the worst cases of uncertainty.

Value transformation and Uncertainty modelling We assume the attacker has the transformed values $\boldsymbol{w}$. Defender does not precisely know $\boldsymbol{w}$ which can be anything from a set $\mathcal{W} \subseteq \mathbb{R}^s$ which we call the *uncertainty set* [3]. $\mathcal{W}$ is obtained by requiring that the transformation from $\boldsymbol{v}$ to $\boldsymbol{w}$ is a monotone increasing and concave function with $\boldsymbol{w}$ normalized to the same range as valuations $\boldsymbol{v}$. WLOG, let v be sorted increasingly in the index. Then, $\mathcal{W}$ is defined by the constraints

$$\mathcal{W} = \left\{ w \,\middle|\, \begin{array}{l} 0 \leq w_1 \leq w_2 \dots w_s = v_s \\ \frac{w_1}{v_1} \geq \frac{w_2}{v_2} \geq \dots \geq \frac{w_s}{v_s} = 1 \end{array} \right\}$$

The first constraints ensure monotonicity, and the second ones convexity. An equivalent formulation can also be obtained by adapting constraints used in [27].

Minmax Regret Formulation Let $U^{\mathrm{a}}(\Phi, i, j, \boldsymbol{w})$ denote the attacker's prospect in terms of $\boldsymbol{w}$ and the player strategies. Similarly, let the defender's wse utility in terms of $\boldsymbol{w}$ be denoted by $U^{\mathrm{wse}}(\Phi, \boldsymbol{w})$ defined analogous to $U^{\mathrm{wse}}(\Phi)$ in (2):

$$\max_{i,j} U^{\mathrm{d}}(\Phi, i, j) \mid U^{\mathrm{a}}(\Phi, i, j, \boldsymbol{w}) \geq U^{\mathrm{a}}(\Phi, i', j', \boldsymbol{w}) \; \forall i' \in \mathcal{S} \; \forall j' \in \mathcal{T}. \tag{5}$$

Then, the *max regret* (MR) of Φ is the worst-case value over all $\boldsymbol{w} \in \mathcal{W}$ of the decrements in defender loss that the optimal $\hat{\Phi}$ achieves over Φ for valuations $\boldsymbol{w}$:

$$\mathrm{MR}(\Phi) = \max_{\boldsymbol{w} \in \mathcal{W}} \max_{\hat{\Phi} \in \mathcal{F}} \left[U^{\mathrm{wse}}(\Phi, \boldsymbol{w}) - U^{\mathrm{wse}}(\hat{\Phi}, \boldsymbol{w}) \right]. \tag{6}$$

The minmax regret (MMR) approach looks to compute the Φ that minimizes $\mathrm{MR}(\Phi)$, i.e., solving the following OP:

$$\min_{\Phi \in \mathcal{F}, \beta} \quad \beta \mid \beta \geq U^{\mathrm{wse}}(\Phi, \boldsymbol{w}) - U^{wse}(\hat{\Phi}, \boldsymbol{w}) \;\; \forall (\boldsymbol{w}, \hat{\Phi}) \in \mathcal{W} \times \mathcal{F}. \tag{7}$$

OP (7) has a constraint for each $(\boldsymbol{w}, \hat{\Phi}) \in \mathcal{W} \times \mathcal{F}$ making it a *semi-infinite program* as $\mathcal{W}$ is infinite, and difficult to solve also due to $\mathcal{F}$ being large. Hence, we adopt the well-studied approach of using *constraint sampling* [9] with *constraint generation* [5], to devise Algorithm 2. It iteratively computes successively tighter upper and lower bounds on MMR until they converge to the objective value. For the lower bound, we compute a relaxed version of OP (7), i.e., *relaxed MMR* by computing its objective subject to constraints corresponding to a sampled subset $\mathcal{S} = \{(\boldsymbol{w}_{(n)}, \hat{\varPhi}_{(n)})\}_n$ instead of $\mathcal{W} \times \mathcal{F}$ directly, giving an interim solution $\varPhi$ (line 4). Since only partial constraints were considered, the regret thus computed must be a lower bound on the true MMR. Next, if this interim solution is not optimal, there must be a constraint of OP (7) not satisfied by $\varPhi$. In particular, such a violated constraint can be found by computing the max regret (MR) of the interim solution $\varPhi$ (as per OP (6)) and by definition of max regret, must be an upper bound on the overall MMR (line 5). We use the new sample $(\boldsymbol{w}, \hat{\varPhi})$ thus computed and add to $\mathcal{S}$ (line 6) and repeat. We get successively tighter lower bounds as $\mathcal{S}$ grows and finally meets the tightest upper bound so far, which marks the convergence of the algorithm (line 3).

Algorithm 2: minmax regret computation

```
1 Initialize $u \leftarrow \infty, l \leftarrow 0$
2 Randomly generate samples $\mathcal{S} = \{(\boldsymbol{w}_{(n)}, \hat{\varPhi}_{(n)})\}_n$
3 while $u > l$ do
4     $l \leftarrow$ relaxed MMR w.r.t $\mathcal{S}$; giving interim solution $\varPhi$.
5     $u \leftarrow$ MR for $\varPhi$; giving a new sample $s = (\boldsymbol{w}, \hat{\varPhi})$.
6     Update $\mathcal{S} = \mathcal{S} \cup \{s\}$
7 Return incumbent solution as the true solution.
```

Next, we look at the two main subroutines of the algorithm.

