Efficient Strategy Computation in Zero-Sum Asymmetric information Repeated Games

Lichun Li and Jeff S. Shamma

Abstract—Zero-sum asymmetric information games model decision making scenarios involving two competing players who have different information about the game being played. A particular case is that of nested information, where one (informed) player has superior information over the other (uninformed) player. This paper considers the case of nested information in repeated zero-sum games and studies the computation of strategies for both the informed and uninformed players for finite-horizon and discounted infinite-horizon nested information games. For finite-horizon settings, we exploit that for both players, the security strategy, and also the opponent’s corresponding best response, depend only on the informed player’s history of actions. Using this property, we formulate an LP computation of player strategies that is linear in the size of the uninformed player’s action set. For the infinite-horizon discounted game, we construct LP formulations to compute the approximated security strategies for both players, and show that the worst case performance difference between the approximated security strategies and the security strategies converges to zero exponentially. Finally, we illustrate the results on a network interdiction game between an informed system administrator and an uniformed intruder.

I. INTRODUCTION

There are many competitive settings in which players have asymmetric information about the underlying state of the game. Examples include cyber security problems [1], [2], resource competitions in air transportation systems [3], [4], national defense [5], [6], economic systems [7], power networks [8] and so on. In these systems, because of the non-cooperation between players, a player usually holds private information that is not shared with the other players, which causes the information asymmetry.

This paper focuses on two player zero-sum repeated games with asymmetric information, where one (informed) player knows the underlying state of the game while the other (uninformed) player only knows the prior distribution of the state. At the beginning of the game, the state of the game is initially selected, once and for all, at random according to the prior distribution. Repeated interactions means that players play over stages and can make observations about past play. Here we assume the case of full monitoring, i.e., a player can observe the actions taken by both players, and perfect recall, i.e., a player records the history actions of both players. The one-stage payoff of the informed player, i.e. the loss of the uninformed player, is decided by the state of the game and the actions of both players. While neither player can directly observe the one-stage payoff, the informed player can derive the one-stage payoff since it knows the state and the current action pair. The fact that the one-stage payoff is unavailable to the uninformed player happens in practice. For example, when jamming wireless sensor networks, an attacker may observe which channel the network uses (action of the network), but cannot measure the throughput in the channel (payoff of the network) [2].

We first study finite stage (N-stage) repeated asymmetric information games. The finite stage repeated asymmetric information game can be expressed as a finite game tree. With perfect recall, the security strategies of both players can be computed by solving a linear program whose size is linear with respect to the size of the game tree, i.e., polynomial in the size of the action sets of both players, linear in the size of the state set, and exponential in the time horizon [9]. The challenge in finite stage games is how to reduce the computational complexity. Based on the fact that both players have a security strategy that is independent of the history actions of the uninformed player [6], [10], [11], this paper shows that if a player uses a strategy independent of the uninformed player’s history actions, the opponent’s best response is also independent of the uninformed player’s history actions. It implies that a player does not need to record the history actions of the uninformed player to compute its security strategy. Our main contribution in finite stage games is developing LP formulations whose sizes are only linear in the size of the uninformed player’s action set to compute the security strategies for both players.

A similar setting was studied in [12] with a random payoff. Besides actions of players, the one stage payoff is also common information in [12]. Based on the common information, the authors design an asymptotically optimal strategy for the uninformed player such that the worst case relative loss of the uninformed player is either finite or increasing with rate $\log(N)$ as $N$ goes to infinity. This paper adopts the game model in [6], [13] with fixed payoff that is not directly observable. We are interested in computing security strategies of both players, i.e., the Nash equilibrium of the finite stage game. Notice that with the assumption of mutually absolute continuity in [12], the game model in [12] is neither more nor less general than the one in this paper.

We also extend the time horizon to infinity, and study discounted repeated asymmetric information games. Compared with finite stage games, discounted repeated asymmetric information games have two main challenges. The first challenge is to find out a fixed sized sufficient statistic for each player,
since history based strategy requires prohibitive memory to record the history actions as the time horizon goes to infinity. The second challenge is to find a computationally tractable approximated security strategy for each player with guaranteed performance since computing the security strategies is non-convex [14], [15]. Our main contributions in discounted games are providing both players with approximated security strategies whose performance converges to the game value exponentially, and constructing LP formulations to compute the approximated security strategies for both players.

For the informed player, the first challenge has been addressed in the previous work [6], and the sufficient statistic of the informed player is the posterior probability of the state of the game conditioned on the history action of the informed player, which is also called the belief. For the second challenge, we first use the game value of a finite stage \((N\)-stage) discounted game, a truncated version of the infinite stage discounted game, to approximate the game value of the infinite stage discounted game, and then construct an LP formulation to compute an approximated security strategy based on the approximated game value. Such an approximated game value is shown to converge to the true game value exponentially fast with respect to \(N\), and the difference between the worst case payoff of the approximated security strategy and the game value and has a finite upper bound which converges to zero exponentially fast with respect to \(N\). The technique used in this section is adopted from our previous work [16] which focused on the informed player’s approximated security strategy in discounted stochastic asymmetric information games. This paper further studies the strategies of the uninformed player.

For the uninformed player, the first challenge is partially addressed in previous work [10], [11] which states that the sufficient statistic of the uninformed player is a real vector, which is called anti-discounted regret in this paper, of the same size of the game state, and provides a recursive formulation to update the sufficient statistic stage by stage. While [10], [11] states that the initial regret is an optimal solution to a specified optimal problem, the physical meaning of the initial regret is still unclear, and a tractable computation method is missing. This paper shows that the special initial regret is the difference between 0 and the worst case payoff of the uninformed player’s security strategy given every possible state of the game in the infinite stage discounted game. Computing the special initial regret is non-convex. Therefore, we use the worst case payoff of the uninformed player’s security strategy given every possible game state in finite stage discounted game to approximate the one in the infinite stage discounted game, and construct a linear program to compute it. To address the second challenge, we use a similar technique that is used to derive the approximated security strategy for the informed player to construct an approximated security strategy for the uninformed player, and provide an LP formulation to compute the approximated strategy. The difference between the worst case performance of the uninformed player’s approximated security strategy and the game value is always finite, and converges to zero exponentially fast with respect to \(N\).

This paper extends the results in our previous work [17] from finite stage to infinite stage formulations. For infinite stage discounted games, our previous work [16] studies approximated security strategies in stochastic discounted games for the informed player only. This paper adopts the same technique to develop approximated security strategies for the informed player in repeated discounted games, and further studies approximated security strategies for the uninformed player. As mentioned above, for the uninformed player, our main contributions are clarifying the physical meaning of the initial condition of the uninformed player’s sufficient statistic, constructing a linear program to compute an approximated initial condition, developing an approximated security strategy whose worst case performance converges to the game value exponentially fast, and providing an LP formulation to compute the approximated security strategy.

The remainder of this paper is organized as follows. Section II presents the main results for finite stage games. Section III discusses discounted infinite horizon games. Section IV illustrates the results on a network interdiction game. Finally, Section V presents some concluding remarks.

II. Finite Stage Asymmetric Repeated Games

Notation. Let \(\mathbb{R}^n\) denote \(n\)-dimensional real space. Given a finite set \(K\), its cardinality is denoted by \(|K|\), and \(\Delta(K)\) is the set of probability distributions over \(K\). The vectors \(1\) and \(0\) are appropriately dimensioned column vectors with all elements being 1 and 0, respectively. For a sequence of real numbers \(v(0), v(1), v(2), ...,\), we adopt the convention that \(\sum_{t=1}^{N} v(t) = 0\), and \(\prod_{t=1}^{N} v(t) = 1\). The supreme norm of a function \(f : D \to \mathbb{R}\) is defined as \(\|f\|_{\sup} = \sup_{x \in D} |f(x)|\), where \(D\) is a non-empty set.

A. Setup

A two-player zero-sum asymmetric repeated game is specified by a five-tuple \((K, A, B, M, p_0)\), where

- \(K\) is a non-empty finite set, called the state set, the elements of which are called states.
- \(A\) and \(B\) are non-empty finite sets, called player 1 and 2’s action sets, respectively.
- \(M : K \times A \times B \to \Delta(K)\) is the one-stage payoff function of player 1, or the one-stage penalty function of player 2. Hence, this is a zero-sum game. \(M^k\) indicates the payoff matrix given state \(k \in K\). The matrix element \(M_{a,b}^k\), also denoted as \(M(k, a, b)\), is the payoff given state \(k \in K\), player 1’s action \(a \in A\), and player 2’s action \(b \in B\). The notation \(M_{a,b}^k\) indicates the row vector payoff given state \(k\) and player 1’s action \(a \in A\).
- \(p_0 \in \Delta(K)\) is the initial probability on \(K\). We assume that \(p^k_0 > 0\) for any \(k \in K\).

An \(N\)-stage asymmetric repeated game is played as follows. Let \(a_t, b_t\) denote the actions of player 1 and player 2 at stage \(t \in \{1, 2, ..., N\}\), respectively. At stage \(t = 1\), a state \(k\) is chosen once and for all according to the probability distribution \(p_0\), and communicated to player 1 only. Player 1 and 2 are called the informed and the uninformed player, respectively. Each player chooses its action independently, and the pair \((a_1, b_1)\) is observed by both players. At stage \(t = 2\), both players again simultaneously choose their actions which
are observable by both players. The payoff of player 1 and player 2 at stage \( t \) is \( M^k_{a_t,b_t} \) and \( -M^k_{a_t,b_t} \), respectively. Since the sum of two players’ payoffs is zero, this is a zero-sum game. The process is repeated for the remaining stages. The above setting is publicly known. At every stage \( t \), player 2 observes both players’ actions, and player 1 observes not only both players’ actions but also the state. Therefore, although neither player can observe the one-stage payoff directly, player 1 can derive the one-stage payoff \( M^k(a_t, b_t) \) based on its observed information.

More formally, we will use the concept of behavior strategies. For any stage \( t = 1, \ldots, N \), the histories of player 1 and 2’s actions prior to time \( t \) are denoted by \( h^A_t = \{a_1, \ldots, a_{t-1}\} \) and \( h^B_t = \{b_1, \ldots, b_{t-1}\} \), respectively. For \( t = 1 \), the null histories are denoted by \( h^A_1 = h^B_1 = \emptyset \). The corresponding sets of possible history actions are denoted by \( H^A_t = A^{t-1} \) and \( H^B_t = B^{t-1} \). With its available information \( K \times H^A_t \times H^B_t \) at time \( t \), player 1’s behavior strategy is a collection of mappings \( \sigma = (\sigma_t)_{t=1}^N \), where each \( \sigma_t \) is a map from \( K \times H^A_t \times H^B_t \) to \( \Delta(A) \). Similarly, but taking into account the lack of information on the state \( k \in K \), a behavior strategy of player 2 is a collection of mappings \( \tau = (\tau_t)_{t=1}^N \), where \( \tau_t \) is a map from \( H^A_t \times H^B_t \) to \( \Delta(B) \). Denote by \( \Sigma \) and \( \mathcal{T} \) the sets of behavior strategies of player 1 and 2, respectively. The values \( \sigma^\ell_t(k, h^A_t, h^B_t) \) for \( a \in A \) and \( \tau^\nu_t(h^A_t, h^B_t) \) for \( b \in B \) denote the probabilities of playing \( a \) and \( b \) at stage \( t \), respectively, given the histories \( h^A_t \in H^A_t \) and \( h^B_t \in H^B_t \), and realized state, \( k \in K \).

Play proceeds as follows. As previously stated, at stage \( t = 1 \), a state \( k \) is chosen once and for all according to the probability distribution \( p_0 \). The action \( a_1 \) is a randomized outcome according to the behavior strategy distribution \( \sigma_1(k, \emptyset, \emptyset) \in \Delta(A) \), and the action \( b_1 \) is a randomized outcome according to the behavior strategy distribution \( \tau_1(\emptyset, \emptyset) \). At stage \( t = 2, \ldots, N \), the action \( a_t \) is a randomized outcome according to the behavior strategy distribution \( \sigma_t(k, h^A_t, h^B_t) \in \Delta(A) \), and the action \( b_t \) is a randomized outcome according to the behavior strategy distribution \( \tau_t(h^A_t, h^B_t) \), where we assume that these outcomes are conditionally independent given \( h^A_t \) and \( h^B_t \).

