Bayesian Repeated Zero-Sum Games with Persistent State, with Application to Security Games

Vincent Conitzer[1], Yuan Deng[1](✉), and Shaddin Dughmi[2]

[1] Duke University, Durham, USA
{conitzer,ericdy}@cs.duke.edu
[2] University of Southern California, Los Angeles, USA
shaddin@usc.edu

Abstract. We study infinitely-repeated two-player zero-sum games with one-sided private information and a persistent state. Here, only one of the two players learns the state of the repeated game. We consider two models: either the state is chosen by nature, or by one of the players. For the former, the equilibrium of the repeated game is known to be equivalent to that of a one-shot public signaling game, and we make this equivalence algorithmic. For the latter, we show equivalence to one-shot team max-min games, and also provide an algorithmic reduction. We apply this framework to repeated zero-sum security games with private information on the side of the defender and provide an almost complete characterization of their computational complexity.

Keywords: Bayesian repeated game · Equilibrium characterization · Equilibrium computation · Computational complexity

1 Introduction

Private information can give one a strategic advantage over other players in a game. However, if play is repeated, then taking advantage of one's private information through one's actions risks leaking that information and thereby the advantage. This is nicely illustrated in the movie *The Imitation Game*, in which British intelligence, having cracked the *Enigma* code, strategically decides not to act on some of its information, in order to preserve its informational advantage [12]. Less dramatically, consider a buyer and a seller that interact repeatedly. The seller has a higher-quality and a lower-quality version of the item for sale, and offers these at different prices. The buyer may, at the current prices, prefer the higher-quality version – but worry that choosing this option will reveal her (persistently) high valuation/type, causing the seller to raise prices in the future, and therefore choose the cheaper low-quality version instead.

V. Conitzer—Supported by NSF Award IIS-1814056.
S. Dughmi—Supported by NSF CAREER Award CCF-1350900.

X. Chen et al. (Eds.): WINE 2020, LNCS 12495, pp. 444–458, 2020.
https://doi.org/10.1007/978-3-030-64946-3_31

In equilibrium, to what extent should a party with an informational advantage refrain from acting on this information? This is the question we set out to address in this paper. It is, in its most general form, a challenging question to answer. The state of the game may change over time; there may be a multiplicity of equilibria; the discount factor matters; and so on. Thus, answering the question in general would require us to simultaneously resolve a number of fundamental questions in (algorithmic) game theory. In this paper, in order to stay focused on the question at hand, we focus on the following special case:

- The state of the game is *persistent*, i.e., it does not change over time (the game is repeated rather than stochastic).
- Only one player has private information, and it does not change.
- The game is two-player and zero-sum.
- Each agent cares about their long-term average payoff.

Even in this setting, it is easy to see that the optimal answer is in general not one of the two extremes – either exploit information fully, or never use it. Some information may not be actionable for the adversary so that one can simply take advantage of it and not worry about revealing it. On the other hand, for other information, it is possible that the adversary would be able to make even better use of it than the initially better-informed player. In that case, the benefits of getting to use the information for one round, without the adversary being able to use it in that particular round, will be completely wiped out by the infinitely many remaining rounds in which the adversary can use the information better.

The technical and conceptual foundations for the study of repeated games of incomplete information with persistent state were laid by [2]. They consider a persistent state of the game drawn by nature from a common prior, and agents who receive private signals regarding this state. [14] provides an in-depth accounting of the special case of this model with two players and zero-sum payoffs. The aforementioned texts reveal that the even-more-special case we consider, that of repeated two-player zero-sum games with one-sided private information, admits an essentially-unique equilibrium (in the sense of payoff equivalence) with an elegant, simple, and instructive characterization which is robust to modeling assumptions. In particular, the equilibrium of the repeated two-player game is equivalent, in a precise technical sense, to the equilibrium of a one-shot *public signaling game* with three players. Moreover, this characterization is robust to how one chooses to model long-term payoffs; say through using a discount factor, taking the limit of the finite repeated game as the number of stages grows to infinity, or considering the infinite game directly. Even mild generalizations of this special case, for example to more players, non-zero-sum payoffs, or incomplete information on both sides, lead to the collapse of this characterization, and such settings are not yet fully understood to the best of our knowledge. This further cements our model as the timely choice for algorithmic study.

1.1 Our Contributions

We examine repeated two-player zero-sum games with one-sided private information from the perspective of algorithmic game theory, both in general and as exemplified by application to the influential domain of security games [16]. We consider both the case when the state is drawn by nature—this is the classical model in [2,14]—as well as a natural, and to our knowledge novel, variant in which the (typically randomized) state is chosen by one of the players, who is therefore the informed party. We refer to this variant as the *allocation* model.

The domain-agnostic part of the paper is organized as follows. For the classical model, where the game state is drawn by nature, we first provide (a) our own exposition of the previously-described equilibrium characterization in terms of one-shot public signaling games, one that is particularly tailored to an algorithmic game theory audience and makes explicit the connection to recent work on public signaling games (e.g., [6–8]). Then, we turn to our novel contributions. We provide (b) an efficient reduction to equilibrium computation in the related one-shot public signaling game to make the equilibrium characterization constructive. For the allocation model, where one of the players determines the (persistent) state, we provide (a') a characterization of the equilibrium of the repeated game as equivalent, in a precise technical sense, to the equilibrium of a one-shot three-player *team max-min game*, as first studied by [15]; (b') an efficient reduction to computing the equilibrium of the associated team max-min game. We note that, in both (b) and (b'), the uninformed player's strategy is particularly nontrivial, and involves efficiently solving a related instance of *Blackwell's approachability* [1,4]. We also note that the reductions in (b) and (b') are "reversible", since both the repeated game and the associated one-shot game share the same game value. Finally, we (c) show that the allocation model is computationally easier than the classical model by way of a polynomial time reduction. We note that this is not reversible, and the complexity relationship is strict, as evidenced by our results for security games which we describe next.