(i) Relaxed MMR Computation. OP (7) has constraints for each $(\boldsymbol{w}, \hat{\varPhi}) \in \mathcal{W} \times \mathcal{F}$. Instead, considering a small subset of samples $\{(\hat{\varPhi}_{(n)}, \boldsymbol{w}_{(n)})\}_n \subseteq \mathcal{W} \times \mathcal{F}$ to generate a subset of constraints in (7) yields

$$\min_{\beta \in \mathbb{R}, \varPhi \in \mathcal{F}} \; \beta \mid \beta \geq U^{\text{wse}}(\varPhi, \boldsymbol{w}_{(n)}) - U^{\text{wse}}(\hat{\varPhi}_{(n)}, \boldsymbol{w}_{(n)}) \;\; \forall n. \tag{8}$$

This yields a lower bound on MMR since we consider fewer constraints. For sample n, let $\gamma_n = U^{\text{wse}}(\varPhi, \boldsymbol{w}_{(n)})$. Then, minimizing β translates to minimizing γ_n and this can be achieved by adding constraints analogous to (3a)$\sim$(3e) corresponding to each n, to obtain the following OP:

$$\begin{array}{ll} \min\limits_{\varPhi, \beta} & \beta \\ \text{s.t.} & \beta \in \mathbb{R}, \; \varPhi \in \mathcal{F} \\ & \left.\begin{array}{l} q_n \in \{0,1\}^{s \times t}, \; \alpha_n, \gamma_n \in \mathbb{R} \\ \beta \geq \gamma_n - U^{\text{wse}}(\hat{\varPhi}_{(n)}, \boldsymbol{w}_{(n)}) \\ \sum_{i,j} q_{nij} \geq 1 \end{array}\right\} \qquad\qquad \forall n \\ & \left.\begin{array}{l} \epsilon(1 - q_{nij}) \leq \alpha_n - U^{\text{a}}(\varPhi, i, j, \boldsymbol{w}_{(n)}) \\ M(1 - q_{nij}) \geq \alpha_n - U^{\text{a}}(\varPhi, i, j, \boldsymbol{w}_{(n)}) \\ U^{\text{d}}(\varPhi, i, j) \leq \gamma_n + (1 - q_{nij})M \\ q_{nij} \leq m_j(\varPhi). \end{array}\right\} \forall i \in \mathcal{S} \;\; \forall j \in \mathcal{T} \;\; \forall n \end{array}$$

(ii) Max Regret Computation. Here, we consider a candidate solution Φ, and compute a sample $(\Phi', \boldsymbol{w})$ which yields $MR(\Phi)$ as per (6), giving an upper bound on MMR by definition. Since $U^{\text{wse}}(\Phi, \boldsymbol{w})$ is defined via an optimization problem itself (given by (5)), (6) becomes a bilevel problem. To reduce it to single-level problems, we let $(i', j'), (i'', j'')$ be the attacked targets at WSE for the two defender strategies Φ' and Φ (the candidate solution) resp. Introducing these allows us to write the required defender utility expressions simply as:

$$U^{\text{wse}}(\Phi', \boldsymbol{w}) = U^{\text{d}}(\Phi', i', j') \quad \text{and} \quad U^{\text{wse}}(\Phi, \boldsymbol{w}) = U^{\text{d}}(\Phi, i'', j'').$$

We then iterate over all tuples (i', j', i'', j'') ($O(s^2 t^2)$ many of them) to compute the max regret corresponding to each pair (via OP described momentarily), and the tuple leading to maximum objective gives the solution to (6).

Previous works using a similar approach, such as, [23] assume mixed strategies and compute the SSE. In our model, however, computing WSE presents the challenge of capturing the worst-case tiebreaking, requiring an entirely different formulation. For given pair of targets $(i', j'), (i'', j'')$ as described above and for input strategy Φ, we compute the regret maximizing sample $(\Phi', \boldsymbol{w})$ as follows:

$$\begin{aligned}
\max_{\Phi', \boldsymbol{w}, \beta} \quad & \beta \\
\text{s.t.} \quad & \Phi' \in \mathcal{F},\ \boldsymbol{w} \in \mathcal{W},\ \beta \in \mathbb{R},\ \boldsymbol{q} \in \{0,1\}^{s \times t} \\
& \beta \leq U^{\text{d}}(\Phi, i'', j'') - U^{\text{d}}(\Phi', i', j') \\
& \left.\begin{aligned}
& M q_{ij} \geq U^{\text{d}}(\Phi', i, j) - U^{\text{d}}(\Phi', i', j') \\
& U^{\text{a}}(\Phi', i', j', \boldsymbol{w}) \geq U^{\text{a}}(\Phi', i, j, \boldsymbol{w}) + \epsilon q_{ij} \\
& U^{\text{a}}(\Phi, i'', j'', \boldsymbol{w}) \geq U^{\text{a}}(\Phi, i, j, \boldsymbol{w}).
\end{aligned}\right\} \forall\ i \in \mathcal{S}, j \in \mathcal{T}
\end{aligned}$$

The objective β is the the regret to be maximized, while the remaining constraints ensure that $(i', j'), (i'', j'')$ are indeed the respective attacked targets, as follows. The fourth constraint requires (i'', j'') to be the attacker best-response against Φ, and the worst-case tiebreaking is ensured by the first constraint since maximizing objective β requires maximizing $U^{\text{d}}(\Phi, i'', j'')$. For (i', j') on the other hand, the third constraint ensures that it is a best response to Φ'. Moreover, ϵ is a small positive constant used there which sets $q_{ij} = 0$ for each ϵ-optimal OC j. As explained previously for computing (3), choosing a small enough ϵ sets $q_{ij} = 0$ for precisely every optimal attack j. Consequently, the defender loss for every such (i, j) is more than for (i', j') (by the second constraint, where M is a large positive constant), thus capturing the worst-case tiebreaking.