A triple \((p_0, \sigma, \tau)\) induces a probability distribution \( P_{p_0,\sigma,\tau} \) on the set \( \Omega = K \times (A \times B)^N \) of plays. Let \( E_{p_0,\sigma,\tau} \) denote the corresponding expectation. The payoff with initial probability \( p_0 \) and strategies \( \sigma \) and \( \tau \) of the \( N \)-stage asymmetric information repeated game is defined as

\[
\gamma_N(p_0, \sigma, \tau) = E_{p_0,\sigma,\tau} \left[ \sum_{t=1}^{N} M(k, a_t, b_t) \right].
\]

The \( N \)-stage game \( \Gamma_N(p_0) \) is defined as the two-player zero-sum asymmetric repeated game equipped with initial distribution \( p_0 \), strategy spaces \( \Sigma \) and \( \mathcal{T} \), and payoff function \( \gamma_N(p_0, \sigma, \tau) \). In game \( \Gamma_N(p_0) \), the informed player seeks to maximize the payoff \( \gamma_N(p_0, \sigma, \tau) \), while the uninformed player seeks to minimize it.

For the \( N \)-stage game \( \Gamma_N(p_0) \), the security level \( \overline{V}_{N}(p_0) \) of the informed player is defined as

\[
\overline{V}_{N}(p_0) = \max_{\sigma \in \Sigma} \min_{\tau \in \mathcal{T}} \gamma_N(p_0, \sigma, \tau),
\]

and the strategy \( \sigma^* \in \Sigma \) which achieves the security level is called the security strategy of the informed player. Similarly, the security level \( \underline{V}_{N}(p_0) \) of the uninformed player is defined as

\[
\underline{V}_{N}(p_0) = \min_{\sigma \in \Sigma} \max_{\tau \in \mathcal{T}} \gamma_N(p_0, \sigma, \tau),
\]

and the strategy \( \tau^* \in \mathcal{T} \) which achieves the security level is called the security strategy of the uninformed player. When \( \overline{V}_{N}(p_0) = \underline{V}_{N}(p_0) \), we say game \( \Gamma_N(p_0) \) has a value, i.e., there exists a Nash equilibrium. Since the game \( \Gamma_N(p_0) \) is a finite game, the game value always exists, and is denoted by \( V_N(p_0) \) (see page 60 in [11]). While [11] showed the existence of the game value and the recursive formula to compute it, no explicit method was provided to compute the game value and the security strategies.

### B. \( H^B \) independent strategies

A fundamental difference between a repeated asymmetric information game and a one-shot asymmetric information game is that in the repeated asymmetric information game, the uninformed player can learn the game state from the informed player’s history actions. Indeed, the uninformed player’s belief about the state plays a crucial role for both players to make decisions [6]. Since only the informed player’s actions are directly related to the game state, the uninformed player’s history actions do not provide extra information about the game state given the informed player’s history actions. Therefore, it is not surprising to see that given informed player’s history actions, both players’ security strategies are independent of the uninformed player’s history actions [11]. In our setting, the feedback information for the uninformed players to figure out its security strategy is the informed player’s actions which reflect the state information.

Let’s define an \( H^B \) independent behavior strategy of player 1 as a collection of mappings \( \bar{\sigma} = (\bar{\sigma}_t)_{t=1}^N \) where each \( \bar{\sigma}_t \) is a map from \( K \times H^A_t \) to \( \Delta(A) \). Similarly, an \( H^B \) independent behavior strategy of player 2 is a collection of mappings \( \bar{\tau} = (\bar{\tau}_t)_{t=1}^N \) where \( \bar{\tau}_t \) is a map from \( H^A_t \) to \( \Delta(B) \). Denote by \( \bar{\Sigma} \) and \( \bar{T} \) the sets of \( H^B \) independent behavior strategies of player 1 and 2. Clearly, \( \bar{\Sigma} \) and \( \bar{T} \) are subsets of \( \Sigma \) and \( \mathcal{T} \), respectively.

**Proposition 1** ([18], [10], [11]). Consider a two-player zero-sum \( N \)-stage asymmetric repeated game \( \Gamma_N(p_0) \). Each player has a security strategy in game \( \Gamma_N(p_0) \) that is independent of player 2’s history actions, i.e.,

\[
\max_{\sigma \in \bar{\Sigma}} \min_{\tau \in \bar{T}} \gamma_N(p_0, \sigma, \tau) = \max_{\sigma \in \bar{\Sigma}} \min_{\tau \in \bar{T}} \gamma_N(p_0, \sigma, \tau),
\]

\[
\min_{\tau \in \bar{T}} \gamma_N(p_0, \sigma, \tau) = \min_{\tau \in \bar{T}} \gamma_N(p_0, \sigma, \tau).
\]

If one player’s behavior strategy is independent of the uninformed player’s history actions, the other player’s best response is independent of the uninformed player’s history actions also.

**Proposition 2.** Consider a two-player zero-sum \( N \)-stage asymmetric repeated game \( \Gamma_N(p_0) \). For any \( \bar{\sigma} \in \bar{\Sigma} \), and any
\[ \tau \in \bar{T}, \]

\[ \min_{\tau \in \bar{T}} \gamma_N(p_0, \bar{\sigma}, \tau) = \min_{\bar{\tau} \in \bar{T}} \gamma_N(p_0, \bar{\sigma}, \bar{\tau}) \quad (1) \]

\[ \max_{\sigma \in \Sigma} \gamma_N(p_0, \sigma, \bar{\tau}) = \max_{\bar{\sigma} \in \bar{\Sigma}} \gamma_N(p_0, \bar{\sigma}, \bar{\tau}). \quad (2) \]

Proof: Since \( \bar{T} \in T \), we have \( \min_{\tau \in \bar{T}} \gamma_N(p_0, \bar{\sigma}, \tau) \leq \min_{\bar{\tau} \in \bar{T}} \gamma_N(p_0, \bar{\sigma}, \bar{\tau}) \). Meanwhile, for any \( \tau \in T \), we can design \( \frac{\partial}{\partial h} (h_i^A) = \sum_{h_i^A \in H_i} \sum_{h_{i+1}^A \in H_{i+1}} \sigma_{x_{i+1}}^A (h_{i+1}^A) \) for all \( t = 1, \ldots, N \), such that \( \gamma_N(p_0, \sigma, \bar{\tau}) = \gamma_N(p_0, \sigma, \bar{\tau}). \) Hence, we have \( \min_{\tau \in T} \gamma_N(p_0, \bar{\sigma}, \tau) \leq \min_{\bar{\tau} \in \bar{T}} \gamma_N(p_0, \bar{\sigma}, \bar{\tau}) \). Therefore, equation (1) is shown.

Similarly, \( \Sigma \in \Sigma \) implies that \( \max_{\sigma \in \Sigma} \gamma_N(p_0, \sigma, \bar{\tau}) \geq \max_{\bar{\sigma} \in \bar{\Sigma}} \gamma_N(p_0, \bar{\sigma}, \bar{\tau}) \). Meanwhile, for any \( \sigma \in \Sigma \), we can design \( \frac{\partial}{\partial h} \gamma_N(p_0, \sigma, \bar{\tau}) \) for all \( t = 1, \ldots, N \), such that \( \gamma_N(p_0, \sigma, \bar{\tau}) = \gamma_N(p_0, \sigma, \bar{\tau}). \) Therefore, equation (2) is shown.

Notice that Proposition 1 and 2 are different in the sense that while Proposition 1 states that the security strategies are \( H^B \) independent, Proposition 2 shows that the best responses to \( H^B \) independent strategies are also \( H^B \) independent.

Proposition 1 and 2 imply that when computing players’ security strategies, we can ignore the uninformed player’s history actions, which greatly reduces the number of both players’ information sets in the extensive game tree, and hence reduces the computational complexity of the security strategies.

C. LP formulations of security strategies

An \( N \)-stage asymmetric repeated game, as a finite game, can always be expressed as a finite extensive game tree [19]. Assuming perfect recall, i.e., each player can record all history actions of both players, we can use sequence form to construct a linear program to compute the security strategy [9]. Roughly speaking, in two-player zero-sum games, with sequence form, the total payoffs at leaf nodes of the game tree are provided first, then the probability of the sequence from the root node to the leaf node is characterized in the form of every player’s realization plan, i.e., production of a player’s strategies along the sequence, and finally an LP formulation is derived to compute the security strategies of two players based on the strong duality theorem [20]. Interested readers can check [9] for more details. The size of the corresponding linear program is linear in the size of the game tree, and hence polynomial in the size of the uninformed player’s action set [9]. In this subsection, we will adopt realization plan in the sequence form, and take advantage of the \( H^B \) independent strategies to develop LP formulations with reduced computational complexity to compute the \( H^B \) independent security strategies.

As in the sequence form, we define the realization plan \( q_i(h^A_t; k) \) of the informed player’s history action sequence \( h^A_t \) given state \( k \) at stage \( t \) as

\[ q_i(h^A_t; k) = \prod_{s=1}^{t-1} \sigma_{x_s}^A (k, h^A_s), \quad (3) \]

where \( a_s \) and \( h^A_s \) are the informed player’s action and history actions at stage \( s \) in the history action sequence \( h^A_t \), denoted by \( a_s, h^A_s \in h^A_t \). Therefore, the realization plan \( q \) satisfies the following constraints:

\[ q_i(h^A_t; k) = 1, \quad \forall k \in K, \quad (4) \]

\[ \sum_{a_i \in \mathcal{A}} q_{i+1}(h^A_{i+1}; a_{i+1}) = q_i(h^A_i; k), \quad \forall k \in K, h^A_i \in H^A_i, \quad (5) \]

\[ q_i(h^A_t; k) \geq 0, \quad \forall k \in K, h^A_i \in H^A_i, \quad (6) \]

where \( (h^A_t, a_t) \) indicates concatenation. A realization plan of the informed player is a collection of the informed player’s realization plans \( q_i \) at all stages. Indeed, the realization plan \( q_i(h^A_t; k) \) is the conditional probability \( \mathbb{P}[h^A_t | k] \).

The set of realization plans of the informed player is denoted by \( Q \), including all properly dimensioned real vectors satisfying (4-6).

A very important difference between a one-shot game and a repeated game is that the uninformed player can learn the game state from the informed player’s history actions. The informed player can characterize his revelation of information by the posterior probability \( \mathbb{P}[k | h^A_t] \), which is also called the belief. Let \( p_t \in \Delta(K) \) denote the posterior probability over the system state \( k \in K \) at stage \( t \) given \( h^A_t \), i.e., \( p_t^k(h^A_t) = \mathbb{P}[k|h^A_t] \). The belief \( p_{t+1} \) at stage \( t + 1 \) can be computed recursively as a function of \( p_t \), the informed player’s strategy \( x^t_k \), and the informed player’s realized action \( a_t \) based on the Bayesian law. Therefore, we have

\[ p_{t+1}^k(h^A_{t+1}) = \mathbb{P}[k|p_t, x_t, a_t] = \frac{p_t^k(h^A_t) x^t_k(a_t)}{\bar{F}_{x_t}(a_t)}, \quad (7) \]

with \( p_1 = p_0 \) in game \( \Gamma_N(p_0) \). Here, \( x^t_k(a_t) = \sigma_{x_t}^A (k, h^A_t) \), and \( \bar{F}_{x_t}(a_t) = \sum_{k \in K} p_t^k(h^A_t) x^t_k(a_t) \). Based on the belief, the value function \( V_N(p) \) satisfies a backward recursive equation which is similar to the Bellman’s equation [6], [11].

\[ V_i(p) = \max_{x \in \Delta(A)^K} \min_{y \in \Delta(B)} \sum_{k \in K} p^k x^T M^k y + \sum_{a \in \mathcal{A}} x_{p, a}(a) V_{i-1}(\pi(p, x, a)). \quad (8) \]

Based on the realization plan \( q \) and the backward recursive formula (8), we construct a linear program to compute the security strategy for the informed player.

