Efficient Computation of Discounted Asymmetric Information Zero-Sum Stochastic Games

Lichun Li and Jeff S. Shamma

Abstract— **In asymmetric information zero-sum games, one player has superior information about the game over the other. Asymmetric information games are particularly relevant for security problems, e.g., where an attacker knows its own skill set or alternatively a system administrator knows the state of its resources. In such settings, the informed player is faced with the tradeoff of exploiting its superior information at the cost of revealing its superior information. This tradeoff is typically addressed through randomization, in an effort to keep the uninformed player informationally off balance. A lingering issue is the explicit computation of such strategies. This paper, building on prior work for repeated games, presents an LP formulation to compute suboptimal strategies for the informed player in discounted asymmetric information stochastic games in which state transitions are not affected by the uninformed player. Furthermore, the paper presents bounds between the security level guaranteed by the sub-optimal strategy and the optimal value. The results are illustrated on a stochastic intrusion detection problem.**

I. Introduction

Games of incomplete information are motivated by settings in which decision makers have imperfect knowledge about the underlying state of the environment/game being played. Information is in the form of signals that are correlated to the state of the environment (e.g., as in the classical setting of Bayesian games [1]). In asymmetric information games, as the terminology implies, players observed different signals, and have different information about the world.

Two-player zero-sum asymmetric information repeated games were analyzed in [2], [3] in the late 1960s. In this model, there are several candidate games characterized by different payoff matrices. At the outset, one candidate game is chosen according to an *a priori* probability, and this game is to be played repeatedly over a specified horizon. The selected game is known only to the informed player, hence the asymmetry in information. At each stage, players choose their actions simultaneously. Actions are observable by both player, but the payoffs are not known to the uninformed player. The work of [2], [3] examines the zero-sum setting and characterizes the value of the resulting repeated game in terms of an associated Bellman equation. In particular, the informed player randomizes in order to balance exploiting its information versus revealing its information, and the underlying Bellman equation characterizes the payoff in terms of the posterior beliefs of the uninformed player.

Asymmetric information stochastic games are slightly different from asymmetric repeated games in that the game being played is not repeated, but may change stage by stage according to specified transition probabilities. This difference, however, does not change the heart of the phenomenon in asymmetric information games: the uninformed player must learn what he can about the concealed information from the actions of the other. It is shown that if the transition law is not affected by the uninformed player, the asymmetric information stochastic game has a value [4], [5]; otherwise, the max-min and min-max values still exist but may differ.

A main impediment towards computing optimal policies is the underlying complicated computational complexity, especially in the infinite horizon case which is non-convex [6], [7]. Finite multi-stage games, both repeated and stochastic, can be expressed as finite extensive form games. The notion of the sequence form of extensive games was developed in [8], which also derived efficient computational methods for the sequence form of the game. The work of [8] establishes that the size of the sequence form is linear in the size of extensive form games. However this size is exponential over the time horizon at a growth rate of the product of the sizes of both player's action sets and, in the case of stochastic games, the size of the state set. Because of the large size of the sequence form, abstractions of the sequence form were proposed [9], [10], [11], but there was no guarantee about how close the optimal strategies of the abstractions are from the true optimal strategy of the original game. Realizing that players in asymmetric information games are Bayesian players, and the uninformed player's belief about the game being played is an essential variable for both players to make decisions [12], [13], [4], [5], we presented a nested LP formulation to compute the optimal strategy of the original asymmetric information stochastic games in [14]. The size of the nested LP formulation was linear over the size of the state set and the size of the uninformed player's action set, but still exponential over the time horizon at a growth rate of the size of the informed player's action set.

Based on the work in [14], this paper develops an efficient approach to compute a sub-optimal strategy in discounted asymmetric information stochastic games whose transition probabilities are not affected by the uninformed player. Furthermore, the paper bounds the error between the security level guaranteed by the sub-optimal strategy and the true game value. An approximated game value is first computed

Lichun Li is with the department of Electrical and Computer Engineering, Georgia Institute of Technology, Atlanta, GA 30308, USA. lichun.li@ece.gatech.edu

J.S. Shamma is with the School of Electrical and Computer Engineering, Georgia Institute of Technology, shamma@gatech.edu, and with King Abdullah University of Science and Technology (KAUST), jeff.shamma@kaust.edu.sa.