5 GEBRA: Exploiting Bounded Rationality Model-free

Here, we aim to tackle bounded rationality without any assumptions on the attacker model. One simple approach is to use (3) (where ϵ was set very small for full rationality), and set an appropriate ϵ to reflect the extent of sub-optimality — akin to the COBRA algorithm [26] for SSGs. Another previous approach for SSGs is MATCH [25] which bounds the defender's loss due to attacker's deviation from rationality, by a (pre-set) constant β times the attacker's utility reduction. Thus, it guarantees that if the attacker is *close* to rationality, the defender is *close* to optimal utility. We adapt this principle to propose our solution GEBRA (Guaranteed Exploitation against Boundedly Rational Attackers).

Strictly Competitive Games In the security domain, having attack choices favorable to both the attacker and the defender is rather unlikely. A very practical class of games here is the *Strictly competitive games* [11], where all outcomes are pareto optimal. In particular, if the attacker deviates to lower utility, the defender gets a smaller loss, thus, the attacker playing rationally is the worst case for the defender. Hence, the previous approaches COBRA and MATCH merely reduce to the conservative Minimax solution, rendering them unavailing as the desired robustness is intrinsically present. Hence, we aim to exploit bounded rationality in this setting, by requiring that the defender loss must improve by at least (a factor of) the reduction in the attacker utility, as explained momentarily.

Note that, checking if a game is strictly competitive is challenging due to the compact representation via polytopal strategy spaces in our game. We show an MILP formulation to determine if a game is strictly competitive (see Appendix).

Optimization problem for GEBRA: In the strictly competitive setting, if the attacker deviates from his optimal utility, the defender is guaranteed to get a smaller loss. To have guaranteed exploitation, we require that the decrement in defender's loss, is lower-bounded by β times the decrement in attacker utility, where β is a positive constant. Then, this can be computed by modifying (3) as:

$$\begin{aligned}
\min_{\boldsymbol{\Phi},\boldsymbol{q},\boldsymbol{h},\gamma,\alpha}\ & \gamma && (9)\\
\text{s.t.}\ & \alpha,\gamma\in\mathbb{R},\ \Phi\in\mathcal{F},\ \boldsymbol{q},\boldsymbol{r}\in\{0,1\}^{s\times t} && \\
& q_{11}+\ldots+q_{st}\geq 1 && (9a)\\
& r_{11}+\ldots+r_{st}\geq 1 && (9b)\\
& \left.\begin{array}{ll}
\epsilon(1-q_{ij})\leq\alpha-U^{\mathrm{a}}(\Phi,j,i) & (9c)\\
M(1-q_{ij})\geq\alpha-U^{\mathrm{a}}(\Phi,j,i) & (9d)\\
U^{\mathrm{d}}(\Phi,j,i)\leq\gamma+M(1-q_{ij}) & (9e)\\
\gamma\leq U^{\mathrm{d}}(\Phi,j,i)+M(1-r_{ij}) & (9f)\\
r_{ij}\leq q_{ij}\leq\Phi_{ij} & (9g)\\
M(1-h_{ij})+\gamma-U^{\mathrm{d}}(\Phi,j,i)\geq\beta(\alpha-U^{\mathrm{a}}(\Phi,j,i)) & (9h)\\
h_{ij}\leq\Phi_{ij}\leq Mh_{ij}. & (9i)
\end{array}\right\}\forall i\in\mathcal{S},j\in\mathcal{T}
\end{aligned}$$

Similar to (3a)∼(3e), constraints (9a)∼(9g) enforce α,γ as wse utilities. Here, we have binary variables h_{ij} for any attack (i,j) the attacker can deviate to instead of the best response. Constraint (9i) ensures $h_{ij}=1$ iff Φ_{ij} is nonzero, i.e., (i,j) is a valid attack (for a deviation). The gist of GEBRA is captured by (9h). For any deviation (i,j), the attacker's utility decreases by $\alpha-U^{\mathrm{a}}(\Phi,j,i)$ relative to optimal. The corresponding decrease in defender loss is $\gamma-U^{\mathrm{d}}(\Phi,j,i)$ which we require to be at least β-fold (whenever $h_{ij}=1$, i.e. for every valid deviation). The constant β represents the magnitude of exploitation guarantee.

Remark 4. Setting $\beta=0$ makes the key constraint of GEBRA always true, and the last constraint is redundant (since $\boldsymbol{h}$ is not tied to any other variables). Hence, GEBRA reduces to computing WSE which always exists in this case.

Note that for strictly competitive games (by definition), (9) is guaranteed to have a feasible solution for some β strictly positive. Importantly, however, the converse is not true, and in fact, in our numerical results, we use a class of games that generalizes strictly competitive games, and GEBRA still always finds a feasible solution. Further, setting $\beta < 0$ in (9), we can rearrange and reinterpret it as to require that the increase in defender loss is at most $|\beta|$ times the attacker's utility decrement — same robustness guarantee as MATCH, which we resort to in games where attacker suboptimality can severely increase defender loss.