Theorem 3. Consider a two-player zero-sum \( N \)-stage asymmetric repeated game \( \Gamma_N(p) \). The game value \( V_N(p) \) satisfies

\[ V_N(p) = \max_{q, \ell \in \mathcal{Q}, \ell} \sum_{i=0}^{N} \sum_{h^A_i \in H^A_i} \ell_{h^A_i} \quad (9) \]

\[ s.t. \quad \sum_{k \in K, a \in \mathcal{A}} p^k q_{i+1}(h^A_{i+1}; a) M_{a, \ell} \geq \ell_{h^A_i} 1^T, \]

\[ \forall t = 1, \ldots, N, \forall h^A_t \in H^A_i. \quad (10) \]

where \( Q \) is a set including all properly dimensioned real vectors satisfying (4-6), \( L \) is a properly dimensioned real
space, and \((h_A, a)\) indicates concatenation. The informed player’s security strategy \(\sigma^*\) is
\[
\sigma^*_t(k, h_A^t) = q_{t+1}^k((h_A^t, a); k)/q_t^k(h_A^t; k), \forall a \in A. \tag{11}
\]

**Proof:** By the strong duality theorem\(^1\), it is easy to see that equation (9-10) is true for \(N = 1\). Let’s assume that \(V_{t-1}(p)\) satisfies (9-10) for all \(t = 2, \ldots\). According to Theorem III.1 of [22], we have \(\bar{x}_{p, x}(a)V_{t-1}(\pi(p, x, a_1)) = V_{t-1}(\bar{x}_{p, x}(p, x, a_1))\). Therefore, the second term of (8) satisfies
\[
\sum_{a_1 \in A} x_{p, x}(a)V_{t-1}(\pi(p, x, a_1)) = \sum_{a_1 \in A} \max_{q, e_1 \in Q, e_1, \sum_{s' = 1}^{t-1} \sum_{h_A^s \in H_A^s}} \ell_{h_A^s}(a_1, h_A^s) \text{ s.t. } p^k q_{s+1}((h_A^s, a); k) \geq \ell_{h_A^s} 1^T, \forall s' = 2, \ldots, t, h_A^s \in H_A^s.
\]

Let \(s = s' + 1\) and \(h_A^s = (a_1, h_A^{s'})\). We have
\[
\sum_{a_1 \in A} x_{p, x}(a)V_{t-1}(\pi(p, x, a_1)) = \max_{q, e_1 \in Q, e_1, \sum_{s = 2}^{t} \sum_{h_A^s \in H_A^s}} \ell_{h_A^s} \text{ s.t. } p^k q_{s+1}((h_A^s, a); k) \geq \ell_{h_A^s} 1^T, \forall s = 2, \ldots, t, h_A^s \in H_A^s.
\]

By the strong duality theorem, it is easy to verify that
\[
\min_{y \in \Delta(B)} \sum_{k \in K} p^k x^k M^k y \geq \max_{\ell_{h_A^s} \in \mathbb{R}} \ell_{h_A^s} \text{ s.t. } \sum_{k \in K} p^k x^k M^k \geq \ell_{h_A^s} 1^T.
\]

According to equation (8), and with the fact that \(x^k(a_1) = q_2(a_1; k)\), we show that equation (9-10) still holds for \(V_{t}(p)\) for \(t = 2, \ldots\).

Once we get the optimal solution \(q^*\), according to (3), the security strategy of the informed player can be computed according to (11).

Our LP formulation of informed player’s security strategy has its size linear in the size of the state set and the size of uninformed player’s action set, polynomial in the size of informed player’s action set, and exponential in time horizon. Let’s first analyze the variable size. Variable \(q\) consists of \((q_t)_{t=1}^{N+1}\), where \(q_t\) is of size \(|K| \times |H_A^t| = |K| \times |A_t|\), and hence \(q\) consists of \(|K|(|1 + |A| + \cdots + |A|^N|) = O(|K||A|^{N+1})\) scalars. Variable \(\ell\) consists of \((1 + |A| + \cdots + |A|^N) = O(|A|^N)\) scalars. In all, we see that the LP formulation has \(|K||A|^N\) scalar variables. Next, let’s take a look at the constraint size. Constraint (4) includes \(|K|\) equations. Constraint (5) includes \(|K|(|1 + |A| + \cdots + |A|^N|) = O(|K||A|^N)\) equations. Constraint (6) includes \(|K|(|1 + |A| + \cdots + |A|^N|) = O(|K||A|^{N+1})\) equations. Constraint (10) includes \(|B|(|1 + |A| + \cdots + |A|^N|) = O(|B||A|^N)\) equations. In all, there are \(|K||A|^N\) equations. Therefore, the size of the LP formulation to compute the informed player’s security strategy is linear in \(|K|\) and \(|B|\), polynomial in \(|A|\), and exponential in \(N\).

Next, let’s take a look at the uninformed player’s security strategy. Define the conditional expected total payoff \(u(\bar{\tau}; k, h_{N+1}^{A})\) given uninformed player’s strategy \(\bar{\tau} \in \mathcal{T}\), state \(k \in K\), and informed player’s history action sequence \(h_{N+1}^{A} \in H_{N+1}^{A}\) as
\[
u(k, h_{N+1}^{A}) = \min_{\ell_{h_A^s} \in \mathbb{R}} \sum_{t=1}^{N} M(k, a_t, b_t)[k, h_{N+1}^{A}] . \tag{12}
\]

It is easy to show that
\[
u(k, h_{N+1}^{A}) = \sum_{t=1}^{N} M_{a_t, b_t} h_{N+1}^{A} . \tag{13}
\]

where \(h_{N+1}^{A} = \bar{\tau}_t(h_{t}^{A}), a_t, h_t^{A} \in H_{t+1}^{A}\). We notice that \(u(\bar{\tau}; k, h_{N+1}^{A})\) is a linear function of \(\bar{\tau}\), or in other words, \(y\).

**Theorem 4.** Consider a two-player zero-sum N-stage asymmetric repeated game \(V_N(p)\). The game value \(V_N(p)\) satisfies
\[
V_N(p) = \min_{y \in Y, \bar{\tau} \in \mathcal{T}} p^T \ell \text{ s.t. } u(y; k, \bar{\tau}) \leq \ell_k 1, \forall k \in K, 1^T y h_{t}^{A} = 1, \forall t = 1, \ldots, N, y_{h_{t}^{A}} \geq 0, \forall t = 1, \ldots, N, \ldots, N, (15)
\]

where \(Y\) is a properly dimensioned real space, and \(u(y; k, \bar{\tau}) = (h_{N+1}^{A})\) is a \(|H_{N+1}^{A}|\) dimensional column vector whose element is \(u(y; k, h_{N+1}^{A})\), a linear function of \(y\) satisfying equation (13). The uninformed player’s security strategy \(\sigma^*(h_{t}^{A})\) is \(y^*(h_{t}^{A})\).

\[
\sigma^*(h_{t}^{A}) = \max_{\bar{\tau} \in \mathcal{T}} u_{\bar{\tau}}(h_{t}^{A}) = \max_{\bar{\tau} \in \mathcal{T}} \sum_{t=1}^{N} M_{a_t, b_t} h_{t}^{A} . \tag{18}
\]

According to the strong duality theorem, we have
\[
u_N(\bar{\tau}) = \min_{\ell_k \in \mathbb{R}} \ell_k 1. \tag{19}
\]

\(^1\)Consider a primal LP problem and the corresponding dual LP problem. If solutions to both problem exist, then optimal feasible solutions to both problems exist, and the optimal values of the two problems are equal [21].

\[\text{This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TAC.2019.2933396, IEEE Transactions on Automatic Control} \]

---

\(0018-9286\) (c) 2019 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.
The game value $V_N(p)$ satisfies
\[
V_N(p) = \min_{\tau \in T} \sum_{k \in K} p^k \nu^k(\tau) \\
= \min_{y \in Y, t : \{k\} \subseteq K} \sum_{k \in K} p^k \ell^k \\
s.t. \ u(y; k, \cdot) \leq 1\ell^k, \forall k \in K.
\]

The LP formulation of the uninformed player’s security strategy has its size linear in the size of the state set and its own action set, polynomial in the size of the informed player’s action set, and exponential in time horizon. We first analyze the variable size. Variable $y$ includes $(y_t)_{t=1}^N$, where $y_t$ is of size $|B||A|^{N-1}$, and hence $y$ has $|B|(1 + |A| + \cdots + |A|^{N-1}) = O(|B||A|^N)$ scalar variables. Variable $\ell$ is of size $|K|$. In all, the variable size is in the order of $|B||A|^N + |K|$. We then study the constraint size. Constraint (16) consists of $(1 + |A| + \cdots + |A|^{N-1}) = O(|A|^N)$ equations. Constraint (17) consists of $|B|(1 + |A| + \cdots + |A|^{N-1}) = O(|B||A|^N)$ equations. Constraint (15) consists of $|A|^N|K|$ equations. In all, the constraint size is of order $O((|B| + |K| + 1)|A|^N)$. Therefore, the size of the LP formulation to compute the uninformed player’s security strategy is linear in $|K|$ and $|B|$, polynomial in $|A|$, and exponential in $N$.

III. $\lambda$-Discounted Asymmetric Repeated Games

A. Setup

A two-player zero-sum $\lambda$-discounted asymmetric repeated game is specified by the same five-tuple $(K, A, B, M, p_0)$ and played in the same way as described in the two-player zero-sum $N$-stage asymmetric repeated game. The payoff of player 1, i.e., the penalty of player 2, at stage $t$ is $\lambda(1 - \lambda)^{t-1}M(k, a_t, b_t)$ for some $\lambda \in (0, 1)$, and the game is played for infinite stages. The payoff of the $\lambda$-discounted asymmetric repeated game with initial probability $p_0$ and strategies $\sigma$ and $\tau$ is defined as
\[
\gamma_\lambda(p_0, \sigma, \tau) = E_{p_0, \sigma, \tau}\left(\sum_{t=1}^{\infty} \lambda(1 - \lambda)^{t-1} M(k, a_t, b_t)\right).
\]

The two-player zero-sum $\lambda$-discounted asymmetric repeated game $\Gamma_\lambda(p_0)$ is defined as a two-player zero-sum asymmetric repeated game equipped with initial distribution $p_0$, strategy spaces $\Sigma$ and $T$, and payoff function $\gamma_\lambda(p_0, \sigma, \tau)$. The security strategies $\sigma^*$ and $\tau^*$, and security levels $\underline{V}_\lambda(p_0)$ and $\overline{V}_\lambda(p_0)$ are defined in the same way as in an $N$-stage game in Section II for player 1 and 2, respectively. Since $\gamma_\lambda(p_0, \sigma, \tau)$ is bilinear over $\sigma$ and $\tau$, $\Gamma_\lambda(p_0)$ has a value $V_\lambda(p_0)$ according to Sion’s minimax Theorem, i.e., $V_\lambda(p_0) = \underline{V}_\lambda(p_0) = \overline{V}_\lambda(p_0)$ [11]. Our objective is to find a computationally tractable approximated security strategy with guaranteed performance for every player.

There are two main challenges in $\lambda$-discounted asymmetric repeated games. The first challenge is that security strategies based on history actions need prohibitive memory to record history actions as the time horizons in $\lambda$-discounted games go to infinity. Therefore, it is necessary to find out fixed-sized sufficient statistics for both players to save the memory.

The second challenge is that computing the game values and the security strategies in $\lambda$-discounted games is non-convex. Therefore, we need to find a computationally tractable approximated security strategy with guaranteed performance for each player.

The first challenge is fully addressed for the informed player in prior work [6]. For the uninformed player, this challenge is partially addressed in prior work [10] which provides a recursive formula to compute the uninformed player’s sufficient statistic. But the physical meaning of the initial condition and a tractable computation method are missing. A main contribution of this section is to show that the initial condition of the uninformed player’s sufficient statistic is the difference between 0 and the worst case payoff of the uninformed player’s security strategy given every possible game state, and providing an LP formulation to compute an approximated initial condition.

The second challenge is still an open problem for both players. The other main contribution of this section is providing both players with approximated security strategies whose guaranteed performances converge to the game value exponentially fast, and constructing LP formulations to compute the approximated security strategies of both players.

B. The informed player

1) The informed player’s security strategy: The belief $p_t$ in (7) plays an important role in decision making of the informed player. Indeed, in a $\lambda$-discounted asymmetric repeated game $\Gamma_\lambda(p)$, the belief state $p_t$ is sufficient statistic of the informed player.