We then examine repeated zero-sum security games with private information on the side of the defender. In the security games we consider, the state is a deployment of "treasures" to "locations", a defender strategy is a deployment of "defensive resources" to the locations, and the attacker's strategy is a location to attack. Such security games are particularly versatile exemplars for both the classical and allocation models of repeated games with persistent state. The classical model abstracts challenges faced in recent applications to environmental protection [9,17,18], where the locations of environmental assets (the treasures) are determined by nature and slow to change over time. The allocation model can be applied to armed conflict scenarios in which supply-chain assets (the treasures) must be deployed covertly to locations early on in the conflict, and can not be easily moved from stage to stage. We show that the classical model of repeated security games is strongly NP-hard even when treasures, locations, and defensive resources are homogeneous. A more nuanced picture emerges for the allocation model of repeated security games: the fully homogeneous case is tractable, as is the case where only the treasures are heterogeneous. The fully heterogeneous case is strongly NP-hard. Remaining cases are either weakly or strongly NP-hard, and we provide an almost complete accounting of the computational complexity of all combinations.

2 Preliminaries

2.1 One-Shot Games

A one-shot two-player zero-sum game of complete information is described by a utility function $\mathcal{U} : S_1 \times S_2 \to \mathbb{R}$, where S_i is the family of *pure strategies* for player i, and $\mathcal{U}(s_1, s_2)$ is the utility of player 1 when player 1 plays $s_1 \in S_1$ and player 2 plays $s_2 \in S_2$. Implicitly, the utility of player 2 is $-\mathcal{U}(s_1, s_2)$. A *mixed strategy* for player i is $s_i \in \Delta(S_i)$, where $\Delta(S_i)$ is the set of distributions over S_i. A one-shot two-player Bayesian zero-sum game with incomplete information on one side $\left(\Pi, \{\mathcal{U}^\theta\}_{\theta\in\Theta}\right)$ is given by: (1) pure strategy sets S_1 and S_2 for players 1 and 2 respectively; (2) a family Θ of *states of nature*; (3) for each state $\theta \in \Theta$, a one-shot two-player zero-sum game of complete information $\mathcal{U}^\theta$; and (4) a *prior distribution* Π over states of nature Θ.

In such a game, nature draws θ from Θ according to the prior Π and then player 1 learns the state θ while player 2 is uninformed about the state. Both players simultaneously choose their strategies s_i (while s_1 can depend on θ but s_2 cannot), which results in a utility of $\mathcal{U}^\theta(s_1, s_2)$ to player 1 and $-\mathcal{U}^\theta(s_1, s_2)$ to player 2. Moreover, given a distribution Π over Θ, we denote by $\mathcal{U}^\Pi$ the game induced by Π such that player 1's payoff is $\mathcal{U}^\Pi(s_1, s_2) = \sum_{\theta\in\Theta} \Pi(\theta) \cdot \mathcal{U}^\theta(s_1, s_2)$. We restrict attention to games where Θ, S_1, S_2 are finite, or at least compact. All *mixed* Nash equilibria of such a game are payoff equivalent to the Nash equilibrium in which each player employs their maximin mixed strategy [11].

2.2 Bayesian Repeated Games

We now describe the classical model of Bayesian repeated games that we consider, henceforth just *Bayesian repeated games* for convenience. Here, a Bayesian zero-sum game is repeated infinitely many times, with incomplete information on one side. We call the one-shot game the *stage game*, and refer to each iteration as a *stage*. We replicate the standard assumptions made by [2,14], as follows. We assume that the state of nature is *persistent*: it does not change from stage to stage.[1] Moreover, we assume that players observe each others' pure strategies after each stage, but do not observe the payoffs directly. This assumption is necessary for the model to be interesting: If players can observe the payoffs directly, then the uncertainty in the game is superfluous, as players can eventually reconstruct relevant entries of the game matrix and the state of nature. Obscuring

[1] If the state of nature is drawn afresh at each stage, then repetition is superfluous for a zero-sum game: the folk theorem and minimax theorem imply that repeating the minimax equilibrium at each stage is the essentially unique equilibrium of the repeated game (up to payoff equivalence).

payoffs in this manner can be viewed as abstracting a situation where payoffs are delayed till the end of the (long, many stage) game. Formally, given a two-player Bayesian zero-sum stage game $G_{\texttt{repeated}} = \left(\Pi, \left\{\mathcal{U}^\theta\right\}_{\theta\in\Theta}\right)$ as described above, the Bayesian repeated game proceeds as follows:

1. θ is drawn by nature from Π and player 1 learns θ while player 2 does not;
2. The stage game $\mathcal{U}^\theta$ is repeated infinitely many times. After each stage, each player observes the pure strategy played by the other player, but does not directly observe the utility gained.

A *history* of play with T stages $H_T = \left((s_1^1, s_2^1), (s_2^2, s_2^2), \ldots, (s_1^T, s_2^T)\right)$ is a finite sequence, where s_i^t is player i's pure strategy at stage t. For convenience, we will use the vectorized form without superscript $\boldsymbol{s}_i = (s_i^1, \cdots, s_i^T)$ to represent the strategy of player i. A pure strategy for player 1 in the repeated game is a function which maps the state θ and an observed history H to player 1's strategy in the next stage of the repeated game, while a pure strategy for player 2 simply maps the observed history H to player 2's strategy in the next stage. A mixed strategy is naturally a distribution over such functions.

2.3 Bayesian Allocation Games

In addition to classical Bayesian repeated games, we introduce a novel variant, the *Bayesian allocation game*, in which the distribution Π of the states is determined by player 1 instead of the nature. Formally, given one-shot games $G_{\texttt{alloc}} = \left(\left\{\mathcal{U}^\theta\right\}_{\theta\in\Theta}\right)$, the Bayesian allocation game proceeds as follows:

1. Player 1 selects a prior Π over Θ that player 2 cannot observe;
2. θ is drawn by nature from Π and player 1 learns θ while player 2 does not;
3. The stage game $\mathcal{U}^\theta$ is repeated infinitely many times. After each stage, each player observes the pure strategy played by the other player, but does not directly observe the utility gained.