6 Numerical results

Setup We keep the game parameters small[8] for numerical analysis and it suffices to clearly highlight their efficacy. We use 5 TCs, OCs each and 15 machines. A game instance is created by randomly creating constraints, player valuations for TCs and the assignment of machines to TCs. To compute aggregates or averages across games or attacker populations, we keep the sample size 50 in each case.

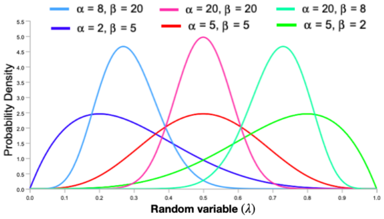


Fig. 1: Distributions $Beta(\alpha, \beta)$

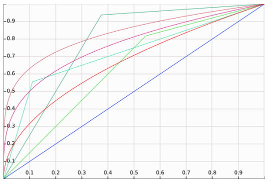


Fig. 2: Two-piecewise Linear (in green) Vs Polynomial (in red) transformations (normalized to be from $[0, 1]$ to $[0, 1]$)

Parametrized Prospect Theoretic Model Here, we compare our Prospect theory based solution (PT) against WSE (i.e., the solution assuming a rational attacker with worst-case tiebreaking). We consider a population of risk-averse attackers governed by a parameter λ drawn from a distribution $Beta(\alpha, \beta)$ (density functions as shown in Figure 1). PT estimates λ by computing the MLE and best-responds to it. We vary the parameters α, β so as to cover a spectrum of the average degree of risk-aversion (captured by distribution mean $\frac{\alpha}{\alpha+\beta}$), and the homogeneity (captured by distribution variance $\approx \frac{\alpha\beta}{(\alpha+\beta)^3}$) of the population.

As shown in Table 1, PT does significantly better for populations with low variance, as compared to high variance. Intuitively, this is because the learned parameter λ can represent the population better when there is more homogeneity (i.e., low variance). Within each of the sub-tables 1c, 1b, 1a, when the degree of risk-aversion is high (i.e., low mean λ; left column), the improvement of WSE over PT is higher, than when the population mean is high (i.e., smaller overall risk-aversion; right column), as expected. At the extreme with small risk-aversion on average and low homogeneity, PT does worse than WSE (Table 1c - column 3). For such cases, and others where the parametrization hypothesis may not be accurate, we show that the model-free algorithms are valuable as shown next.

Prospect-theoretic attackers with arbitrary transformations PT relies on the assumption of polynomial transformations and homogeneous populations, which may not hold. Here, we consider a family of *Two-piecewise linear* (2PL)

[8] Essential for quickly solving many instances (to get averaged numbers). Bigger parameters can be handled when solving a specific instance for real-world deployment.

	Distribution		
	(32,80)	(80,80)	(80,32)
WSE	2.712	2.827	2.941
PT	**2.178**	**2.432**	**2.580**

(a) Low variance

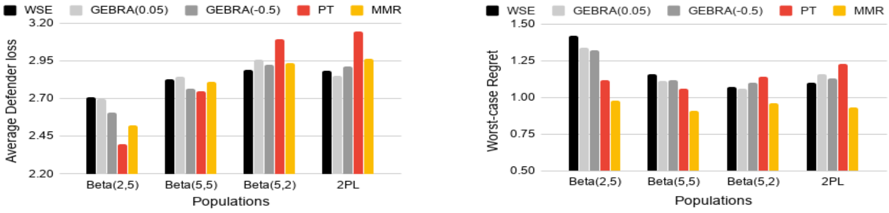
Distribution		
(8,20)	(20,20)	(20,8)
2.724	2.832	2.919
2.272	**2.662**	**2.739**

(b) Medium variance

Distribution		
(2,5)	(5,5)	(5,2)
2.710	2.829	**2.892**
2.396	**2.749**	3.093

(c) High variance

Table 1: Average Defender loss of WSE and PT



(a) Comparing Average Defender Loss

(b) Comparing Worst-case Regret

Fig. 3: Comparing PT, MMR, GEBRA and WSE for prospect theoretic attackers

payoff transformations shown in Figure 2 in contrast with the polynomial transformations that PT hypothesizes for parametrization. We compare the average defender loss of PT, MMR and GEBRA (with overall best parameters $\beta = 0.05$ among positive, and $\beta = -0.5$ among negative), against attacker populations with 2PL transformations, and polynomial transformations with high variance.

Figure 3a shows that against $Beta(5, 2)$ and $2PL$, PT has a much higher loss than WSE which is greatly mitigated with MMR and GEBRA. For $Beta(2, 5)$ and $Beta(5, 5)$, PT has a smaller loss than WSE as seen before, and so do MMR and GEBRA(-0.5), even though the reduction margin is lower, while GEBRA(0.05) does not show much difference. In conclusion, in populations with high risk-aversion and parametrized populations, PT has an edge, however, in other cases where PT suffers, MMR and GEBRA perform much better. To compare the robustness quality, we compare the worst-case regret. Figure 3b shows that the worst-case regret is reduced with MMR compared to WSE and PT in all 4 cases, by up to 40%, 30% respectively, while GEBRA has a worst-case regret a little lower for $Beta(2, 5)$ and not much different than WSE in other cases.