The authors acknowledge the financial support of ARO project #W911NF-09-1-0553 and the AFOSR/MURI project #FA9550-10-1-0573.

based on value iteration and shown to converge to the game value exponentially fast. A sub-optimal strategy is then derived from the approximated game value with a guaranteed security level from the game value. Finally, an LP formulation is proposed to compute the sub-optimal strategy.

II. The model

Let $\mathbb{R}^n$, $\mathbb{R}_0^+$ and $\mathbb{Z}^+$ indicate the n-dimensional real space, non-negative real numbers, and positive integers, respectively. For a finite set S, $|S|$ denotes its cardinality, and $\Delta(S)$ indicates the set of probability distributions over S. $(p^s)_{s\in\mathcal{S}}$ is a $|S|$-dimensional column vector whose element is p^s. Given $x \in \mathbb{R}^n$, $(x)^k$ is the the kth element of x. $\mathbf{1}$ and $\mathbf{0}$ are appropriately dimensional column vectors with all their elements to be 1 and 0, respectively.

A two-player zero-sum asymmetric information stochastic game with the informed player controlling the transition is specified by the six-tuple (S, I, J, Q, G, p_0), where

- S is a non-empty finite set, called the state set, the elements of which are called states;
- I is a non-empty finite set, called player 1's action set.
- J is a non-empty finite set, called player 2's action set.
- $Q : S \times I \to \Delta(S)$ is the transition probability. $Q^i \in \mathbb{R}^{|S|\times|S|}$ indicates the transition matrix given player 1's action $i \in I$, and $Q^i_{s',s}$ is the transition probability over $s' \in S$ given $(s, i) \in S \times I$.
- $G : S \times I \times J \to \mathbb{R}$ is the one-stage payoff function. G^s indicates the payoff matrix given state $s \in S$, and $G^s_{i,j}$ is the payoff given state $s \in S$, player 1's action $i \in I$ and player 2's action $j \in J$.
- $p_0 \in \Delta(S)$ is the initial probability on S.

Notice that a two-player zero-sum asymmetric repeated game is a special case of the asymmetric stochastic game described above by letting Q^i be an identity matrix for all $i \in I$.

The two-player zero-sum asymmetric information stochastic game with the informed player controlling the transition is played in the following way. Let s_t, i_t, j_t denote the state, player 1's action, and player 2's action at stage $t \in \mathbb{Z}^+$, respectively. At stage 1, s_1 is chosen according to p_0, and communicated to player 1 only. Each player chooses his action independently, and the couple (i_1, j_1) is announced publicly. At stage $t \geq 2$, s_t is chosen according to $Q^{i_{t-1}}_{:,s_{t-1}}$, and communicated to player 1 only. Both players choose their actions independently. The pair (i_t, j_t) is, then, told to both. At every stage $t \geq 1$, after both players take their actions, player 2 pays $G(s_t, i_t, j_t)$ to player 1. The payoff is not announced to either player.

The state history space at stage t is defined as $H_t^S = S^t$ for $t \in \mathbb{Z}^+$. Player 1 and player 2's action history spaces at stage t are defined as $H_t^I = I^{t-1}$ and $H_t^J = J^{t-1}$ for $t \in \mathbb{Z}^+$, respectively. H_1^I and H_1^J are both $\{\emptyset\}$.

A behavior strategy for player 1 is an element of $\sigma = (\sigma_t)_{t\in\mathbb{Z}^+}$, where σ_t is a map from $H_t^S \times H_t^I \times H_t^J$ to $\Delta(I)$. Similarly, but taking into account his lack of information on S, a strategy for player 2 is an element of $\tau = (\tau_t)_{t\in\mathbb{Z}^+}$, where τ_t is a map from $H_t^I \times H_t^J$ to $\Delta(J)$. Denote by Σ and $\mathcal{T}$ the sets of strategies of player 1 and 2, respectively.