In the Bayesian allocation game, in addition to choosing the actions to play at each stage, player 1's strategy also includes a choice of the prior $\Pi \in \Delta(\Theta)$.

2.4 Utility and Equilibrium Model

We consider the utility/equilibrium models deduced from the infinitely-repeated game perspective for agents that are interested in their long-term payoffs. Each player's expected utility is the limit, as $T \to \infty$, of his average expected utility over the first T stages alone. Though this limit may not exist in general, we can nevertheless define a value and equilibrium as in [2,14]. The max-min value of the game is the supremum over all player 1's mixed strategies, of the infimum over player 2's mixed strategies, of the limit infimum as $T \to \infty$ of player 1's average expected utility. Player 1's max-min strategy is that attaining this supremum. We can similarly define the min-max value of the game and Player 2's min-max strategy. When both the max-min and min-max values are equal we refer to them as the value of the game, and the corresponding max-min and min-max strategies form the equilibrium. For a Bayesian repeated game $G_{\texttt{repeated}}$ and a Bayesian allocation game $G_{\texttt{alloc}}$, we denote their game value by $\nu_{\texttt{repeated}}(G_{\texttt{repeated}})$ and $\nu_{\texttt{alloc}}(G_{\texttt{alloc}})$, respectively. Several other natural utility/equilibrium models are equivalent to this one, and we defer the detailed discussions to the full version.

Example 1. Consider a zero-sum security game with 3 identical locations (denoted by ℓ_A, ℓ_B, ℓ_C) and 2 identical treasures, in which the defender can defend 1 location. The defender determines how to allocate the treasures to the locations (once) and how to defend them (every round). The attacker earns one unit of payoff if she attacks an undefended location with a treasure, and zero otherwise. For comparison, in the one-shot Bayesian allocation game (i.e., if there is only a single round), it is straightforward to verify that the optimal strategy for the defender is to allocate two treasures uniformly at random, and for each realization, defend each of the two locations with a treasure with probability $\frac{1}{2}$, leading to an expected payoff $\frac{1}{3}$ for the attacker. However, it turns out that in the infinitely-repeated version, an optimal strategy (unique up to symmetries) to allocate the treasures for the defender is as follows:

– Allocate a treasure to ℓ_A with probability 1;
– Allocate the remaining treasure to ℓ_B with probability $\alpha = \frac{\sqrt{5}-1}{2} \approx 0.618$ and to ℓ_C with probability $1 - \alpha = \frac{3-\sqrt{5}}{2} \approx 0.382$.

In each stage of the repeated game, the defender defends ℓ_A with probability α (so that the attacker's utility of attacking this location is $1 - \alpha$), and defends ℓ_B with probability $1 - \alpha$ (so that the attacker's utility of attacking this location is $\alpha^2 = 1 - \alpha$). The defender never defends ℓ_C (so that the attacker's utility for attacking this target is also $1 - \alpha$).

The above example illustrates a fundamental difference between a one-shot Bayesian allocation game and its infinitely-repeated counterpart. In the one-shot version, the optimal strategy for the defender correlates the allocation and the defensive strategy, and thus, the game is reduced to a two-player zero-sum normal-form game so that the minimax theorem can be applied. However, in the infinitely-repeated version, we will show that in the equilibrium, the allocation of treasures and the defensive strategy are *independent*, as in the example above. In other words, there exists no benefit for the defender to correlate the allocation and the defensive strategy in the infinitely-repeated Bayesian allocation game. Note that the attacker's payoff is larger in the infinitely-repeated version as $1 - \alpha = \frac{3-\sqrt{5}}{2} > \frac{1}{3}$. Intuitively, this is because the attacker can observe the defender's historical defensive actions in the infinitely-repeated game. This is disadvantageous for the defender: either the defensive actions over time give away where the treasures are, or these actions have to be chosen in such a way that they do not, which is a costly constraint. We also emphasize that the game value is an irrational number, demonstrating that the infinitely-repeated Bayesian allocation game cannot be solved by a linear program.

3 Reductions from Repeated Games to One-Shot Games

In this section, we discuss the relationship between one-shot games and both our models of infinitely repeated games, so that one can solve the infinitely repeated game by first solving the corresponding one-shot game. The equivalence between classical Bayesian repeated games and public signaling games has already been shown by [2] and [14]; for completeness, we will fully elaborate on this equivalence first in Sect. 3.1. This will set the stage for our novel results on the equivalence between Bayesian allocation games and team max-min games (Sect. 3.2), and on the computational complexity of both models (Sect. 3.3). The omitted proofs in this paper are deferred to the full version.

3.1 Equivalence Between Bayesian Repeated Games and Public Signaling Games (Reproducing Known Results)

We begin with reproducing the known result relating the classical model of Bayesian repeated games to public signaling games [2,14].

Definition 1 (Public Signaling Game [6–8]). *Consider a one-shot two-player zero-sum game* $G_{signal} = \left(\Pi, \{\mathcal{U}^\theta\}_{\theta\in\Theta}\right)$ *where players a-priori know nothing about* θ *besides its prior* Π*. We consider a credible principal who is privy to the realization of* θ*. The principal designs a* public signaling scheme*: a randomized function* $\varphi : \Theta \to \Delta(\Sigma)$ *mapping states of nature to an abstract set of signals* Σ*. The order of events is as follows:*

- *The principal commits to* φ*;*
- *The nature draws* $\theta \sim \Pi$ *and the principal learns* θ*;*
- *The principal invokes the signaling scheme to obtain a signal* $\sigma \sim \varphi(\theta)$*;*
- *Both players learn* σ*, and update their beliefs about the state* θ*, denoted as* $\Pi_{\varphi,\sigma}$*, according to the Bayes' rule:* $\Pi_{\varphi,\sigma}(\theta) = \frac{\mathbf{Pr}[\varphi(\theta)=\sigma]\cdot\Pi(\theta)}{\sum_{\theta'\in\Theta}\mathbf{Pr}[\varphi(\theta')=\sigma]\cdot\Pi(\theta')}$.
- *Players play the equilibrium strategies in the zero-sum game* $\mathcal{U}^{\Pi_{\varphi,\sigma}}$.