Exploiting bounded rationality with GEBRA We want to consider the aforementioned class of strictly competitive games, however, checking this property is non-trivial (requiring to solve an MILP for each game, rather than defined via closed-form constraints). Hence, we consider a slightly more general class of games with *strictly conflicting valuations* - for TCs i and j, $u_i \geq u_j \iff v_i \geq v_j$, i.e. if the attacker gets a higher reward from a TC than the other, the defender suffers a higher loss and vice versa. Unsurprisingly, even for this class, MATCH (i.e., GEBRA with $\beta < 0$) and COBRA achieve an output that differs little from that of WSE. Hence, we only compare GEBRA (with $\beta > 0$) against WSE here.

Having studied risk-averse attackers, we consider a different form of bounded rationality as given by the *Quantal Response* (QR) model — an attacker with a QR parameter ϵ chooses an attack having utility u, with a probability $\propto \exp(\epsilon u)$. Thus, $\epsilon \to \infty$ for a perfectly rational attacker, while $\epsilon = 0$ for a fully random attacker. We consider populations of attackers with varying distributions of ϵ,

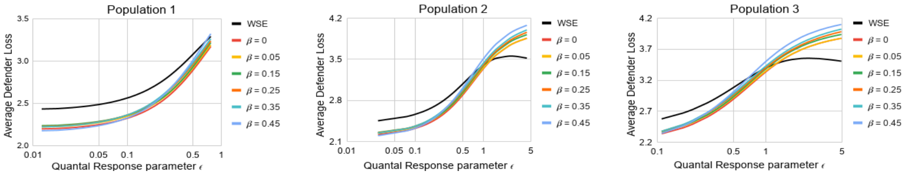


Fig. 4: Average Defender Loss comparison between GEBRA and WSE

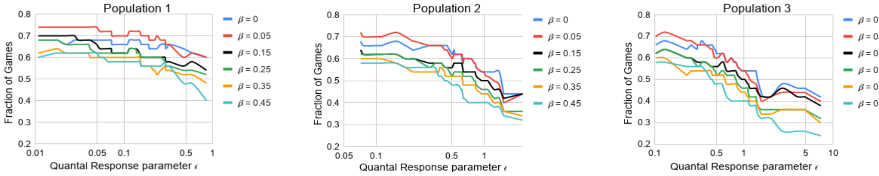


Fig. 5: Fraction of games where GEBRA does at least as good as WSE

namely $LN(-2,1)$, $LN(-1,1)$, $LN(0,1)$ where $LN(\alpha,\beta)$ denotes a LogNormal Distribution with parameters (α,β). These three distributions have an increasing order of means and thus, increasing average degree of rationality.

Figure 4 shows the performance of GEBRA for various settings of β (for illustration, we only show a range for β with sizeable loss reduction). For $LN(-2,1)$ with least average rationality, GEBRA reduces the absolute loss by about 10%. However, the loss gets higher by 10% than WSE for attackers nearly rational.

We also measure the fraction of games in which GEBRA surpasses WSE, shown in Figure 5. With $\beta = 0.05$, it does at least as good as WSE in 75% games for attackers nearly random and over 60% for the ones more rational. As degree of rationality rises, however, this percentage drops in other two populations.

7 Summary

In this paper, we present Risk-based Cyber Camouflage Games (RCCG) to capture the crucial uncertainty in the attack success. First, for rational attackers, we show NP-hardness of equilibrium computation, a pseudo-polynomial time algorithm for the special *unconstrained* setting, and an MILP formulation for the general *constrained* problem. Further, to tackle risk-averse attackers, we propose a Prospect theory based approach (PT) that estimates the attacker behavior from data and a variant that is robust against arbitrary payoff transformations based on Min-Max Regret (MMR). Finally, we also propose a model-free approach (GEBRA) that can exploit arbitrary deviations from rationality.

Our numerical results show that PT shows significant improvement for homogeneous populations and for a high risk-aversion, however, for heterogeneous populations, MMR moderately improves the defender loss while also achieving much lower regret. Finally, GEBRA is valuable in the *Strictly Competitive* [11] setting where previous model-free approaches for handling bounded rationality prove ineffective, particularly for attackers with a high deviation from rationality.

8 Acknowledgements

This work is sponsored by the Army Research Office (grant W911NF-17-1-0370).

Bibliography

[1] P. Aggarwal, O. Thakoor, A. Mate, M. Tambe, E. A. Cranford, C. Lebiere, and C. Gonzalez. An exploratory study of a masking strategy of cyberdeception using cybervan. In *HFES*, 2020.

[2] T. Alpcan and T. Başar. *Network security: A decision and game-theoretic approach.* 2010.

[3] A. Ben-Tal, L. El Ghaoui, and A. Nemirovski. *Robust optimization.* 2009.

[4] D. Berrueta. A practical approach for defeating nmap os- fingerprinting. 2003.

[5] C. Boutilier, R. Patrascu, P. Poupart, and D. Schuurmans. Constraint-based optimization and utility elicitation using the minimax decision criterion. *Artificial Intelligence*, 170(8-9):686–713, 2006.

[6] M. Breton, A. Alj, and A. Haurie. Sequential stackelberg equilibria in two-person games. *Journal of Optimization Theory and Applications*, Oct 1988.