A triple (p_0, σ, τ) induces a probability distribution $P_{p_0,\sigma,\tau}$ on the set $\Omega = (S \times I \times J)^\infty$ of plays. $E_{p_0,\sigma,\tau}$ stands for the corresponding expectation.The λ-discounted payoff with initial probability p_0 and strategies σ and τ is defined as

$$\gamma_\lambda(p_0, \sigma, \tau) = E_{p_0,\sigma,\tau}\left(\sum_{t=1}^{\infty} \lambda(1-\lambda)^{t-1} G^{s_t}_{i_t,j_t}\right),$$

where $\lambda \in [0, 1)$.

The λ-discounted game $\Gamma_\lambda(p_0)$ is defined as a two-player zero-sum asymmetric information stochastic game with the informed player controlling the transition, equipped with initial distribution p_0, strategy spaces Σ and $\mathcal{T}$, and payoff function $\gamma_\lambda(p_0, \sigma, \tau)$.

The λ-discounted game $\Gamma_\lambda(p_0)$ has a value $v_\lambda(p_0)$ iff $\underline{v}_\lambda(p_0) = \bar{v}_\lambda(p_0) \doteq v_\lambda(p_0)$, where $\underline{v}_\lambda(p_0) = \sup_{\sigma\in\Sigma}\inf_{\tau\in\mathcal{T}}\gamma_\lambda(p_0, \sigma, \tau)$ and $\bar{v}_\lambda(p_0) = \inf_{\tau\in\mathcal{T}}\sup_{\sigma\in\Sigma}\gamma_\lambda(p_0, \sigma, \tau)$ are the maxmin and minmax values of the game $\Gamma_\lambda(p_0)$, respectively.

The value $v_\lambda(p_0)$ of the λ-discounted game exists and satisfies a recursive formula (subsection 6.5.3 of [13]). Let $p \in \Delta(S)$ represent player 2's belief on the state at stage t, and assume player 1 and 2 choose their actions according to $x_t = (x_t^s)_{s\in S} \in \Delta(I)^{|S|}$ and $y_t \in \Delta(J)$. The expected payoff of player 1 at stage t is $g(p, x_t, y_t) = \sum_{s\in S} p^s x_t^s{}^T G^s y_t$.

The probability that player 1 plays $i \in I$ at stage t is

$$\bar{x}_{p,x_t}(i) = \sum_{s\in S} p^s x_t^s(i). \tag{1}$$

Given that player 1 played $i \in I$ according to x_t at stage t, the conditional probability over the state at stage $t+1$ is denoted by $p^+(p, x_t, i)$ satisfying

$$p^+(p, x_t, i) = Q^i\left(\frac{p^s x_t^s(i)}{\bar{x}_{p,x_t}(i)}\right)_{s\in S} \tag{2}$$

Define

$$\mathbf{T}_{p,x_t}(v_\lambda) = \sum_{i\in I} \bar{x}_{p,x_t}(i) v_\lambda(p^+(p, x_t, i)). \tag{3}$$

Lemma 2.1 (section 6.5.3 of [13]): For any $p_0 \in \Delta(s)$,

$$\begin{aligned} & v_\lambda(p_0) \\ &= \max_{x\in\Delta(I)^{|S|}} \min_{y\in\Delta(J)} \left(\lambda g(p_0, x, y) + (1-\lambda)\mathbf{T}_{p_0,x}(v_\lambda)\right) \\ &= \min_{y\in\Delta(J)} \max_{x\in\Delta(I)^{|S|}} \left(\lambda g(p_0, x, y) + (1-\lambda)\mathbf{T}_{p_0,x}(v_\lambda)\right), \end{aligned} \tag{4}$$

Player 1 has an optimal stationary strategy in $\Gamma_\lambda(p_0)$ that depends only on player 1's history actions.

III. Value Iteration

The dynamic programming-like equation (4) is generally difficult to solve. This section studies how to use value iteration to approximate the value function, and the convergence rate of the value iteration.