We assume that the principal designs φ *so as to maximize player 1's expected utility, the maximum value of which, denoted by* $\nu_{signal}(G_{signal})$*, is the game value of the public signaling game.*

It turns out the equilibrium in Bayesian repeated games corresponds to the solution of the above signaling problem in a precise sense, stated below [2,14].

Theorem 1. $\nu_{repeated}(G_{repeated}) = \nu_{signal}(G_{signal})$ *when* $G_{repeated} = G_{signal}$.

We will prove Theorem 1 by constructing the equilibrium strategy $\boldsymbol{s}_1^*$, $\boldsymbol{s}_2^*$ for player 1 and 2, respectively in the Bayesian repeated game $G_{\texttt{repeated}}$ from the solution of the public signaling game $G_{\texttt{signal}}$. For convenience, in the Bayesian repeated game, we will refer to player 1 (the informed player) as the *leader* and player 2 (the uninformed player) as the *follower*.

In particular, we will show that in the Bayesian repeated game $G_{\texttt{repeated}}$, if the leader plays strategy $\boldsymbol{s}_1^*$, then no matter how the follower reacts, the leader can guarantee himself an average utility at least the game value $\nu_{\texttt{signal}}(G_{\texttt{signal}})$ in the public signaling game $G_{\texttt{signal}}$ over the first T stages as $T \to \infty$. On the other hand, if the follower plays strategy $\boldsymbol{s}_2^*$, then no matter how the leader reacts, the follower can guarantee the leader an average utility at most $\nu_{\texttt{signal}}(G_{\texttt{signal}})$ over the first T stages as $T \to \infty$.

Lemma 1. *When $G_{repeated} = G_{signal}$, in the Bayesian repeated game $G_{repeated}$, consider the following strategy for the leader:*

- *upon learning the state θ of the nature, the leader invokes the optimal signaling strategy φ of the public signaling game G_{signal} to obtain $\sigma \sim \varphi(\theta)$;*
- *the leader then discards all information other than σ, i.e., behaves as if his belief is $\Pi_{\varphi,\sigma}$, and plays the maximin strategy in the game $\mathcal{U}^{\Pi_{\varphi,\sigma}}$, i.e., $\operatorname{argmax}_{s_1} \min_{s_2} \mathcal{U}^{\Pi_{\varphi,\sigma}}(s_1, s_2)$, repeatedly.*

This strategy can guarantee the leader an average expected utility $\nu_{signal}(G_{signal})$.

Although the strategy for the leader is easy to construct from the signaling scheme of the public signaling game, the follower's strategy is not so straightforward. The main difficulty is that there does not exist a credible principal in the repeated game as in the public signaling game, and therefore, the follower is uncertain about whether the leader exactly follows the scheme. In particular, the leader might have incentive to deviate by sending a different signal: conditioned on his type θ, choose σ^* such that $\sigma^* = \operatorname{argmax}_{\sigma\in\Sigma} \mathcal{U}^{\theta}\left(s_1^*(\sigma), s_2^*(\sigma)\right)$, where

$$s_1^*(\sigma) = \underset{s_1}{\operatorname{argmax}}\,\underset{s_2}{\min}\,\mathcal{U}^{\Pi_{\varphi,\sigma}}(s_1, s_2) \quad \text{and} \quad s_2^*(\sigma) = \underset{s_2}{\operatorname{argmin}}\,\underset{s_1}{\max}\,\mathcal{U}^{\Pi_{\varphi,\sigma}}(s_1, s_2).$$

In other words, the leader can send a signal σ^* that gives himself the maximum utility conditioned on θ. Therefore, the follower's strategy cannot rely on the possibly non-credible signaling scheme.

To circumvent this difficulty, we will construct an adaptive strategy for the follower, which does not depend on the non-credible signal σ but only depends on the prior Π and the history of play. Our approach relies on the solution of the dual program of the public signaling game. For convenience, given a distribution Π over Θ, let $f(\Pi) = \max_{s_1} \min_{s_2} \mathcal{U}^{\Pi}(s_1, s_2)$ be the game value of the induced game $\mathcal{U}^{\Pi}$. The problem of computing the optimal public signaling scheme can be formulated as the following linear program with infinitely many variables $x(\Pi')$ for $\Pi' \in \Delta(\Theta)$ [6–8]:

$$\begin{array}{lll} \max & \sum_{\Pi'\in\Delta(\Theta)} x(\Pi') \cdot f(\Pi') & \\ s.t. & \sum_{\Pi'\in\Delta(\Theta)} x(\Pi') \cdot \Pi'(\theta) = \Pi(\theta) & \forall\theta \in \Theta \\ & x(\Pi') \geq 0 & \forall\Pi' \in \Delta(\Theta) \end{array} \tag{1}$$

Intuitively, a signaling scheme can be viewed as a convex decomposition of the prior Π into a collection of posteriors $\{\Pi'\}$ [8,10]. Based on the primal, we can construct its dual with $|\Theta|$ variables $y(\theta)$ for $\theta \in \Theta$ as follows:

$$\begin{array}{l} \min \sum_{\theta\in\Theta} y(\theta)\cdot \Pi(\theta) \\ s.t. \ \sum_{\theta\in\Theta} y(\theta)\cdot \Pi'(\theta) \geq f(\Pi') \ \forall \Pi' \in \Delta(\Theta) \end{array} \tag{2}$$