Proposition 5 ([11]). Consider a two-player zero-sum $\lambda$-discounted asymmetric repeated game $\Gamma_\lambda(p)$. The game value $V_\lambda(p)$ satisfies
\[
V_\lambda(p) = \max_{x \in \Delta(A)^{|K|}} \min_{y \in \Delta(B)} \left(\lambda \sum_{k \in K} p^k \bar{x}^k M_k y + (1 - \lambda)T_{p,x}(V_\lambda)\right),
\]
where
\[
T_{p,x}(V_\lambda) = \sum_{a \in A} \bar{x}^a_{p,x} V_\lambda(\pi(p, x, a)).
\]

Moreover, the informed player has a security strategy that depends only on the belief state $p_t$ at each stage $t$, and is independent of the uninformed player’s history actions.

First of all, Proposition 5 points out that the informed player’s security strategy is independent of the uninformed player’s history action, just as what it is in an $N$-stage game. Following the same steps, we can show that the uninformed player’s best response to an $H^B$ independent strategy is also $H^B$ independent. Second, Proposition 5 provides the sufficient statistic $p_t$ of the informed player. So the informed player only needs to record $p_t \in \Delta(K)$ instead of all of its own history actions. Finally, given the belief state $p_t$, Proposition 5 gives a Bellman-like equation (20) to compute the informed player’s security strategy.
Unfortunately, computing the value \( V_\lambda(p) \) and the informed player’s corresponding security strategy \( \sigma^* \) is non-convex [14], [15]. Therefore, we need to find an approximated security strategy that is easy to compute, and has some performance guarantee.

2) The informed player’s approximated security strategy: One way to approximate the security strategy is to approximate the game value \( V_\lambda(p) \) first, and then compute the security strategy based on the approximated game value. Here, we will use the game value \( V_{\lambda,N}(p) \) of a \( \lambda \)-discounted \( N \)-stage asymmetric repeated game \( \Gamma_{\lambda,N}(p) \) to approximate the game value \( V_\lambda(p) \).

A \( \lambda \)-discounted \( N \)-stage repeated asymmetric information game \( \Gamma_{\lambda,N}(p_0) \) is a truncated version of a \( \lambda \)-discounted asymmetric repeated game \( \Gamma_{\lambda,N}(p_0) \) with time horizon \( N \). To be more specific, a \( \lambda \)-discounted \( N \)-stage asymmetric repeated game \( \Gamma_{\lambda,N}(p_0) \) is specified by the same five-tuple \((K, A, B, M, p_0)\) and played in the same way as in a \( \lambda \)-discounted asymmetric repeated game \( \Gamma_{\lambda}(p_0) \). The one-stage payoff of player 1, i.e., the one-stage penalty of player 2, is \( \lambda(1 - \lambda)^{t-1} M(k, a_t, b_t) \), the same as the one-stage payoff in \( \Gamma_{\lambda}(p_0) \). The only difference between a \( \lambda \)-discounted \( N \)-stage asymmetric information game \( \Gamma_{\lambda,N}(p_0) \) and a \( \lambda \)-discounted repeated asymmetric information game \( \Gamma_{\lambda}(p_0) \) is that the game is played for \( N \) stages in \( \Gamma_{\lambda,N}(p_0) \), and infinite stages in \( \Gamma_{\lambda}(p_0) \). Therefore, the payoff of game \( \Gamma_{\lambda,N}(p_0) \) is

\[
\gamma_{\lambda,N}(p_0, \sigma, \tau) = E_{p_0, \sigma, \tau} \left( \sum_{t=1}^{N} \lambda(1 - \lambda)^{t-1} M(k, a_t, b_t) \right).
\]

A \( \lambda \)-discounted \( N \)-stage repeated asymmetric information game \( \Gamma_{\lambda,N}(p_0) \) is defined as a two-player zero-sum repeated asymmetric information game equipped with initial probability \( p_0 \), strategy spaces \( \Sigma \) and \( \mathcal{T} \), and payoff function \( \gamma_{\lambda,N}(p_0, \sigma, \tau) \).

Following the standard arguments as in the proof of Proposition 5, we see that the game value \( V_{\lambda,N+1}(p) \) of the \( \lambda \)-discounted \( N \)-stage game \( \Gamma_{\lambda,N+1}(p) \) satisfies the following recursive formula.

\[
V_{\lambda,N+1}(p) = \max_{x \in \Delta(A)^{|K|}} \min_{y \in \Delta(B)} \left( \sum_{k \in K} p_{k} x^{k} M k y + (1 - \lambda) T_{p,x}(V_{\lambda,N}) \right), \tag{22}
\]

with \( V_{\lambda,0}(p) = 0 \).

Before we go ahead to provide the approximated security strategy based on this approximated game value, we are interested in how good the approximated game value is, and how fast it converges to the real game value. To this purpose, we define an operator \( F_{x} \) as

\[
F_{x}^{V}(p) = \min_{y \in \Delta(B)} \left( \sum_{k \in K} p_{k} x^{k} M k y + (1 - \lambda) T_{p,x}(V) \right). \tag{23}
\]

It’s clear that \( V_{\lambda}(p) = \max_{x \in \Delta(A)^{|K|}} F_{x}^{V}(p) \), and \( V_{\lambda,N+1}(p) = \max_{x \in \Delta(A)^{|K|}} F_{x}^{V,N}(p) \). The operator \( F_{x} \) is actually a contraction mapping.

**Lemma 6.** Let \( V \) be the set of functions mapping from \( \Delta(K) \) to \( \mathbb{R} \). Given any \( x \in \Delta(A)^{|K|} \) and \( \lambda \in (0, 1) \), the operator \( F_{x} : V \to V \) defined in (23) is a contraction mapping with contraction constant \( 1 - \lambda \), i.e.,

\[
\| F_{x}^{V} - F_{x}^{V} \|_{\sup} \leq (1 - \lambda) \| V_{1} - V_{2} \|_{\sup}, \forall V_{1}, V_{2} \in V.
\]

**Proof:** Since the second term of mapping \( F_{x} \) in equation (23) is irrelevant to \( y \), we have

\[
F_{x}^{V}(p) = \min_{y \in \Delta(B)} \left( \sum_{k \in K} p_{k} x^{k} M k y + (1 - \lambda) T_{p,x}(V_{1}) \right),
\]

\[
F_{x}^{V}(p) = \min_{y \in \Delta(B)} \left( \sum_{k \in K} p_{k} x^{k} M k y + (1 - \lambda) T_{p,x}(V_{2}) \right).
\]

Therefore, according to the definition of \( T \) in (21),

\[
\| F_{x}^{V} - F_{x}^{V} \|_{\sup} \leq (1 - \lambda) \| V_{1} - V_{2} \|_{\sup} \sum_{a \in A} \| x_{p,a}(\pi(a; p, x)) \| - (1 - \lambda) \| V_{1} - V_{2} \|_{\sup} \sum_{a \in A} \| x_{p,a}(\pi(a; p, x)) \| = (1 - \lambda) \| V_{1} - V_{2} \|_{\sup}
\]

Its supreme norm, hence, satisfies

\[
\| F_{x}^{V} - F_{x}^{V} \|_{\sup} \leq (1 - \lambda) \| V_{1} - V_{2} \|_{\sup} \sum_{a \in A} \| x_{p,a}(\pi(a; p, x)) \|.
\]

**Theorem 7.** Given \( \lambda \in (0, 1) \), the approximated game value \( V_{\lambda,N+1} \) converges to \( V_{\lambda} \) exponentially fast with rate \( 1 - \lambda \), i.e.,

\[
\| V_{\lambda} - V_{\lambda,N+1} \|_{\sup} \leq (1 - \lambda) \| V_{\lambda} - V_{\lambda,N} \|_{\sup} \tag{24}
\]

\[
\leq (1 - \lambda)^{N+1} \| V_{\lambda} \|_{\sup}. \tag{25}
\]

**Proof:** From equation (20) and (22), we have

\[
\| V_{\lambda}(p) - V_{\lambda,N+1}(p) \| \leq \max_{x \in \Delta(A)^{|K|}} F_{x}^{V,N}(p) - \max_{x \in \Delta(A)^{|K|}} F_{x}^{V,N}(p).
\]

Let \( x^* \) and \( x^* \) be the solution to \( \max_{x \in \Delta(A)^{|K|}} F_{x}^{V}(p) \) and \( \max_{x \in \Delta(A)^{|K|}} F_{x}^{V,N}(p) \), respectively.

Given \( p \in \Delta(K) \), if \( V_{\lambda}(p) \geq V_{\lambda,N+1}(p) \), we have

\[
\| V_{\lambda}(p) - V_{\lambda,N+1}(p) \| \leq \max_{x \in \Delta(A)^{|K|}} F_{x}^{V,N}(p) - \max_{x \in \Delta(A)^{|K|}} F_{x}^{V,N}(p) \leq (1 - \lambda) \| V_{\lambda} - V_{\lambda,N} \|_{\sup}.
\]

Given \( p \in \Delta(K) \), if \( V_{\lambda}(p) \leq V_{\lambda,N+1}(p) \), we have

\[
\| V_{\lambda}(p) - V_{\lambda,N+1}(p) \| \leq \max_{x \in \Delta(A)^{|K|}} F_{x}^{V,N}(p) - \max_{x \in \Delta(A)^{|K|}} F_{x}^{V,N}(p) \leq (1 - \lambda) \| V_{\lambda} - V_{\lambda,N} \|_{\sup}.
\]

Therefore, we have for any \( p \in \Delta(K) \), \( \| V_{\lambda}(p) - V_{\lambda,N+1}(p) \| \leq (1 - \lambda) \| V_{\lambda} - V_{\lambda,N} \|_{\sup} \), which further implies equation (24) and (25).
In λ-discounted game $\Gamma_\lambda(p)$, $\bar{\sigma}_{\lambda,N} : K \times \Delta(K) \rightarrow \Delta(A)$ indicates the informed player’s stationary strategy that is computed based on the approximated game value $V_{\lambda,N}$, and satisfies the following formula.

$$\bar{\sigma}_{\lambda,N}(.;p) = \underset{x \in \Delta(A)}{\arg \max} \min_{y \in \Delta(B)} \left( \lambda \left( \sum_{k \in K} p^k x^k T^k y \right) + (1 - \lambda) T_{p,x}(\bar{\sigma}_{\lambda,N}) \right),$$

(26)

where $\bar{\sigma}_{\lambda,N}(.;p)$ is a $|A| \times |K|$ matrix whose $k$th column is $\bar{\sigma}_{\lambda,N}(k;p)$. Clearly, $\bar{\sigma}_{\lambda,N}(.;p)$ can be also seen as player 1’s security strategy at stage 1 in the λ-discounted asymmetric repeated game $\Gamma_{\lambda,N+1}(p)$. Following the same steps as in Theorem 3, we can construct a linear program to compute the approximated game value $V_{\lambda,N+1}(p)$ and the corresponding approximated strategy $\bar{\sigma}_{\lambda,N}(k;p)$.

**Theorem 8.** Consider a two-player zero-sum λ-discounted asymmetric information game $\Gamma_\lambda(p)$. The approximated game value $V_{\lambda,N+1}(p)$ satisfies

$$V_{\lambda,N+1}(p) = \max_{q,t \in Q,L} \sum_{t=1}^{N+1} \sum_{k \in K} \lambda(1 - \lambda)^{t-1} q_t(k, (h_t^A, a_t)) M_{k,a_t} \geq h_t^A T^t,$$

s.t. $q_t(k, (h_t^A, a_t)) M_{k,a_t} \geq h_t^A T^t,$

$$\forall t = 1, 2, \ldots, N + 1, h_t^A \in H_t^A,$$

(27)

where $q \in Q$ is a set including all properly dimensioned real vectors satisfying (4-6), $L$ is a properly dimensioned real space, and $(h_t^A, a_t)$ corresponds to concatenation. The approximated security strategy

$$\bar{\sigma}_{\lambda,N}^A(k;p) = q_2^t(a;k), \forall a \in A.$$  

(29)

3) The performance analysis of the informed player’s approximated security strategy: Now that we can compute the informed player’s approximated security strategy, the next question is which performance this strategy can guarantee. To this purpose, we first define the security level $J^\gamma_{\lambda,N}(p)$ guaranteed by the approximated security strategy $\bar{\sigma}_{\lambda,N}$ as

$$J^\gamma_{\lambda,N}(p) = \min_{\bar{\tau} \in \bar{\Sigma}} \gamma_{\lambda}(p, \bar{\sigma}_{\lambda,N}; \bar{\tau}).$$

(30)

Since $\bar{\sigma}_{\lambda,N}$ is a stationary strategy, according to the standard procedure of dynamic programming, its security level $J^\gamma_{\lambda,N}$ has the following property.