Suppose the initial approximated game value $v_{\lambda|0}(p) \equiv 0$. The approximated game value $v_{\lambda|n}(p)$ at the nth iteration is

updated in the following way.

$$v_{\lambda|n+1}(p) \tag{5}$$
$$= \max_{x_n\in\Delta(I)^{|S|}} \min_{y_n\in\Delta(J)} (\lambda g(p,x_n,y_n) + (1-\lambda)\mathbf{T}_{p,x_n}(v_{\lambda|n})).$$

Let us first define two operators $\mathbf{F}$ and $\mathbf{H}$ as follows:

$$\mathbf{F}^v_x(p) = \min_{y\in\Delta(J)} \{\lambda g(p,x_n,y_n) + (1-\lambda)\mathbf{T}_{p,x}(v)\}, \tag{6}$$

$$\mathbf{H}^v(p) = \max_{x\in\Delta(I)^{|S|}} \mathbf{F}^v_x(p). \tag{7}$$

where $p\in\Delta(S)$, $x\in\Delta(I)^{|S|}$, and $v:\Delta(S)\to\mathbb{R}$. It is easy to see that $v_\lambda(p) = \mathbf{H}^{v_\lambda}(p)$, and $v_{\lambda|n+1}(p) = \mathbf{H}^{v_{\lambda|n}}(p)$. Let $\mathcal{F}$ indicate a function space including all functions from $\Delta(S)$ to $\mathbb{R}$. Define the supreme norm $\|\cdot\|_{\sup}$ in function space $\mathcal{F}$ as $\|v\|_{\sup} = \sup_{p\in\Delta(S)}|v(p)|$ for all $v\in\mathcal{F}$.

Definition 3.1: We say an operator $\mathbf{H}:\mathcal{F}\to\mathcal{F}$ is a contractor with contraction constant α if there exists a constant $\alpha\in[0,1)$ such that $\|\mathbf{H}^v-\mathbf{H}^{\tilde v}\|_{\sup}\le\alpha\|v-\tilde v\|_{\sup}$, for any $v,\tilde v\in\mathcal{F}$.

The rest of this section shows that the functionals $\mathbf{F}_x$ and $\mathbf{H}$ are contractors, and hence the value iteration (5) converges to the game value v_λ exponentially.

Lemma 3.2: Given any $x\in\Delta(I)^{|S|}$ and $\lambda\in(0,1)$, the functional $\mathbf{F}_x:\mathcal{F}\to\mathcal{F}$ defined in (6) is a contractor with contraction constant $1-\lambda$, i.e.

$$\|\mathbf{F}^v_x-\mathbf{F}^{\tilde v}_x\|_{\sup}\le(1-\lambda)\|v-\tilde v\|_{\sup}, \forall v,\tilde v\in\mathcal{F}.$$

Proof: Since the second term of mapping $\mathbf{F}_x$ in equation (6) is irrelevant to y, we have

$$\mathbf{F}^v_x(p) = \min_{y\in\Delta(J)}\{\lambda\sum_{s\in S}p^s(x^s)^TG^sy\} + (1-\lambda)\mathbf{T}_{p,x}(v),$$
$$\mathbf{F}^{\tilde v}_x(p) = \min_{y\in\Delta(J)}\{\lambda\sum_{s\in S}p^s(x^s)^TG^sy\} + (1-\lambda)\mathbf{T}_{p,x}(\tilde v).$$

Therefore, according to the definition of $\mathbf{T}$ in (3),

$$|\mathbf{F}^v_x(p)-\mathbf{F}^{\tilde v}_x(p)|$$
$$=(1-\lambda)\sum_{i\in I}\bar x_{p,x}(i)|v(p^+(p,x,i))-\tilde v(p^+(p,x,i))|.$$

Its supreme norm, hence, satisfies $\|\mathbf{F}^v_x-\mathbf{F}^{\tilde v}_x\|_{\sup}\le(1-\lambda)\|v-\tilde v\|_{\sup}\sup_{p\in\Delta(S)}\sum_{i\in I}\bar x_{p,x}(i) = (1-\lambda)\|v-\tilde v\|_{\sup}$. ∎

Lemma 3.3: Given $\lambda\in(0,1)$, the functional $\mathbf{H}:\mathcal{F}\to\mathcal{F}$ as in (7) is a contractor with contraction constant $1-\lambda$, i.e.

$$\|\mathbf{H}^v-\mathbf{H}^{\tilde v}\|_{\sup}\le(1-\lambda)\|v-\tilde v\|_{\sup}, \forall v,\tilde v\in\mathcal{F}.$$

Proof: From the definition of $\mathbf{H}$ in (7), we have

$$|\mathbf{H}^v(p)-\mathbf{H}^{\tilde v}(p)| = |\max_{x\in\Delta(I)^{|S|}}\mathbf{F}^v_x(p) - \max_{x\in\Delta(I)^{|S|}}\mathbf{F}^{\tilde v}_x(p)|.$$

Let $F^v_{x^*}(p) = \max_{x\in\Delta(I)^{|S|}}\mathbf{F}^v_x(p)$, $F^{\tilde v}_{x^+}(p) = \max_{x\in\Delta(I)^{|S|}}\mathbf{F}^{\tilde v}_x(p)$.