Let x^* and y^* be the solution of the primal and the dual, respectively. By strong duality, $\sum_{\Pi'\in\Delta(\Theta)} x^*(\Pi')\cdot f(\Pi') = \sum_{\theta\in\Theta} y^*(\theta)\cdot \Pi(\theta) = \nu_{\mathtt{signal}}(G_{\mathtt{signal}})$. We will interpret y and Π as vectors such that $\boldsymbol{y} = \big(y(\theta_1), \cdots, y(\theta_{|\Theta|})\big)$ and $\boldsymbol{\Pi} = \big(\Pi(\theta_1), \cdots, \Pi(\theta_{|\Theta|})\big)$. The inner product $\langle \boldsymbol{y}, \boldsymbol{\Pi}\rangle$ is defined as $\sum_{\theta\in\Theta} y(\theta)\cdot\Pi(\theta)$. The next proposition directly follows the feasibility of $\boldsymbol{y}^*$ and strong duality:

Proposition 1. *For any prior $\boldsymbol{\Pi}$ in the public signaling game, there exists $\boldsymbol{y}^*$ such that $\langle \boldsymbol{y}^*, \boldsymbol{\Pi}\rangle = \nu_{signal}(G_{signal})$ and $\forall \Pi' \in \Delta(\Theta)$, $\langle \boldsymbol{y}^*, \boldsymbol{\Pi}'\rangle \geq f(\Pi')$.*

Hence, if the follower can ensure that for any strategy $\boldsymbol{s}_1$ deployed by the leader, there exists an adaptive mixed strategy $\boldsymbol{s}_2$ for the follower such that,

$$\forall \theta \in \Theta, \lim_{T\to\infty} \frac{\sum_{t=1}^{T} \mathcal{U}^{\theta}(s_1^t, s_2^t)}{T} \leq y^*(\theta), \tag{3}$$

then the average utility of the leader as $T \to \infty$ would be

$$\lim_{T\to\infty} \sum_{\theta\in\Theta} \Pi(\theta)\cdot \frac{\sum_{t=1}^{T} \mathcal{U}^{\theta}(s_1^t, s_2^t)}{T} \leq \sum_{\theta\in\Theta} \Pi(\theta)\cdot y^*(\theta) = \nu_{\mathtt{signal}}(G_{\mathtt{signal}}).$$

To prove (3), it is equivalent to show that $\mathcal{R}(\boldsymbol{y}^*) = \{\boldsymbol{v} \mid \boldsymbol{v} \leq \boldsymbol{y}^*\}$ is approachable.

Definition 2 (Blackwell's Approachability [4]). *Given a convex set $\mathcal{R}$ of vectors of utilities, we say $\mathcal{R}$ is* approachable *from the perspective of the follower, if for any strategy of the leader $\boldsymbol{s}_1$, there exists an adaptive strategy $\boldsymbol{s}_2$ for the follower such that* $\lim_{T\to\infty} \operatorname{dist}\left(\frac{1}{T}\sum_{t=1}^{T} \boldsymbol{\mathcal{U}}(s_1^t, s_2^t), \mathcal{R}\right) = 0$ *almost surely, where* $\boldsymbol{\mathcal{U}}(s_1, s_2) = \big(\mathcal{U}^{\theta_1}(s_1, s_2), \cdots, \mathcal{U}^{\theta_{|\Theta|}}(s_1, s_2)\big)$ *and* $\operatorname{dist}(\boldsymbol{u}, \mathcal{R}) = \min_{\boldsymbol{v}\in\mathcal{R}} \|\boldsymbol{v} - \boldsymbol{u}\|$.

Theorem 2 ([2,14]). *$\mathcal{R}(\boldsymbol{y}^*) = \{\boldsymbol{v} \mid \boldsymbol{v} \leq \boldsymbol{y}^*\}$ is approachable.*

To establish the approachability of $\mathcal{R}(\boldsymbol{y}^*)$, we first consider a halfspace $\mathcal{H}(\boldsymbol{\Pi}', b)$ such that $\boldsymbol{v} \in \mathcal{H}(\boldsymbol{\Pi}', b)$ if and only if $\langle \boldsymbol{\Pi}', \boldsymbol{v}\rangle \leq b$.

Lemma 2. *A halfspace $\mathcal{H}(\boldsymbol{\Pi}', b)$ is approachable if $f(\Pi') \leq b$.*

Theorem 3 ([4]). *A convex set $\mathcal{R}$ is approachable if and only if all halfspaces containing $\mathcal{R}$ are approachable.*

All that remains to show is that all halfspaces containing $\mathcal{R}(\boldsymbol{y}^*)$ are approachable.

Lemma 3. *All halfspaces containing $\mathcal{R}(\boldsymbol{y}^*) = \{\boldsymbol{v} \mid \boldsymbol{v} \leq \boldsymbol{y}^*\}$ are approachable.*

Proof. Notice that any minimal halfspace containing $\mathcal{R}(\boldsymbol{y}^*)$ must cross $\boldsymbol{y}^*$ by the construction of $\mathcal{R}(\boldsymbol{y}^*)$. Therefore, such a halfspace can be represented by $\mathcal{H}(\boldsymbol{\Pi}', \langle \boldsymbol{\Pi}', \boldsymbol{y}^* \rangle)$ with $\Pi' \in \Delta(\Theta)$. By Proposition 1, $f(\Pi') \leq \langle \boldsymbol{\Pi}', \boldsymbol{y}^* \rangle$, and therefore, by Lemma 2, $\mathcal{H}(\boldsymbol{\Pi}', \langle \boldsymbol{\Pi}', \boldsymbol{y}^* \rangle)$ is approachable.

Combining Theorem 3 and Lemma 3, we finish the proof of Theorem 2. We can then apply Blackwell's construction [4] to obtain an adaptive strategy for the follower that approaches $R(\boldsymbol{y}^*)$ almost surely.