**Lemma 9.** Let $\bar{\sigma} \in \bar{\Sigma}$ be the informed player’s stationary strategy that depends only on the belief state $\bar{\tau}_t$ besides the state $k \in K$. The security level $J^\gamma$ of $\bar{\sigma}$ satisfies $J^\gamma(p) =$ $F^\gamma_{\bar{\sigma}(.;p)}(p)$.

**Proof:** Since player 1’s strategy is fixed to be $\bar{\sigma}$, the discounted game $\Gamma_\lambda$ becomes a discounted optimization problem, and hence satisfies Bellman’s principle, i.e.,

$$J^\gamma(p) = \min_{y \in \Delta(B)} \left( \lambda \left( \sum_{k \in K} p^k \bar{\sigma}(k,p)^T M^k y \right) + (1 - \lambda) T_{p,x}(\bar{\sigma}(.;p)) \right)$$

$$= \min_{y \in \Delta(B)} \left( \lambda \sum_{k \in K} p^k \bar{\sigma}(k,p)^T M^k y \right) + (1 - \lambda) T_{p,x}(\bar{\sigma}(.;p)(J^\gamma))$$

$$= F^\gamma_{\bar{\sigma}(.;p)}(p).$$

Now, we are ready to show that the difference between the approximated security strategy’s security level $J^\gamma_{\lambda,N}$ and the game value is bounded from above, which is stated in the following theorem.

**Theorem 10.** The security level $J^\gamma_{\lambda,N}$ of the informed player’s approximated security strategy $\bar{\sigma}_{\lambda,N}$ defined in equation (26) converges to the game value exponentially fast in $N$ with rate $1 - \lambda$. To be more specific,

$$\|V_\lambda - J^\gamma_{\lambda,N}\|_{\sup} \leq \frac{2(1 - \lambda)}{\lambda} \|V_\lambda - V_{\lambda,N}\|_{\sup}.$$  

(31)

$$\leq (1 - \lambda)^{N+1} 2\|V_\lambda\|_{\sup}.$$  

(32)

**Proof:** Lemma 9 indicates that

$$\|V_\lambda(p) - J^\gamma_{\lambda,N}(p)\| \leq \|V_\lambda(p) - V_{\lambda,N+1}(p)\| + \|V_{\lambda,N+1}(p) - J^\gamma_{\lambda,N}(p)\|$$

$$= \|V_\lambda(p) - V_{\lambda,N+1}(p)\| + \|F_{V_{\lambda,N}} - F_{J^\gamma_{\lambda,N}}(p)\| + \|F_{J^\gamma_{\lambda,N}}(p)\|.$$

Take the supreme norm on both sides, and use Lemma 6 and Theorem 7. We have

$$\|V_\lambda - J^\gamma_{\lambda,N}\|_{\sup} \leq (1 - \lambda) \left( \|V_\lambda - V_{\lambda,N}\|_{\sup} + \|V_{\lambda,N} - J^\gamma_{\lambda,N}\|_{\sup} \right)$$

(33)

$$\leq (1 - \lambda) \left( \|V_\lambda - V_{\lambda,N}\|_{\sup} + \|V_{\lambda,N} - V_{\lambda,N}\|_{\sup} + \|V_{\lambda,N} - J^\gamma_{\lambda,N}\|_{\sup} \right),$$

which implies equation (31). Together with Theorem 7, equation (32) is shown.

Notice that the difference between the game value and the security level of the approximated security strategy $\bar{\sigma}_{\lambda,N}$ decreases exponentially fast with respect to $N$, and converges to 0 as $N$ goes to infinity.

Based on how the approximated security strategy of the informed player is derived, we provide an algorithm to compute the informed player’s approximated security strategy as follows.

**Algorithm 1.** The informed player’s algorithm to compute $\bar{\sigma}_{\lambda,N}$ in λ-discounted asymmetric repeated game

(i) Initialization

- Read payoff matrices $M$, initial probability $p_0$, and system state $k$.
- Set $N$. 


Let $t = 1$, and $p_1 = p_0$.

(ii) Compute the informed player’s approximated security strategy $\bar{\sigma}_{\lambda,N}$ based on (29) where $q_y^2$ is the optimal solution of LP (27-28) with $p = p_t$.

(iii) Choose an action $a \in A$ according to the probability $\bar{\sigma}_{\lambda,N}(k,p)$, and announce it publicly.

(iv) Update the belief state $p_{t+1}$ according to (7).

(v) Update $t = t + 1$ and go to step (ii).

C. The uninformed player

Because of the lack of access to the informed player’s strategy, the belief state $p_t$ is not available to the uninformed player, and hence can not serve as the uninformed player’s sufficient statistic. De Meyer first introduced the dual game of player 1 (informed player) at stage 1. In the dual game, Player 2 (uninformed player) is still not informed of the game state. The one-stage payoff of player 1, i.e., one-stage penalty of player 2, at stage $t$ is the same as in the primal game. Let $p$ be player 1’s strategy to choose the system state, player 1’s payoff or player 2’s penalty in the dual game $\tilde{V}_\lambda (w)$ is defined as

$$\tilde{\gamma}_\lambda(w,p,\sigma,\tau) = E_{p,\sigma,\tau} \left[ w^k + \sum_{t=1}^{T-1} \lambda(1-\lambda)^{t-1} M_{a_t,b_t} [k, a_t, b_t] \right].$$

The dual game $\tilde{V}_\lambda (w)$ is defined as a $\lambda$-discounted asymmetric repeated dual game equipped with initial regret $w$, strategy spaces $\Delta(K) \times \Sigma$ and $T$, and payoff function $\tilde{\gamma}_\lambda(w,p,\sigma,\tau)$.

The $\lambda$-discounted asymmetric repeated dual game $\tilde{V}_\lambda (w)$ has a game value denoted by $\bar{V}_\lambda (w)$ satisfying [10]

$$\bar{V}_\lambda (w) = \min_{\sigma} \max_{\tau} \bar{\gamma}_\lambda(w,p,\sigma,\tau) = \max_{\sigma} \min_{\tau} \bar{\gamma}_\lambda(w,p,\sigma,\tau).$$

The game values of the dual game and the primal game have the following relation [10], [11].

$$\tilde{V}_\lambda (w) = \max_{p \in \Delta(K)} \left\{ \tilde{V}_\lambda (p) + p^T w \right\},$$

$$\tilde{V}_\lambda (p) = \min_{w \in \mathbb{R}|K|} \left\{ \tilde{V}_\lambda (w) - p^T w \right\}. $$

It was shown that the security strategies of the uninformed player in both the primal and the dual games depend only on the informed player’s history actions [10], [11]. Following the same steps as in Proposition 2, we can show that the informed player’s best responses to an $H^B$ independent strategy in both the primal and the dual games are also $H^B$ independent. Therefore, we only consider $H^B$ independent strategies for both players in the rest of this subsection.

Let’s define the anti-discounted regret $w^k_t$ at stage $t$ with respect to state $k$ given informed player’s history action sequence $h^A_t$ as

$$w^k_t(h^A_t) = \frac{E_{x} \left[ w^k + \sum_{s=1}^{t-1} \lambda(1-\lambda)^{s-1} M_{a_s, b_s} [k, h^A_s] \right]}{(1-\lambda)^t-1}, \forall k \in K.$$ 

The anti-discounted regret $w^k_t(h^A_t)$ can be computed recursively as

$$w^k_{t+1}(h^A_t, a_t) = \frac{w^k_t(h^A_t) + \lambda M_{a_t, h^A_t} - \bar{\tau}(h^A_t)}{1-\lambda}, \forall k \in K, \tag{37}$$

with $w_1 = w$. The anti-discounted regret $w_t$ is indeed the sufficient statistic for the uninformed player in the dual game.

Proposition 11. [10], [11] The value $\bar{V}_\lambda (w)$ of the $\lambda$-discounted asymmetric repeated dual game $\tilde{V}_\lambda (w)$ satisfies

$$\bar{V}_\lambda (w) = \min_{y \in \Delta(B)} \max_{a \in A} (1-\lambda)\bar{\gamma}_\lambda \left( w + \lambda M_{a,b} - \bar{\tau} \right),$$

where $M_a$ is a $|K| \times |B|$ matrix whose kth row is $M_{a,k}$. Moreover, the uninformed player has a security strategy that depends only at each stage $t$ on $w_t$.

Meanwhile, it was also shown in [10], [11] that the security strategy of the uninformed player in the dual game $\tilde{V}_\lambda (w^*)$ with some special initial regret $w^*$ is also the security strategy for the uninformed player in the primal game $\Gamma_\lambda (p)$.

Proposition 12. (Corollary 2.10 and 3.25 in [11]) Consider a $\lambda$-discounted asymmetric repeated game $\Gamma_\lambda (p)$ and its dual game $\tilde{V}_\lambda (w)$. Let $w^*$ be the optimal solution to the optimal problem in (36). A security strategy of the uninformed player in the dual game $\tilde{V}_\lambda (w^*)$ is a security strategy of the uninformed player in the primal game $\Gamma_\lambda (p)$.

Proposition 12 indicates that given an initial probability in the primal game, there exists an initial regret in the dual game such that the security strategies of the uninformed player in the primal and the dual games are the same. Therefore, when playing the primal game $\Gamma_\lambda (p)$, the uninformed player can find out the corresponding initial regret $w^*$ in the dual game first, and then play the dual game instead.
2) The special initial regret $w^*$ and its approximation: The next question is what the special initial regret $w^*$ is. Mathematically speaking, $w^*$ is an optimal solution to the problem $\min_{w \in \mathbb{R}^{|K|}} \{ \tilde{V}_\lambda(w) - p^T w \}$. We are curious about the physical meaning of $w^*$, i.e., what exactly $w^*$ stands for in the primal game $\Gamma_\lambda(p)$. To this purpose, let’s first define the uninformed player’s worst case regret $\mu_\lambda \in \mathbb{R}^{|K|}$ of strategy $\bar{\tau} \in \bar{T}$ in the primal game as

$$\mu^k_\lambda(\tau) = -\max_{\bar{\sigma}(k) \in \bar{S}(k)} \mathbf{E}_{p, \bar{\sigma}, \bar{T}} \left[ \sum_{t=1}^{\infty} \lambda(1 - \lambda)^{t-1} M^k_{\bar{a}_t, b_t} | \right] ,$$

(39)

where $\bar{\sigma}(k)$ indicates informed player’s $H^B$ independent behavior strategy if the system state is $k \in K$, and $\bar{S}(k)$ is the corresponding set including all $\sigma(k)$.

The special initial regret $w^*$ is actually the uninformed player’s worst case regret of his security strategy.

**Theorem 13.** Consider a two-player zero-sum $\lambda$-discounted asymmetric repeated game $\Gamma_\lambda(p)$. Let $\tau^*$ be the uninformed player’s security strategy in $\Gamma_\lambda(p)$. An optimal solution $w^*$ to the optimal problem $\min_{w \in \mathbb{R}^{|K|}} \{ \tilde{V}_\lambda(w) - p^T w \}$ is $w^* = \mu_\lambda(\tau^*)$, i.e.,

$$\min_{w \in \mathbb{R}^{|K|}} \{ \tilde{V}_\lambda(w) - p^T w \} = \tilde{V}_\lambda(\mu_\lambda(\tau^*)) = -p^T \mu_\lambda(\tau^*).$$

(40)

Proof: Equation (36) shows that the left hand side of (40) equals to $V_\lambda(p)$. We will show that the right hand side of (40) equals to $V_\lambda(p)$, too.

First, since $\tau^*$ is the uninformed player’s security strategy in the primal game, its worst case performance meets the game value, i.e.,

$$V_\lambda(p) = \max_{\sigma \in \Sigma} \gamma_\lambda(p, \sigma, \tau^*) = -p^T \mu_\lambda(\tau^*).$$

(41)

The last equality is derived from equation (39).

Next, we show that

$$\tilde{V}_\lambda(\mu_\lambda(\tau^*)) = 0.$$ 

(42)

Equation (35) implies that $\tilde{V}_\lambda(\mu_\lambda(\tau^*)) \geq V_\lambda(p) + p^T \mu_\lambda(\tau^*) = 0$, which is derived from (41).