[7] R. Chadha, T. Bowen, C. J. Chiang, Y. M. Gottlieb, A. Poylisher, A. Sapello, C. Serban, S. Sugrim, G. Walther, L. M. Marvel, E. A. Newcomb, and J. Santos. Cybervan: A cyber security virtual assured network testbed. In *MILCOM 2016 - 2016 IEEE Military Communications Conference*, Nov 2016. https://doi.org/10.1109/MILCOM.2016.7795481.

[8] S. Cooney, K. Wang, E. Bondi, T. Nguyen, P. Vayanos, and et al. Learning to signal in the goldilocks zone: Improving adversary compliance in security games. In *ECML/PKDD*, 2019.

[9] D. P. de Farias and B. Van Roy. On constraint sampling in the linear programming approach to approximate linear programming. In *CDC*, 2003.

[10] F. De Gaspari, S. Jajodia, L. V. Mancini, and A. Panico. Ahead: A new architecture for active defense. In *SafeConfig*, 2016.

[11] J Eatwell, M Milgate, and P Newman. The new palgrave: a dictionary of economics. 1987.

[12] K. Ferguson-Walter, D. LaFon, and T. Shade. Friend or faux: Deception for cyber defense. *Journal of Information Warfare*, 2017.

[13] V. Goel and N. Perlroth. *Yahoo Says 1 Billion User Accounts Were Hacked*, December 2016. https://www.nytimes.com/2016/12/14/technology/yahoo-hack.html.

[14] Q. Guo, J. Gan, F. Fang, L. Tran-Thanh, M. Tambe, and B. An. On the inducibility of stackelberg equilibrium for security games. *CoRR*, abs/1811.03823, 2018.

[15] I Gutzmer. *Equifax Announces Cybersecurity Incident Involving Consumer Information*, 2017. https://investor.equifax.com/news-and-events/news/2017/09-07-2017-213000628.

[16] A. Xin Jiang, H. Chan, and K. Leyton-Brown. Resource graph games: A compact representation for games with structured strategy spaces. In *AAAI*, 2017.

[17] R. Joyce. Disrupting nation state hackers. San Francisco, CA, 2016. USENIX Association.

[18] C. Kiekintveld, J. Marecki, and M. Tambe. Approximation methods for infinite bayesian stackelberg games: Modeling distributional payoff uncertainty. In *AAMAS*, 2011.

[19] C. Kiekintveld, T. Islam, and V. Kreinovich. Security games with interval uncertainty. In *AAMAS*, 2013.

[20] A. Laszka, Y. Vorobeychik, and X. D. Koutsoukos. Optimal personalized filtering against spear-phishing attacks. In *AAAI*, 2015.

[21] Mandiant. Apt1: Exposing one of china's cyber espionage units, 2013.

[22] R. McKelvey and T. Palfrey. Quantal response equilibria for normal form games. *Games and economic behavior*, 10(1):6–38, 1995.

[23] T. H. Nguyen, A. Yadav, B. An, M. Tambe, and C. Boutilier. Regret-based optimization and preference elicitation for stackelberg security games with uncertainty. In *AAAI*, 2014.

[24] A. Peterson. *OPM says 5.6 million fingerprints stolen in cyber-attack, five times as many as previously thought*, September 2015. https://www.washingtonpost.com/news/the-switch/wp/2015/09/23/opm-now-says-more-than-five-million-fingerprints -compromised-in-breaches.

[25] J. Pita, R. John, R. Maheswaran, M. Tambe, and S. Kraus. A robust approach to addressing human adversaries in security games. In *ECAI*, page 660–665, 2012.

[26] J. Pita, R. John, R. Maheswaran, M. Tambe, Rong Yang, and Sarit Kraus. A robust approach to addressing human adversaries in security games. In *AAMAS*, pages 1297–1298, 2012.

[27] Y. Qian, W. Haskell, and M. Tambe. Robust strategy against unknown risk-averse attackers in security games. In *AAMAS*, 2015.

[28] M. Rahman, M. Manshaei, and E. Al-Shaer. A game-theoretic approach for deceiving remote operating system fingerprinting. *CNS*, pages 73–81, 2013.

[29] A. Schlenker, H. Xu, M. Guirguis, C. Kiekintveld, A. Sinha, M. Tambe, S. Sonya, D. Balderas, and N. Dunstatter. Don‘t bury your head in warnings: A game-theoretic approach for intelligent allocation of cyber-security alerts. 2017.

[30] A. Schlenker, O. Thakoor, H. Xu, F. Fang, M. Tambe, L. Tran-Thanh, P. Vayanos, and Y. Vorobeychik. Deceiving cyber adversaries: A game theoretic approach. In *AAMAS*, 2018.

[31] E. Serra, S. Jajodia, A. Pugliese, A. Rullo, and VS Subrahmanian. Pareto-optimal adversarial defense of enterprise systems. *ACM Transactions on Information and System Security (TISSEC)*, 17(3):11, 2015.

[32] A. Sinha, P. Malo, and K. Deb. A review on bilevel optimization: From classical to evolutionary approaches and applications. *IEEE Transactions on Evolutionary Computation*, 22(2):276–295, 2018.

[33] M. Tambe. *Security and game theory: algorithms, deployed systems, lessons learned.* 2011.

[34] O. Thakoor, M. Tambe, P. Vayanos, H. Xu, C. Kiekintveld, and F. Fang. Cyber camouflage games for strategic deception. In *GameSec*, 2019.

[35] Thinkst. *Canary*, 2015. https://canary.tools/.

[36] A. Tversky and D. Kahneman. Prospect theory: An analysis of decision under risk. *Econometrica*, 47(2):263–291, 1979.