Intuitively, at stage t, if $\frac{1}{t-1} \sum_{\tau=1}^{t-1} \boldsymbol{\mathcal{U}}(s_1^\tau, s_2^\tau) \notin \mathcal{R}(\boldsymbol{y}^*)$, then the follower first finds a halfspace $\mathcal{H}(\boldsymbol{\Pi}', \langle \boldsymbol{\Pi}', \boldsymbol{y}^* \rangle)$ that separates $\frac{1}{t-1} \sum_{\tau=1}^{t-1} \boldsymbol{\mathcal{U}}(s_1^\tau, s_2^\tau)$ and $\mathcal{R}(\boldsymbol{y}^*)$. Given such a Π', the follower plays the minimax strategy of $\mathcal{U}^{\Pi'}$ at stage t, and then the distance between the vector of average utilities and $\mathcal{R}(\boldsymbol{y}^*)$ will become smaller after stage t. Observe that the follower's strategy can be computed from the prior Π, the game $G_{\texttt{repeated}}$, and the history of play. In doing so, it guarantees that the expected average utility of the leader is at most $\nu_{\texttt{signal}}(G_{\texttt{signal}})$ in the limit, and Proposition 2 follows:

Proposition 2. *In a Bayesian repeated game $G_{\texttt{repeated}} = \left(\Pi, \left\{\mathcal{U}^\theta\right\}_{\theta \in \Theta}\right)$, given $\boldsymbol{y}^*$ satisfying Proposition 1 and an oracle to compute the minimax strategy of the zero-sum game $\mathcal{U}^{\Pi'}$ for all $\Pi' \in \Delta(\Theta)$, there exists an efficient algorithm to construct the follower's optimal strategy.*

We will elaborate the complexity of computing $\boldsymbol{y}^*$ in Sect. 3.3.

3.2 Equivalence Between Bayesian Allocation Games and Team Max-Min Games

Definition 3 (Team Max-Min Game [15]). *In a zero-sum team max-min game $G_{\texttt{team}} = \left(\left\{\mathcal{U}^\theta\right\}_{\theta \in \Theta}\right)$, in addition to player 1 and 2, there is a player 3 whose set of pure strategies is Θ. Player 1 and player 3 form a team and share the same utility such that when player 1 plays $s_1 \in S_1$, player 2 plays $s_2 \in S_2$, and player 3 plays $\theta \in \Theta$, the utility for both player 1 and player 3 is $\mathcal{U}^\theta(s_1, s_2)$, while the utility for player 2 is $-\mathcal{U}^\theta(s_1, s_2)$. A team max-min equilibrium is a Nash equilibrium that maximizes the team's utility and we denote its game value by $\nu_{\texttt{team}}(G_{\texttt{team}})$: $\nu_{\texttt{team}}(G_{\texttt{team}}) = \max_{s_1 \in \Delta(S_1), \Pi \in \Delta(\Theta)} \min_{s_2 \in \Delta(S_2)} \mathcal{U}^\Pi(s_1, s_2)$.*

We emphasize that player 1's strategy and player 3's strategy are not allowed to be correlated; otherwise, the team max-min game degenerates to a classic two-player zero-sum game in which player 1 and 3 can be treated as a single player. [15] show that a team max-min equilibrium always exists. It turns out the equilibrium in Bayesian allocation games corresponds to the solution of the above team max-min games in a precise sense, stated below.

Theorem 4. *$\nu_{\texttt{alloc}}(G_{\texttt{alloc}}) = \nu_{\texttt{team}}(G_{\texttt{team}})$ when $G_{\texttt{alloc}} = G_{\texttt{team}}$.*

To prove Theorem 4, we will construct strategies for players in the Bayesian allocation game from the equilibrium strategies in the team max-min game.

Lemma 4. *When $G_{alloc} = G_{team}$, let s_1^*, s_2^*, Π^* be the equilibrium strategies for the team max-min game G_{team}. In the Bayesian allocation game G_{alloc}, consider the following strategy for the leader:*

- *set the prior Π to be Π^*; then repeatedly play strategy s_1^* for every stage.*

This strategy can guarantee the leader an average expected utility $\nu_{team}(G_{team})$.

In comparison to the Bayesian repeated games in which the follower knows the prior, the follower does not even know the prior set by the leader in the Bayesian allocation game. To overcome this obstacle, observe that in the Bayesian repeated game, the approachability of a convex set is a property that only depends on the collection of games $\left(\{\mathcal{U}^\theta\}_{\theta\in\Theta}\right)$ but independent of the prior. Motivated by this observation, we show that $\mathcal{R}(\nu_{\mathtt{team}}(G_{\mathtt{team}}) \cdot \mathbf{1}) = \{\boldsymbol{v} \mid \boldsymbol{v} \le \nu_{\mathtt{team}}(G_{\mathtt{team}}) \cdot \mathbf{1}\}$ is approachable where $\mathbf{1}$ is a vector of all ones.

Lemma 5. *$\mathcal{R}(\nu_{team}(G_{team}) \cdot \mathbf{1})$ is approachable.*

It is straightforward to show that, when $\mathcal{R}(\nu_{\mathtt{team}}(G_{\mathtt{team}}) \cdot \mathbf{1})$ is approachable, for any prior $\Pi \in \Delta(\Theta)$, the average utility of the leader is at most $\nu_{\mathtt{team}}(G_{\mathtt{team}})$.

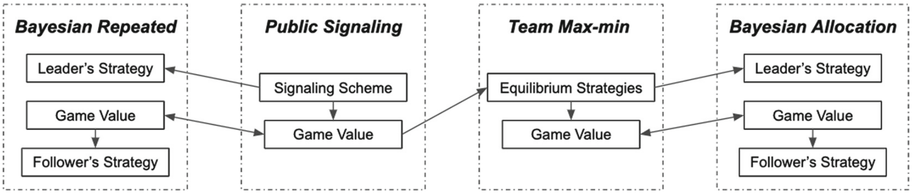

Fig. 1. The relationships of computational problems, assuming the minimax strategy of $\mathcal{U}^\Pi$ can be computed efficiently for all $\Pi \in \Delta(\Theta)$: the arrows point to problems that are computationally easier.