Meanwhile, for any $p' \in \Delta(K)$, we have

$$V_\lambda(p') = \min_{\tau \in \bar{T}} \max_{\sigma \in \Sigma} \gamma_\lambda(p', \sigma, \tau) \leq \max_{\sigma \in \Sigma} \gamma_\lambda(p', \sigma, \tau^*) = -p'^T \mu_\lambda(\tau^*)$$

(43)

where the last equality is derived from equation (41). Notice here that $\tau^*$ is the uninformed player’s security strategy in $\Gamma_\lambda(p)$, and is not necessarily the uninformed player’s security strategy in $\Gamma_\lambda(p')$. The fact that equation (43) holds for any $p' \in \Delta(K)$ implies that $V_\lambda(p') + p'^T \mu_\lambda(\tau^*) \leq 0$.

Therefore, equation (42) is true. Together with (41), we show that the right hand side of (40) equals to $V_\lambda(p)$, which completes the proof.

The uninformed player’s worst case regret $w^*$ of its security strategy can be seen as the dual variable of the initial probability $p$. The production of the two variables recovers the opposite of the game value (see equation (41)). While the informed player’s security strategy depends only on $p$ and its Bayesian update $p_t$, the uninformed player can fully rely on $w^*$ and its anti-discounted update $w_t$ to generate its security strategy. Moreover, the belief state $p_t$ is fully available to the informed player, while the anti-discounted regret $w_t$ is fully available to the uninformed player.

Theorem 13 characterizes the physical meaning of the special initial regret $w^*$. The next question is how to compute it. Unfortunately, computing $w^*$ is difficult, since it relies on the security strategy of the uninformed player and the game value in the primal game, which is non-convex [15]. Therefore, we propose to approximate $w^*$ based on $\lambda$-discounted asymmetric repeated game $\Gamma_{\lambda,N}(p)$, a truncated version of the primal game $\Gamma_\lambda(p)$. Let $\tau^* \in \bar{T}$ be the security strategy for the uninformed player in $\Gamma_{\lambda,N}(p)$. The approximation $w^*$ of the special initial regret is $\mu_{\lambda,N}(\tau^*)$ which is defined as

$$w^* = \mu_{\lambda,N}(\tau^*)$$

(44)

$$=-\max_{\sigma(k) \in \bar{S}(k)} \mathbf{E}_{p, \tau, \tau^*} \left[ \sum_{t=1}^{N} \lambda(1 - \lambda)^{t-1} M^k_{\bar{a}_t, b_t} | \right]$$

(45)

Similarly to the $N$-stage game, in the $\lambda$-discounted $N$-stage game, we define the conditional expected total payoff $u_{\lambda,N}(\tau; k, h_{N+1}^A)$ given uninformed player’s strategy $\tau \in \bar{T}$, state $k \in K$ and informed player’s history action sequence $h_{N+1}^A$ as

$$u_{\lambda,N}(\tau; k, h_{N+1}^A) = \mathbf{E}_{\tau} \left[ \sum_{t=1}^{N} \lambda(1 - \lambda)^{t-1} M^k_{\bar{a}_t, b_t} | k, h_{N+1}^A \right] ,$$

(46)

which satisfies

$$u_{\lambda,N}(\tau; k, h_{N+1}^A) = \sum_{t=1}^{N} \lambda(1 - \lambda)^{t-1} M^k_{\bar{a}_t, b_t} y_{\bar{h}_t^A}.$$ 

(47)

Following the same steps as in Theorem 4, we can construct an LP formulation to compute $V_{\lambda,N}(p)$ and $\mu_{\lambda,N}(\tau^*)$.

**Theorem 14.** Consider a $\lambda$-discounted asymmetric repeated game $\Gamma_\lambda(p)$. The approximated game value $V_{\lambda,N}(p)$ satisfies

$$V_{\lambda,N}(p) = \min_{y \in Y, \bar{\tau} \in \bar{T}} \sum_{k \in K} p^k \ell^k$$

(48)

where $Y$ is a properly dimensioned real space, and $u_{\lambda,N}(y; k, \cdot)$ is a $|H_{N+1}^A|$ dimensional column vector whose element is $u_{\lambda,N}(y; k, h_{N+1}^A)$, a linear function of $y$ satisfying equation (47). The approximated regret $w^*$ is $-\ell^\tau$.

3) The Uninformed player’s approximated security strategy: Now that the approximated initial regret $w^*$ for the dual game $\Gamma_\lambda(w^*)$ is computed, the next step is to compute the uninformed player’s security strategy in the dual game, which is again non-convex [15]. Similar to what we do in approximating the informed player’s security strategy, we use the
game value of a \( \lambda \)-discounted \( N \)-stage dual game \( \Gamma_{\lambda,N}(w^*) \) to approximate \( \tilde{V}_{\lambda}(w^*) \), and derive the uninformed player’s approximated security strategy based on the approximated game value.

A two-player zero-sum \( N \)-stage \( \lambda \)-discounted asymmetric repeated dual game \( \Gamma_{\lambda,N}(w) \) is a truncated version of a \( \lambda \)-discounted asymmetric repeated dual game \( \Gamma_{\lambda}(w) \) with time horizon \( N \), and is specified by a five tuple \((K, A, B, M, w)\) which is defined in the same way as in the infinite stage dual game \( \Gamma_{\lambda}(w) \). The playing rule and the one-stage payoff of player 1, i.e., one-stage penalty of player 2, is the same as in the infinite stage dual game \( \Gamma_{\lambda}(w) \). The only difference between \( \Gamma_{\lambda,N}(w) \) and \( \Gamma_{\lambda}(w) \) is that the dual game is played for \( N \)-stages in \( \Gamma_{\lambda,N}(w) \) and infinite stages in \( \Gamma_{\lambda}(w) \). The payoff function \( \tilde{\gamma}_{\lambda,N}(w, p, \sigma, \tau) \) is defined as

\[
\tilde{\gamma}_{\lambda,N}(w, p, \sigma, \tau) = E_{p, \sigma, \tau} \left[ w + \sum_{i=1}^{N} (1-\lambda)^{i-1} M_{(a_{i}, b_{i})} \right].
\]

A \( \lambda \)-discounted \( N \)-stage asymmetric repeated dual game \( \Gamma_{\lambda,N}(w) \) is defined as a two-player zero-sum \( N \)-stage asymmetric repeated dual game equipped with initial regret \( w \), strategy spaces \( \Delta(K) \times \Sigma \) and \( T \), and payoff function \( \tilde{\gamma}_{\lambda,N}(w, p, \sigma, \tau) \).

Since \( \Gamma_{\lambda,N}(w) \) is a finite game, it has a value denoted by \( \tilde{V}_{\lambda,N}(w) \), i.e.,

\[
\tilde{V}_{\lambda,N}(w) = \min_{p \in P} \max_{\sigma \in \Sigma} \tilde{\gamma}_{\lambda,N}(w, p, \sigma, \tau) = \max_{\tau \in T} \min_{\sigma \in \Sigma} \tilde{\gamma}_{\lambda,N}(w, p, \sigma, \tau)
\]

Following the same steps as in the proof of Proposition 3.23 in [11], we derive that the game value \( \tilde{V}_{\lambda,N+1}(w) \) of dual game \( \Gamma_{\lambda,N+1}(w) \) satisfies the following recursive formula.

\[
\tilde{V}_{\lambda,N+1}(w) = \min_{y \in \Delta(B)} \max_{a \in A} \left( w + \lambda M_{ab} y \right)
\]

with \( \tilde{V}_{\lambda,0}(w) = \max_{k \in K} w^k \). Moreover, since \( \Gamma_{\lambda,N}(w) \) is a dual game of \( \Gamma_{\lambda,N}(p) \), their game values have the following relations.

\[
\tilde{V}_{\lambda,N}(w) = \max_{p \in \Delta(K)} \{V_{\lambda,N}(p) + p^T w\},
\]

\[
\tilde{V}_{\lambda,N}(p) = \min_{w \in \mathbb{R}^{|K|}} \{\tilde{V}_{\lambda,N}(w) - p^T w\}.
\]

Based on the relations between the game values of the \( \lambda \)-discounted game \( \Gamma_{\lambda}(p) \), the \( \lambda \)-discounted \( N \)-stage games \( \Gamma_{\lambda,N}(p) \) and their dual games, we have the following lemmas.

**Lemma 15.** Consider a two-player zero-sum \( \lambda \)-discounted \( N \)-stage asymmetric repeated game \( \Gamma_{\lambda,N}(p) \) and its dual game \( \Gamma_{\lambda,N}(w) \). Let \( w^* \) be defined as in (45). Their game values satisfy

\[
\tilde{V}_{\lambda,N}(w^*) = V_{\lambda,N}(p) + p^T w^*.
\]

**Proof:** According to the definition of \( w^* \) in (45), we have \( V_{\lambda,N}(p) + p^T w^* = 0 \).

Following the same steps as in equation (42), we can show that \( \tilde{V}_{\lambda,N}(w^*) = 0 \), which completes the proof.

**Lemma 16.** Consider a two-player zero-sum \( \lambda \)-discounted asymmetric repeated game \( \Gamma_{\lambda}(p) \) and its dual game \( \Gamma_{\lambda}(w) \), and a two-player zero-sum \( \lambda \)-discounted \( N \)-stage asymmetric repeated game \( \Gamma_{\lambda,N}(p) \) and its dual game \( \Gamma_{\lambda,N}(w) \). Their game values satisfy

\[
\|V_{\lambda} - \tilde{V}_{\lambda,N} \|_{\sup} \leq \|V_{\lambda} - \tilde{V}_{\lambda,N} \|_{\sup}.
\]

**Proof:** First, we show \( \|V_{\lambda} - \tilde{V}_{\lambda,N} \|_{\sup} \leq \|V_{\lambda} - \tilde{V}_{\lambda,N} \|_{\sup} \). According to equation (36) and (57), we have

\[
\|V_{\lambda}(p) - \tilde{V}_{\lambda,N}(p)\|_{\sup} \leq \min_{w \in \mathbb{R}^{|K|}} \{V_{\lambda}(w) - \tilde{V}_{\lambda,N}(w)\}. \]

Let \( w^* \) and \( w^* \) be the optimal solution to the problem \( \min_{w \in \mathbb{R}^{|K|}} \|V_{\lambda}(w) - \tilde{V}_{\lambda,N}(w)\|_{\sup} \) and \( \min_{w \in \mathbb{R}^{|K|}} \|V_{\lambda}(w) - \tilde{V}_{\lambda,N}(w)\|_{\sup} \), respectively. If \( \min_{w \in \mathbb{R}^{|K|}} \|V_{\lambda}(w) - \tilde{V}_{\lambda,N}(w)\|_{\sup} \geq \min_{w \in \mathbb{R}^{|K|}} \|V_{\lambda}(w) - \tilde{V}_{\lambda,N}(w)\|_{\sup} \), then we have \( \|V_{\lambda}(p) - \tilde{V}_{\lambda,N}(p)\|_{\sup} \leq \|V_{\lambda}(w^*) - \tilde{V}_{\lambda,N}(w^*)\|_{\sup} \). Otherwise, we have \( \|V_{\lambda}(p) - \tilde{V}_{\lambda,N}(p)\|_{\sup} \leq \|V_{\lambda}(w^*) - \tilde{V}_{\lambda,N}(w^*)\|_{\sup} \). Therefore, for any \( p \in \Delta(K) \), \( \|V_{\lambda}(p) - \tilde{V}_{\lambda,N}(p)\|_{\sup} \leq \|V_{\lambda} - \tilde{V}_{\lambda,N}\|_{\sup} \), which implies that \( \|V_{\lambda} - \tilde{V}_{\lambda,N}\|_{\sup} \leq \|V_{\lambda} - \tilde{V}_{\lambda,N}\|_{\sup} \).

Following the same steps, based on equation (35) and (56), we derive that \( \|V_{\lambda} - \tilde{V}_{\lambda,N}\|_{\sup} \leq \|V_{\lambda} - \tilde{V}_{\lambda,N}\|_{\sup} \). Therefore, equation (59) is shown.