[37] B. von Stengel and S. Zamir. Leadership with commitment to mixed strategies. Technical report, 2004.

[38] R. Yang, C. Kiekintveld, F. Ordonez, M. Tambe, and R. John. Improving resource allocation strategy against human adversaries in security games. In *ICJAI*, 2011.

A RCCG for rational attackers

Lemma 1. *Under a given defender strategy $\varPhi$, let j_1, j_2 be OCs which are masking subsets of machines $\mathcal{K}_1$ and $\mathcal{K}_2$ respectively. Let $\varPhi'$ be constructed from $\varPhi$ by merging the machines in $\mathcal{K}_1$ and $\mathcal{K}_2$ and masking with a single OC, say j'. Then, $U^{\mathrm{a}}(\varPhi', i, j') \leq \max(U^{\mathrm{a}}(\varPhi, i, j_1), U^{\mathrm{a}}(\varPhi, i, j_2)) \;\forall\; i \in \mathcal{S}$.*

Proof. For an arbitrary TC i, for brevity, let's denote $a = \varPhi_{i,j_1}$, $b = \varPhi_{i,j_2}$. WLOG, let $a/|\mathcal{K}_1| \leq b/|\mathcal{K}_2|$ (these are the probabilities that the attacks on (i, j_1), (i, j_2) are successful, resp.). Then,

$$U^{\mathrm{a}}(\varPhi', i, j') = \frac{(a+b)}{|\mathcal{K}_1| + |\mathcal{K}_2|} v_i \leq \frac{b}{|\mathcal{K}_2|} v_i = U^{\mathrm{a}}(\varPhi, i, j_2). \qquad \square$$

This shows that merging the machines in any two OCs under one, cannot increase the attacker utility for any target, which prompts the following results.

Proposition 1. *Unconstrained zero-sum RCCG always has an optimal strategy that uses just one OC, thus computable in $O(1)$ time.*

Proof. Consider an optimal strategy $\varPhi$ that uses two or more OCs, with i^*, j^* being the attacker best response. In the unconstrained setting, OCs can be freely merged. Say we merge machines from OC $\hat{j}$ to j^* to obtain $\varPhi'$. By Lemma 1, the attacker utility from any (i, j^*) under $\varPhi'$ is at most the utility from (i, j^*) or $(i, \hat{j})$ under $\varPhi$, and thus, at most the best response attacker utility against $\varPhi$. As the remaining attack options have an unchanged utility, it follows that the best response attacker utility against $\varPhi'$ is at most that against $\varPhi$. Since the game is zero-sum, the same applies for the defender loss, making $\varPhi'$ also optimal while it uses fewer OCs. It follows via inductive reasoning that there exists an optimal strategy which uses a single OC to mask all the machines. □

Theorem 1. *Zero-sum RCCG is NP-hard.*

Proof. We reduce from the problem "Exact Cover by 3-Sets" ($ExC3$ for brevity) which is NP-complete. In this problem, we are given a set X, with $|X| = 3q$ (so, the size of X is a multiple of 3), and a collection C of 3-element subsets of X. The decision problem is whether $\exists C' \subset C$ where every element of X occurs in exactly one member of C'. Given such an instance, construct an RCCG instance as follows. Construct TCs $1, \ldots, 3q$ corresponding to elements of X. Let the value of each TC be 0 and let there be exactly one machine of each. Let there be TC $3q + 1$ of value $V > 0$ and q machines of it. Let there be $|C|$ OCs corresponding the subsets in C. Suppose OC corresponding to any $S \in C$ can mask exactly the 3 TCs in S & TC $3q + 1$. Let all costs be 0. This is a poly-time reduction by construction. We claim that an $ExC3$ instance is YES iff the minimum defender loss in the constructed RCCG is exactly $V/4$.

Consider a strategy Φ. Let $J' \subseteq J$ be the OCs which mask at least one machine of some TC $i \in \{1, \ldots, 3q\}$. By construction, J' must have q OCs. Further, machines of TC $3q + 1$ must be masked by OCs in j' to minimize the defender loss since otherwise the defender loss is V. Now, for an OC j that masks a machine of TC $3q+1$, it must mask only one to minimize the defender loss. For each such OC j, if it masks $x_j (\leq 3)$ machines from TCs $1, \ldots, 3q$, we can write $U^{\mathrm{d}}(\Phi, 3q+1, j) = \frac{1}{1+x_j} V \geq V/4$ which attains the minimum of $V/4$ when $x_j = 3$. Since the attacker chooses to attack (i, j) which maximizes it , the defender loss is lower bounded by $v/4$. Now, if the given instance of $ExC3$ is a YES instance, it is possible to find q OCs which cover all the TCs, and use them to mask the 3 machines of the corresponding TCs along with one machine of TC $3q + 1$ each, thus achieving the minimum loss of $V/4$. Conversely, if the minimum defender loss is $V/4$, the defender loss when attacked at any OC and TC $3q + 1$ (if so valid) must be at most $V/4$, which implies that it must contain only 1 machine of $3q+1$, and thus i) there should be q such OCs used, and ii) each of them must have at least 3 machines of TCs $1, \ldots, 3q$. So, there must be exactly q such OCs each with exactly 3 machines. Hence, the subsets corresponding to these OCs form the exact cover of the given $ExC3$ making it a YES instance. □

Proposition 2. *Unconstrained RCCG always has an optimal strategy that uses just two OCs.*