3.3 Computational Complexity of the Follower's Optimal Strategy

As demonstrated before, constructing the follower's optimal strategy in Bayesian repeated games requires a solution to the dual program (2). Hence, it is not immediate that one can efficiently construct the follower's optimal strategy if the public signaling game is efficiently solvable. Here, we say an algorithm is efficient if the running time of the algorithm is polynomial in terms of the number of states $|\Theta|$, and the number of pure strategies $|S_1| + |S_2|$.

We manage to show that, when the minimax strategy of $\mathcal{U}^\Pi$ can be computed efficiently for all $\Pi \in \Delta(\Theta)$, in both Bayesian repeated games and Bayesian allocation games, the follower's optimal strategy can be efficiently constructed if the corresponding game values are given. We further show that team max-min game is computationally easier than the public signaling game, and therefore, Bayesian allocation game is computationally easier than the Bayesian repeated game. Figure 1 summarizes the relationships of the computational problems discussed in this section, while the proofs are deferred to the full version.

4 Bayesian Repeated Security Games

In Sect. 3, we have shown that Bayesian repeated games can be reduced to public signaling games, while Bayesian allocation games can be reduced to team max-min games. However, it has been shown that both public signaling games and team max-min games are computationally intractable for general zero-sum games and even worse, no FPTAS is possible [5,8]. Particularly, public signaling games do not even admit PTAS [3,13].

Motivated by the applications in the domain of repeated security games, we will concern ourselves with repeated games where the stage game is a *security game* of a particularly simple form. The one-shot complete-information security games are described by a set L of *locations*, a set M of *treasures*, and a set R of *defensive resources*. For convenience, we use $\bot$ to denote a *null* treasure or a *null* defensive resource. $v : L \times (M \cup \bot) \rightarrow \mathbb{R}_{\geq 0}$ is a *location-treasure importance* function such that $v(\ell, m)$ characterizes the utility loss of the defender if location $\ell \in L$ with treasure $m \in M$ allocated is attacked without defense. In addition, there is a *defense-quality* function $q : L \times (M \cup \bot) \times (R \cup \bot) \rightarrow \{0, 1\}$ such that $q(\ell, m, r)$ characterizes the effectiveness of allocating defensive resource $r \in R$ to defend location $\ell \in L$ that hosts treasure m. Note that in our setting, a defensive resource is either 100% effective for a combination of location and treasure or totally useless. For a *null* treasure, we have $v(\ell, \bot) = 0$ for all ℓ, and for a *null* defensive resource, we have $q(\ell, m, \bot) = 0$ for all ℓ and m.

A state of nature is a matching $\theta : L \rightarrow M$ that maps the locations to treasures such that for any $i, j \in L$ with $i \neq j$, $\theta(i) \neq \bot$, and $\theta(j) \neq \bot$, we have $\theta(i) \neq \theta(j)$. A pure strategy for the defender is also a matching $D : L \rightarrow R$ that maps the locations to the defensive resources such that for any $i, j \in L$ with $i \neq j$, $D(i) \neq \bot$, and $D(j) \neq \bot$, we have $D(i) \neq D(j)$. Finally, a pure strategy for the attacker is a single location $a \in L$ to attack. A mixed strategy is naturally a distribution over such functions. The defender's utility under θ when the defender plays D and the attacker plays a is $\mathcal{U}^{\theta}(D, a) = -\big(1 - q\big(a, \theta(a), D(a)\big)\big) \cdot v\big(a, \theta(a)\big)$, while the attacker's utility is simply $-\mathcal{U}^{\theta}(D, a)$.

We say the treasures are *homogeneous* if for all $m \in M$, $v(\ell, m)$ equals to some constant for all $\ell \in L$; the locations are *homogeneous* if for all $\ell \in L$, $v(\ell, m)$ equals to some constant for all $m \in M$; and the defensive resources are *homogeneous* if $q(\ell, m, r) = 1$ for all $\ell \in L$, $m \in M$, and $r \in R$. If the condition of homogeneity is not satisfied, we say they are *heterogeneous*.

We analyze the complexity of repeated security games under the contexts of both Bayesian repeated games and Bayesian allocation games. In Bayesian repeated games, an algorithm is efficient if its running time is in polynomial of $|\Theta|$, $|L|$, $|M|$, and $|R|$; while in Bayesian allocation games, an algorithm is efficient if its running time is in polynomial of $|L|$, $|M|$, and $|R|$.

Proposition 3. *Given the marginals of Π, the optimal strategies for both the defender and the attacker in the security game $\mathcal{U}^{\Pi}$ can be computed efficiently.*

However, for our class of security games with a general prior Π, computing the game value of the Bayesian repeated games is computationally intractable.

Theorem 5. *It is strongly NP-hard to compute the game value of the Bayesian repeated games with a security game as the stage game, even when all of treasures, locations, and defensive resources are homogeneous. Moreover, no FPTAS is possible. Consequently, it is strongly NP-hard to compute any representation of the equilibrium which permits computing the game value.*

5 Bayesian Allocation Security Games

We turn to Bayesian allocation games with a security game as the stage game. It turns out that a Bayesian allocation game with a security game as the stage game can be efficiently solved when only the treasures are heterogeneous (Fig. 2).