Before we derive the uninformed player’s approximated security strategy based on the approximated game value \( \tilde{V}_{\lambda,N}(w^*) \), we are interested in how far away the approximated game value is from the real game value. To this purpose, we define an operator \( \tilde{F}_y \) as

\[
\tilde{F}_y(w) = (1 - \lambda) \max_{a \in A} \tilde{V} \left( w + \lambda M_{ab} y \right) \left( 1 - \lambda \right),
\]

where \( y \in \Delta(B) \), \( w \in \mathbb{R}^{|K|} \), and \( \tilde{V} : \mathbb{R}^{|K|} \rightarrow \mathbb{R} \). With the same technique as in Lemma 6, we can show that \( \tilde{F} \) is also a contraction mapping.

**Lemma 17.** Given any \( y \in \Delta(B) \) and \( \lambda \in (0, 1) \), the operator \( \tilde{F}_y \) defined as in (60) is a contraction mapping with contraction constant \( 1 - \lambda \), i.e.,

\[
\|\tilde{F}_y \tilde{V}_1 - \tilde{F}_y \tilde{V}_2\|_{\sup} \leq (1 - \lambda) \|\tilde{V}_1 - \tilde{V}_2\|_{\sup},
\]

where \( \tilde{V}_{1,2} : \mathbb{R}^{|K|} \rightarrow \mathbb{R} \).

Lemma 17 further implies that the approximated value \( \tilde{V}_{\lambda,N} \) converges to the real game value \( V_{\lambda} \) exponentially fast with respect to \( N \). The proof is similar to the proof of Theorem 7.

**Theorem 18.** Consider the game value \( \tilde{V}_{\lambda,N}(w) \) of a \( \lambda \)-discounted asymmetric repeated dual game \( \Gamma_{\lambda}(w) \) and the game value \( \tilde{V}_{\lambda,N}(w) \) of a \( \lambda \)-discounted \( N \)-stage asymmetric repeated dual game \( \Gamma_{\lambda,N}(w) \). The game value \( \tilde{V}_{\lambda,N} \) converges to \( \tilde{V}_{\lambda} \) exponentially fast with respect to the time horizon \( N \) with convergence rate \( 1 - \lambda \), i.e.,

\[
\|\tilde{V}_{\lambda} - \tilde{V}_{\lambda,N+1}\|_{\sup} \leq (1 - \lambda) \|\tilde{V}_{\lambda} - \tilde{V}_{\lambda,N}\|_{\sup}.
\]
Applying the approximated game value $\tilde{V}_{\lambda,N}$ in equation (38), we derive the uninformed player’s approximated security strategy $\bar{\tau}_{\lambda,N}(w_t)$ in dual game $\bar{\Gamma}_\lambda(w)$ as

$$\bar{\tau}_{\lambda,N}(w_t) = \arg\min_{y \in \Delta(B)} \max_{a \in A} (1 - \lambda) \tilde{V}_{\lambda,N} \left( \frac{w_t + \lambda M_a y}{1 - \lambda} \right),$$

(63)

where $w_t$ is updated according to (37). Comparing equation (63) and (55), we see that the approximated security strategy $\bar{\tau}_{\lambda,N}(w_t)$ can be seen as the uninformed player’s security strategy at stage 1 in a $\lambda$-discounted $N + 1$-stage dual game $\bar{\Gamma}_{\lambda,N+1}(w_t)$. Similar to the LP formulation computing the game value of $\Gamma_{\lambda,N}(p)$, we construct an LP formulation to compute the game value of $\bar{\Gamma}_{\lambda,N+1}(w_t)$ and the uninformed player’s approximated security strategy $\bar{\tau}_{\lambda,N}$.

**Theorem 19.** Consider a two-player zero-sum $\lambda$-discounted $N + 1$-stage dual game $\bar{\Gamma}_{\lambda,N+1}(w)$. Its game value $\bar{\tilde{V}}_{\lambda,N+1}(w)$ satisfies

$$\bar{\tilde{V}}_{\lambda,N+1}(w) = \min_{y \in Y, \ell \in [K]} \max_{L \in \mathbb{R}} L$$

$$s.t. w + \ell \leq L \mathbb{1}$$

$$u_{\lambda,N+1}(y; k, \cdot) \leq \ell^k \mathbb{1}, \quad \forall k \in K,$$

$$1^T y_{h_t^k} = 1, \quad \forall t = 1, \ldots, N + 1,$$

$$y_{h_t^k} \geq 0, \quad \forall t = 1, \ldots, N + 1,$$

(64) \hspace{1cm} (65) \hspace{1cm} (66) \hspace{1cm} (67)

where $Y$ is a properly dimensioned real space, and $u_{\lambda,N+1}(y; k, \cdot)$ is a $[H_{N+2}^A]$ dimensional column vector whose element is $u_{\lambda,N+1}(y; k, h^k_{N+2})$, a linear function of $y$ satisfying equation (47).

Moreover, suppose in dual game $\bar{\Gamma}_\lambda(w_0)$, at stage $t$, the anti-discounted regret $w_t = w$. The uninformed player’s approximated security strategy $\bar{\tau}_{\lambda,N}(w)$ is $y_{h_t^k}^\star$.

**Proof:** According to equation (54), we have

$$\bar{\tilde{V}}_{\lambda,N+1}(w) = \min_{\tau \in T} \max_{p \in \Delta(K)} \sum_{k \in K} p^k (w^k - \lambda \mu_{\lambda,N+1}(\tau)).$$

Similar to how we derive equation (18), we have

$$-\mu_{\lambda,N+1}(\tau) = \min_{\ell^k \in \mathbb{R}} \ell^k,$$

$$s.t. u_{\lambda,N+1}(\tau; k, \cdot) \leq \ell^k \mathbb{1}.$$

Therefore, we have

$$\bar{\tilde{V}}_{\lambda,N+1}(w) = \min_{\tau \in T} \max_{p \in \Delta(K)} \min_{\ell^k \in \mathbb{R}} \sum_{k \in K} p^k (w^k + \ell^k),$$

$$s.t. u_{\lambda,N+1}(\tau; k, \cdot) \leq \ell^k \mathbb{1}, \forall k \in K.$$

Since $\sum_{k \in K} p^k (w^k + \ell^k)$ is bilinear in $p$ and $\ell$, according to Sion’s minimax theorem, we have

$$\bar{\tilde{V}}_{\lambda,N+1}(w) = \min_{\tau \in T} \max_{p \in \Delta(K)} \min_{\ell^k \in \mathbb{R}} \sum_{k \in K} p^k (w^k + \ell^k),$$

$$s.t. u_{\lambda,N+1}(\tau; k, \cdot) \leq \ell^k \mathbb{1}, \forall k \in K.$$

According to the strong duality theorem, given any $\bar{\tau} \in \bar{T}$ and $\ell \in \mathbb{R}^{|K|}$, we have

$$\max_{p \in \Delta(K)} \sum_{k \in K} p^k (w^k + \ell^k)$$

$$s.t. u_{\lambda,N+1}(\tau; k, \cdot) \leq \ell^k \mathbb{1}, \forall k \in K,$$

$$\min_{L \in \mathbb{R}}$$

$$s.t. w + \ell \leq L \mathbb{1},$$

$$u_{\lambda,N+1}(\tau; k, \cdot) \leq \ell^k \mathbb{1}, \forall k \in K,$$

which completes the proof.

Now, we know how to compute the approximated special initial regret $w^\star$ and the uninformed player’s approximated security strategy $\bar{\tau}_{\lambda,N}$ in the dual game $\bar{\Gamma}_\lambda(w^\star)$. This approximated security strategy $\bar{\tau}_{\lambda,N}$ is also the uninformed player’s approximated security strategy in the primal game $\Gamma_\lambda(p)$.

According to how the the uninformed player’s approximated security strategy is derived, we provide an algorithm to compute $\bar{\tau}_{\lambda,N}$ in the $\lambda$-discounted asymmetric repeated game $\Gamma_\lambda(p)$.

**Algorithm 2.** The uninformed player’s algorithm to compute $\bar{\tau}_{\lambda,N}$ in $\lambda$-discounted asymmetric repeated game $\Gamma_\lambda(p)$

(i) Initialization

- Read payoff matrices $M$ and initial probability $p_0$.
- Set $N$.
- Solve the LP problem (49-50) with $p = p_0$, and let $w^\star = -\mathbb{1}$.
- Let $t = 1$ and $w_1 = w^\star$.

(ii) Solve the LP problem (64-68) with $w = w_t$ and the uninformed player’s approximated security strategy $\bar{\tau}(w_t)$ is $y_{h_t^k}^\star$.

(iii) Choose an action $b \in B$ according to the probability $\bar{\tau}_{\lambda,N}(w_t)$, and announce it publicly.

(iv) Read the informed player’s action, and update the anti-discounted regret $w_{t+1}$ according to (37).

(v) Update $t = t + 1$ and go to step (ii).

4) The performance difference between the suboptimal strategy and the security strategy: With the uninformed player’s approximated security strategy $\bar{\tau}_{\lambda,N}$, we are interested in the worst case cost guaranteed by this strategy, which is also called the security level of $\bar{\tau}_{\lambda,N}$. Given an uninformed player’s strategy $\tau \in \bar{T}$, the security level $J^\tau(p)$ in game $\Gamma_\lambda(p)$ is defined as

$$J^\tau(p) = \max_{\sigma \in \Sigma} \gamma_\lambda(p, \sigma, \tau).$$

(69)

Since the uninformed player’s approximated security strategy is derived from his approximated security strategy in the dual game, the security levels of the approximated security strategy in the primal and dual games are highly related. Hence, we would also like to define the security level $\bar{J}^\tau(w)$ of $\tau \in \bar{T}$ in the dual game $\bar{\Gamma}_\lambda(w)$ as

$$\bar{J}^\tau(w) = \max_{p \in \Delta(K)} \max_{\sigma \in \Sigma} \bar{\gamma}_\lambda(p, w, \sigma, \tau).$$

(70)

Following the same steps as in the proof of (35-36) in [10], [11], we can show that $J^\tau(p)$ and $\bar{J}^\tau(w)$ have the following
relations.

\[
\hat{J}^\tau(w) = \max_{p \in \Delta(K)} \{ J^\tau(p) + p^T w \}, \quad (71)
\]

\[
J^\tau(p) = \min_{w \in \mathbb{R}^{|K|}} \{ \hat{J}^\tau(w) - p^T w \}. \quad (72)
\]

Meanwhile, we also notice that in dual game \( \tilde{\Gamma}_\lambda(w) \), the security level \( J^\tau \) of a stationary strategy \( \tau \) that depends only on \( w_t \) satisfies \( J^\tau(w) = \bar{F}_{\tau(w)}(w) \).

**Lemma 20.** Let \( \tau \in \mathcal{T} \) be the uninformed player’s stationary strategy that depends only on the anti-discounted regret \( w_t \). The security level \( J^\tau \) of \( \tau \) in a \( \lambda \)-discounted asymmetric information repeated game \( \Gamma_\lambda(p) \) satisfies \( J^\tau(w) = \bar{F}_{\tau(w)}(w) \), where \( \bar{F}_{\tau(w)} \) is defined in (60).

**Proof:** According to Bellman’s principle, we have

\[
J^\tau(w) = \max_{p \in \Delta(k)} \max_{x \in \Delta(A)^{|K|}} \left( \sum_{a \in A, k \in K} p^k x^k(a) w^k + \sum_{a \in A, k \in K} \lambda p^k x^k(a) M_{a, \tau}(w) + (1 - \lambda) \sum_{a \in A} \bar{x}_p(a) \max_{\gamma \in \Sigma} \gamma_\lambda(\pi(p, x, a), \sigma, \tau) \right)
\]

\[
= \max_{p \in \Delta(k)} \max_{x \in \Delta(A)^{|K|}} \left( 1 - \lambda \right) \sum_{a \in A} \bar{x}_p(a) \left( \sum_{k \in K} \pi(p, x, a)(w^k + \lambda M_{a, \tau}(w)) + \max_{\sigma \in \Sigma} \gamma_\lambda(\pi(p, x, a), \sigma, \tau) \right)
\]

\[
= \max_{x \in \Delta(A)} \left( 1 - \lambda \right) \sum_{a \in A} \bar{x}(a) \max_{p \in \Delta(k)} \max_{\sigma \in \Sigma} \left( \sum_{k \in K} p^k w^k + \lambda M_{a, \tau}(w) \right) + \max_{\sigma \in \Sigma} \gamma_\lambda(p^+ \sigma, \tau) \right)
\]

\[
= \max_{a \in A} \bar{J}^\tau \left( \frac{w + \lambda M_{a, \tau}(w)}{1 - \lambda} \right) = \bar{F}_{\tau(w)}(w)
\]

Now, we are ready to analyze the performance difference between the approximated security strategy \( \tilde{\tau}_{\lambda,N} \) and the security strategy \( \tilde{\tau}^\tau \).