Proof. Consider an optimal strategy Φ that uses three or more OCs, with i^*, j^* being the attacker best response. In the unconstrained setting, OCs can be freely merged. Say we merge machines from OC j_1 to j_2 (with j_1, $j_2 \neq j^*$) to obtain Φ'. By Lemma 1, the attacker utility of any (i, j_2) under Φ' is at most the utility of (i, j_1) or (i, j_2) under Φ, and thus, i^*, j^* must still be the best response for the attacker against Φ. In particular, this also ensures that it remains the worst-case for the defender in case of tie-breaks. It follows via inductive reasoning that given an optimal strategy with two or more OCs, another using fewer OCs can be constructed. Thus, there exists an optimal strategy which uses a single OC. □

SOBRE Algorithm

SOBRE uses the subroutine DPBRF (Dynamic Programming for Best Response Feasibility) which given the input (i, n^*, m^*), computes if the machines can be masked so that OC 1 has m^* total machines with n^* of TC i^*, and $(i^*, 1)$ is the attacker best response. Function $f(i, m)$, (memoized: Line 2), computes if such a strategy exists with additional property that TCs $1, \ldots, i$ in total have m machines in $OC1$. To compute $f(i, m)$, we consider n out of n_i machines of TC $i (\neq i^*)$ to be put in OC 1 (Line 7). If doing so keeps $(i^*, 1)$ at a higher utility than $(i,1),(i,2)$ (Line 8), and similarly recursively for all smaller-indexed TCs (Line 9), $f(i,m)$ is true. Lines 5,6 mark the base cases. DPBRF returns true if $f(s,m^*)$ is true (Line 3) by definition.

$f(i,m)$ is computable in $O(n_i)$ (Line 7), hence, DPBRF takes $O(km^*)$ using $\sum_{i=1}^{s} n_i = k$. Summing over the loops of SOBRE gives its runtime as $O(k^4)$.

Algorithm 3: DPBRF(i^*, n^*, m^*)

```
1  for $i = 1,\ldots,s; m = 0,\ldots,m^*$
2      $A[i,m] \leftarrow f(i,m)$
3  Return $A[s,m^*]$
4  Function $f(i,m)$
5      if $(i = 0)$ Return $m = 0$
6      if $(i = i^*)$ Return $A[i-1, m-n^*]$
7      for $n = 0,\ldots,n_i$
8          if $(\max\{\frac{n}{m^*}v_i, \frac{n_i-n}{k-m^*}v_i\} < \frac{n^*}{m^*}v_{i^*})$ // '<=' if lower defender loss
9              if $(A[i-1, m-n])$ Return $true$
10     Return $false$
```

B Sensitivity to learning error.

Suppose the estimated parameter is λ^* and the computed optimal solution is Φ, yielding a defender utility u^*. We want to provide an error interval around λ^* s.t. the defender loss does not increase (at all, or beyond a desired margin ϵ), if the true λ is within this interval. Equivalently, we compute the least perturbation needed s.t. the defender loss increases. We consider all pairs (i,j) s.t. $U^{\mathrm{d}}(\Phi,i,j) > u^* + \epsilon$. Thus, if the attacker best response is any such (i,j), then the defender loss increases beyond the desired threshold. We compute the minimum deviation (of true λ from estimated λ^*) that causes this (if it exists) by solving

$$\min_{\lambda} \quad |\lambda - \lambda^*| \quad \text{s.t.} \;\; \log f_\lambda(i,j) \geq \log f_\lambda(i',j') \;\; \forall\, i' \in \mathcal{S} \;\forall\, j' \in \mathcal{T} \tag{10}$$

The constraint here ensures that (i,j) is indeed the prospect-maximizing response, where we use log on both sides to get an LP, for efficient computation. Then, solving (10) for all (i,j) pairs for which $U^{\mathrm{d}}(\Phi,i,j) > u^* + \epsilon$, and taking the minimum of all the perturbations, gives us the required tolerance.

C Computing Strict Competitiveness

We formulate an MILP that is feasible iff for a strategy Φ, deviating from some (i,j) to (i',j') is beneficial to both players — game is not strictly competitive.

$$\left.\begin{aligned} M(1-q_{ij}) + U^{\mathrm{a}}(\Phi,i,j) > \alpha, &\quad M(1-r_{ij}) + \alpha > U^{\mathrm{a}}(\Phi,i,j) \\ M(1-r_{ij}) + U^{\mathrm{d}}(\Phi,i,j) > \beta, &\quad M(1-q_{ij}) + \beta > U^{\mathrm{d}}(\Phi,i,j) \\ r_{ij} \leq \Phi_{ij} \leq M r_{ij}, &\quad q_{ij} \leq \Phi_{ij} \leq M q_{ij} \end{aligned}\right\} \forall\, i \in \mathcal{S}, j \in \mathcal{T}$$

$$\Phi \in \mathcal{F}, \;\; \boldsymbol{q}, \boldsymbol{r} \in \{0,1\}^{s \times t}, \quad q_{11} + \ldots + q_{st} = 1, \quad r_{11} + \ldots + r_{st} = 1$$

Here, binary variables q, r capture (i,j) and (i',j') respectively which define the aforementioned attacker deviation. Line 4 ensures they are unique and Line 3 ensures they are valid attacks. Attacker and defender both prefer (i,j) over (i',j') as per Lines 1,2 respectively.