Theorem 6. *There exists an efficient algorithm to compute the game value and the defender's optimal strategy of a Bayesian allocation game with a security game as the stage game, when only the treasures are heterogeneous.*

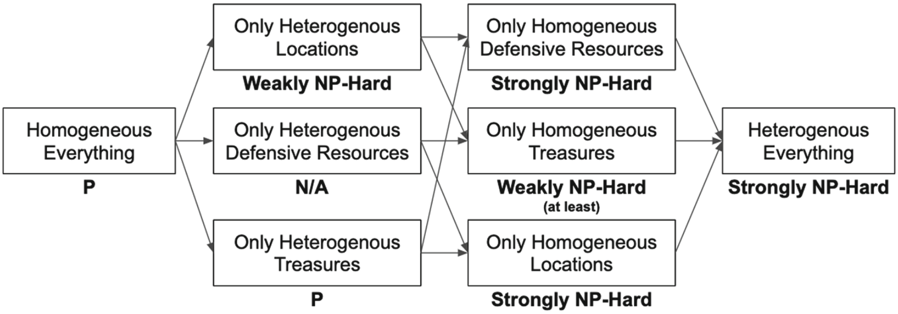
Homogeneous Everything
P
Only Heterogenous Locations
Weakly NP-Hard
Only Heterogenous Defensive Resources
N/A
Only Heterogenous Treasures
P
Only Homogeneous Defensive Resources
Strongly NP-Hard
Only Homogeneous Treasures
Weakly NP-Hard
(at least)
Only Homogeneous Locations
Strongly NP-Hard
Heterogenous Everything
Strongly NP-Hard

Fig. 2. The computational complexity of Bayesian allocation games with a security game as the stage game: the arrows point to more general versions of the problem.

Moreover, the following lemma illustrates that one can efficiently construct the attacker's strategy when the game value is given.

Lemma 6. *Given the game value of a Bayesian allocation game with a security game as the stage game, there exists an efficient algorithm to compute the attacker's optimal strategy.*

Therefore, one can efficiently construct both the defender's optimal strategy and the attacker's optimal strategy when only the treasures are heterogeneous. However, going beyond, the problem becomes computationally intractable.

Theorem 7. *It is weakly NP-hard to compute the game value of the Bayesian allocation games with a security game as the stage game, when only the locations are heterogeneous. Moreover, there exists a pseudo-polynomial time algorithm that can compute the game value.*

Theorem 8. *It is strongly NP-hard to compute the game value of the Bayesian allocation games with a security game as the stage game, when only the defensive resources are homogeneous, or only the locations are homogeneous.*

There are three other settings that have not been discussed: (1) heterogeneous everything; (2) only treasures are homogeneous; and (3) only defensive resources are heterogeneous. For the setting in which everything is heterogeneous, it is also strongly NP-hard to compute the game value since it is a more general setting than the settings in which only defensive resources are homogeneous or only locations are homogeneous. As for the setting in which only treasures are homogeneous, it is at least weakly NP-hard to compute the game value since it is a more general setting than the case in which only locations are heterogeneous. We leave it as an open question to settle whether it is strongly NP-hard. Finally, for the setting in which only defensive resources are heterogeneous, this setting is not well-defined: since the locations and the treasures are homogeneous, a defensive resource should be either effective or ineffective for any combination of the locations and the treasures. Consequently, the defender can simply eliminate the ineffective defensive resources to focus on effective ones, which reduces the problem to the case in which everything is homogeneous.

References

1. Abernethy, J., Bartlett, P.L., Hazan, E.: Blackwell approachability and no-regret learning are equivalent. In: Proceedings of the 24th Annual Conference on Learning Theory, pp. 27–46 (2011)
2. Aumann, R.J., Maschler, M.: Repeated games with incomplete information. MIT Press, Cambridge (1995)
3. Bhaskar, U., Cheng, Y., Ko, Y.K., Swamy, C.: Hardness results for signaling in Bayesian zero-sum and network routing games. In: Proceedings of the 2016 ACM Conference on Economics and Computation, pp. 479–496 (2016)
4. Blackwell, D.: An analog of the minimax theorem for vector payoffs. Pac. J. Math. **6**(1), 1–8 (1956)
5. Borgs, C., Chayes, J., Immorlica, N., Kalai, A.T., Mirrokni, V., Papadimitriou, C.: The myth of the folk theorem. Games Econ. Behav. **70**(1), 34–43 (2010)
6. Cheng, Y., Cheung, H.Y., Dughmi, S., Emamjomeh-Zadeh, E., Han, L., Teng, S.H.: Mixture selection, mechanism design, and signaling. In: 2015 IEEE 56th Annual Symposium on Foundations of Computer Science. pp. 1426–1445. IEEE (2015)
7. Dughmi, S.: Algorithmic information structure design: a survey. ACM SIGecom Exchanges **15**(2), 2–24 (2017)
8. Dughmi, S.: On the hardness of designing public signals. Games Econ. Behav. **118**, 609–625 (2019)
9. Fang, F., Nguyen, T.H.: Green security games: apply game theory to addressing green security challenges. ACM SIGecom Exchanges **15**(1), 78–83 (2016)
10. Kamenica, E., Gentzkow, M.: Bayesian persuasion. Am. Econ. Rev. **101**(6), 2590–2615 (2011)
11. Neumann, J.V.: Zur theorie der gesellschaftsspiele. Mathematische annalen **100**(1), 295–320 (1928)

12. Rockmore, D.: What's missing from "the imitation game", November 2014. https://www.newyorker.com/tech/annals-of-technology/imitation-game-alan-turing. Accessed 23 Jan 2020
13. Rubinstein, A.: Eth-hardness for signaling in symmetric zero-sum games. CoRR abs/1510.04991 (2015)
14. Sorin, S.: A First Course on Zero-Sum Repeated Games, vol. 37. Springer, Heidelberg (2002)
15. von Stengel, B., Koller, D.: Team-maxmin equilibria. Games Econ. Behav. **21**(1–2), 309–321 (1997)
16. Tambe, M.: Security and Game Theory: Algorithms, Deployed Systems, Lessons Learned. Cambridge University Press, Cambridge (2011)
17. Wang, Y., et al.: Deep reinforcement learning for green security games with real-time information. In: Proceedings of the AAAI Conference on Artificial Intelligence, vol. 33, pp. 1401–1408 (2019)
18. Xu, H., et al.: Optimal patrol planning for green security games with black-box attackers. In: Rass, S., An, B., Kiekintveld, C., Fang, F., Schauer, S. (eds.) International Conference on Decision and Game Theory for Security. GameSec 2017. Lecture Notes in Computer Science, vol. 10575, pp. 458–477. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-68711-7_24