**Theorem 21.** Consider a two-player zero-sum \( \lambda \)-discounted asymmetric information repeated game \( \Gamma_\lambda(p) \) and the uninformed player’s approximated security strategy \( \tilde{\tau}_{\lambda,N} \) defined in (63). The security level \( J_{\tilde{\tau}_{\lambda,N}}(p) \) of \( \tilde{\tau}_{\lambda,N} \) in game \( \Gamma_\lambda(p) \) satisfies

\[
\| J_{\tilde{\tau}_{\lambda,N}} - V_\lambda \|_{\text{sup}} \leq \frac{2}{\lambda} \| V_\lambda - V_{\lambda,N} \|_{\text{sup}} \quad (73)
\]

\[
\leq \frac{2(1 - \lambda)^N}{\lambda} \| V_{\lambda,N} \|_{\text{sup}} \quad (74)
\]

**Proof:** Since \( J_{\tilde{\tau}_{\lambda,N}}(p) \geq V_\lambda(p), \) according to equation (72), we have for any \( p, \)

\[
\| J_{\tilde{\tau}_{\lambda,N}} - V_\lambda \|_{\text{sup}} \leq \min_{w \in \mathbb{R}^{|K|}} \{ \tilde{J}^\tau_{\lambda,N}(w) - p^T w \} - V_\lambda(p)
\]

\[
\leq \tilde{J}^\tau_{\lambda,N}(w^*) - p^T w^* - V_\lambda(p)
\]

\[
= \tilde{J}^\tau_{\lambda,N}(w^*) - V_{\lambda,N}(p) - p^T w^* + V_{\lambda,N}(p) - V_\lambda(p)
\]

\[
\leq |\tilde{J}^\tau_{\lambda,N}(w^*) - V_{\lambda}(w^*)| + |V_{\lambda}(w^*) - V_{\lambda,N}(w^*)| + |V_{\lambda,N}(p) - V_\lambda(p)|
\]

\[
\leq \| J_{\tilde{\tau}_{\lambda,N}} - V_\lambda \|_{\text{sup}} + \| V_{\lambda,N} - V_\lambda \|_{\text{sup}} + \| V_{\lambda,N} - V_\lambda \|_{\text{sup}}
\]

The third equality is derived from equation (58). Notice that the first inequality implies how the approximated initial regret \( w^* \) influences the performance of the approximated security strategy \( \tilde{\tau}_{\lambda,N} \). Since the equation above holds for all \( p, \) we have

\[
\| J_{\tilde{\tau}_{\lambda,N}} - V_\lambda \|_{\text{sup}} \leq \| J_{\tilde{\tau}_{\lambda,N}} - V_{\lambda,N} \|_{\text{sup}} + \| V_{\lambda,N} - V_\lambda \|_{\text{sup}} + \| V_{\lambda,N} - V_\lambda \|_{\text{sup}}
\]

Following the same steps as in the proof of Theorem 10, we can show that \( \| J_{\tilde{\tau}_{\lambda,N}} - V \|_{\text{sup}} \leq \frac{2(1 - \lambda)}{\lambda} \| V_{\lambda,N} - V_{\lambda} \|_{\text{sup}}. \)

Together with (59), we have

\[
\| J_{\tilde{\tau}_{\lambda,N}} - V_\lambda \|_{\text{sup}} \leq \frac{2}{\lambda} \| V_{\lambda,N} - V_\lambda \|_{\text{sup}}
\]

Equation (74) can be further derived from Theorem 7.

Equation (74) implies that the difference between the security level of the uninformed player’s approximated security strategy \( \tilde{\tau}_{\lambda,N} \) and the game value decreases exponentially fast with respect to \( N \), and converges to \( 0 \) as \( N \) goes to infinity.

**IV. CASE STUDY: A NETWORK INTERDICATION PROBLEM**

This section uses game theoretic tools to study a network interdiction problem developed in [2], and provides security strategies and approximated security strategies for both players (attacker and network) in finite-horizon game and discounted game, respectively.

Consider a network with a source node and a sink node. There are two channels from the source node to the sink node. One of them has high capacity of 3, and the other one has low capacity of 1. Only the network knows which channel has high capacity. The network needs to choose a channel to use at each stage to maximize the throughput over a certain horizon. Meanwhile, the attacker will either block one channel with cost 1 or observe the usage of channels with cost 0 to minimize the throughput over the same horizon. Notice that the attacker can only detect whether a channel is in use, but cannot measure the capacity of a channel. Our objective is to design security or approximated security strategies for both players.

The network interdiction problem is modeled as an asymmetric repeated game with the network to be the informed player and the attacker to be the uninformed player. The network’s action is to choose channel 1 (1) or use channel 2 (2), and the attacker’s action is to observe (o), block channel 1 (1), or block channel 2 (2). The payoff matrices are provided.
as in Table I. The initial probability that channel 1 has high capacity is 0.5.

We first compute the security strategies and security levels for both the network and the attacker in a 3-stage asymmetric information game according to Theorem 3 and 4, respectively. The linear program used to compute the network’s security strategy has 65 constraints and 35 variables, while the attacker’s LP formulation has 44 constraints and 23 variables. The security level of the network is 6.57 which meets the security level of the attacker.

The security strategy of the network is given in Table II. Consider the case in which channel 1 has high capacity. At stage 1, the network uses the high capacity channel with probability 0.64 instead of 1, because if the network reveals the high capacity channel at stage 1, the attacker will block the high capacity channel for the next two stages. At stage 2, if channel 1 was used at stage 1, then the network thinks that the attacker may guess that channel 1 has high capacity, and hence the network reduces its probability of using channel 1 to 0.56. Otherwise, the probability of using channel 1 is increased to 0.8. At the final stage, unless channel 1 is continuously used, the network will use high capacity channel for sure.

The security strategy of the attacker is shown in Table III. Notice that because the cost of blocking a channel is low compared with the gain of blocking the high capacity channel, the attacker prefers blocking channels to observing channels. Therefore, for many cases, the attacker launches attacks instead of observing channels unless he is almost sure which channel has high capacity. In this case, because the loss of blocking low capacity channel is higher than the loss of observing channels (see Table I), the attacker would prefer observing channels to blocking low capacity channel. At stage 1, since the initial probability over the states is [0.5 0.5], the attacker will block either channel with equal probability. At stage 2, the attacker will increase the probability of blocking channel 1 by 0.04 if channel 1 is used at stage 1. Otherwise, the probability of blocking channel 1 is decreased by 0.04. At stage 3, if one channel was used continuously, the attacker’s realized loss in the case that this channel has high capacity is already high, so his strategy focuses more on minimizing the payoff if the continuously used channel has high capacity, as if he is playing only a single game.

The security strategies of both players are, then, used in the 3-stage network interdiction game. We ran the 3-stage game for 5000 times, and the average total payoff of the network was 6.58 which was approximately the game value 6.57 computed according to Theorem 3 and 4.

Next, we compute the approximated security strategies for both players in a 0.7-discounted asymmetric repeated game. According to Theorem 8, the network computes his approximated security strategy based on the approximated game value $V_{\lambda,4}$. The game values from the discounted 1-stage game to the discounted 4-stage game are presented in the top plot of Figure 1. We see that the approximated game value converges, and that the more unsure the attacker is about the high capacity channel, the higher throughput the network can get, and the highest approximated game value is 2.24 when the initial probability is [0.5 0.5]. The approximated security strategy is given in the bottom plot of Figure 1. For both cases, the probability of using channel 1 is lower if the network thinks that the attacker has stronger belief that channel 1 has higher capacity. Meanwhile, compared to the case in which channel 2 has high capacity (green dots), it is more possible for the network to use channel 1 if channel 1 has high capacity (blue crosses).
To compute the attacker’s approximated security strategy, we first need to compute the approximated special initial regret $w^*$ according to Theorem 14, which is $[-2.24; -2.24]$ for $N = 4$. At each stage, the attacker computes his approximated security strategy based on the anti-discounted regret $w_1$ with $w_1 = w^*$. We assume that the anti-discounted regret $w_1^2$ if channel 2 has high capacity is 2, and use the approximated value $\hat{V}_{\lambda}^A(w_1) \leq 2$ to compute the approximated security strategy. The approximated game values $\hat{V}_{\lambda,N}$ where $N$ varies from 1 to 4 are presented in the top plot of Figure 2. We see that the approximated game value converges over $N$, and increases with respect to $w_1^2$. The attacker’s approximated security strategy is shown in the bottom plot of Figure 2. When $w_1^2$ is relatively low compared with $w_1^2$, the attacker will block channel 2 with higher probability to balance the payoffs of both cases, as if he believes that it is more possible for channel 2 to have high capacity. Contrarily, when $w_1^2$ is larger than $w_1^2$, the attacker will block channel 1 with higher probability to balance the payoffs of both cases, as if he believes that it is more possible for channel 1 to have high capacity.

The approximated security strategies of both players are, then, used in a 0.7-discounted network interdiction game. Before running the game, we first anticipate the average payoff of the game. From equation (24), we have that $\|V_{\lambda,N}\|_{\infty}^A \leq \|V_{\lambda}\| \leq \frac{\|V_{\lambda,N}\|_{\infty}}{1+(1-\lambda)^N}$. Together with equation (31) and (24), the network can guarantee a payoff $|J_{\lambda,N}^A(p)| = (1 - \frac{2(1-\lambda)^{N+1}}{\lambda}) \frac{\|V_{\lambda,N}\|_{\infty}}{1+(1-\lambda)^N} = 2.21$. Together with equation (73) and (24), the attacker can guarantee a payoff $|J_{\lambda,N}^A(p)| \leq (1 + \frac{2(1-\lambda)^N}{\lambda}) \frac{\|V_{\lambda,N}\|_{\infty}}{1+(1-\lambda)^N} = 2.31$. Therefore, we anticipate that the payoff is in the interval $[2.21, 2.31]$. When running the game, we stopped at stage 10 since the sum of the payoff after stage 10 is less than $10^{-3}$. The 10-stage 0.7-discounted game was run for 100 times, and the average payoff is 2.30 which is within our anticipated interval, and demonstrates our main results.

V. CONCLUDING REMARKS

This paper studies asymmetric repeated games in which one player has superior information about the game over the other, and provides LP formulations to compute both player’s security strategies in finite-horizon games and approximated security strategies in discounted games. In the future, we will generalize these results to the case in which one player has superior knowledge of one part of the information, while the other player has superior knowledge of the other part.

REFERENCES


Lichun Li is a postdoctoral researcher in Coordinated Science Lab at University of Illinois at Urbana-Champaign. Before that, she was also a postdoctoral in School of Electrical and Computer Engineering at Georgia Institute of Technology. Li received her Ph.D. in Department of Electrical Engineering from University of Notre Dame in 2013. Li’s research interests include game theory, algorithmic game theory, optimization, and discrete event systems, with the application in cyber-physical systems.

Jeff Shamma is a Professor and Chair of Electrical Engineering at the King Abdullah University of Science and Technology (KAUST) in Thuwal, Saudi Arabia, where he is also the director of the Robotics, Intelligent Systems & Control laboratory (RISC). He is the former Julian T. Hightower Chair in Systems & Control in the School of Electrical and Computer Engineering at Georgia Tech. He also has held faculty positions at the University of Minnesota, The University of Texas at Austin, and the University of California, Los Angeles. Shamma received a Ph.D. in systems science and engineering from MIT in 1988. He is the recipient of an NSF Young Investigator Award, the American Automatic Control Council Donald P. Eckman Award, and the Mohammed Dahleh Award. Shamma is a Fellow of the IEEE and the International Federation of Automatic Control (IFAC) and a current distinguished lecturer of the IEEE Control Systems Society. He is the deputy editor-in-chief for the IEEE *Transactions on Control of Network Systems* and an associate editor for the journal *Games*. Shamma’s research is in the general area of feedback control and systems theory. His most recent research has been in decision and control for distributed multi-agent systems and the related topics of game theory and network science, with applications to cyber-physical and societal network systems.