If $F^v_{x^*}(p)\ge F^{\tilde v}_{x^+}(p)$, then $|\mathbf{H}^v(p)-\mathbf{H}^{\tilde v}(p)|\le|F^v_{x^*}(p)-F^{\tilde v}_{x^*}(p)|\le\max_{x\in\Delta(I)^{|S|}}|\mathbf{F}^v_x(p)-\mathbf{F}^{\tilde v}_x(p)|$. If $F^v_{x^*}(p)\le F^{\tilde v}_{x^+}(p)$, then $|\mathbf{H}^v(p)-\mathbf{H}^{\tilde v}(p)|\le|F^v_{x^+}(p)-F^{\tilde v}_{x^+}(p)|\le\max_{x\in\Delta(I)^{|S|}}|\mathbf{F}^v_x(p)-\mathbf{F}^{\tilde v}_x(p)|$. Therefore, $|\mathbf{H}^v(p)-\mathbf{H}^{\tilde v}(p)|\le\max_{x\in\Delta(I)^{|S|}}|\mathbf{F}^v_x(p)-\mathbf{F}^{\tilde v}_x(p)|\le\max_{x\in\Delta(I)^{|S|}}\|\mathbf{F}^v_x-\mathbf{F}^{\tilde v}_x\|_{\sup}\le(1-\lambda)\|v-\tilde v\|_{\sup}n$ The second inequality holds because Lemma 3.2 holds for any $x\in\Delta(I)^{|S|}$. Its supreme norm: $\|\mathbf{H}^v-\mathbf{H}^{\tilde v}\|_{\sup}\le(1-\lambda)\|v-\tilde v\|_{\sup}$. ∎

Theorem 3.4: Given $\lambda\in(0,1)$, the approximated value function $v_{\lambda|n}$ satisfying (5) converges to v_λ exponentially with rate $1-\lambda$, i.e.

$$\|v_\lambda-v_{\lambda|n}\|_{\sup}\le(1-\lambda)\|v_\lambda-v_{\lambda|n-1}\|_{\sup}$$
$$\le(1-\lambda)^n\|v_\lambda\|_{\sup}.$$

Proof: Equation (4-5) imply $|v_\lambda(p)-v_{\lambda|n+1}(p)| = |\mathbf{H}^{v_\lambda}(p)-\mathbf{H}^{v_{\lambda|n}}(p)|$. Therefore, $\|v_\lambda-v_{\lambda|n+1}\|_{\sup} = \|\mathbf{H}^{v_\lambda}-\mathbf{H}^{v_{\lambda|n}}\|_{\sup}\le(1-\lambda)\|v_\lambda-v_{\lambda|n}\|_{\sup}$. ∎

IV. Sub-optimal Policy and Its Error Bound

Based on the approximated game value $v_{\lambda|n}$, a sub-optimal strategy $\sigma_{\lambda|n}$ is computed using the following formula.

$$\sigma_{\lambda|n} = \arg\max_{\sigma\in\Sigma}\min_{y\in\Delta(J)}(\lambda g(p,\sigma(p),y) + (1-\lambda)T_{p,\sigma(p)}(v_{\lambda|n})) \tag{8}$$

The *security level* $J^{\sigma_{\lambda|n}}(p)$ guaranteed by the sub-optimal strategy $\sigma_{\lambda|n}$ is, then,

$$J^{\sigma_{\lambda|n}}(p) = \min_{\tau\in\mathcal{T}}\gamma_\lambda(p,\sigma_{\lambda|n},\tau). \tag{9}$$

Note that the security level $J^{\sigma_{\lambda|n}}(p)$ of policy $\sigma_{\lambda|n}$ is different from the approximated game value $v_{\lambda|n}$.

This section studies the difference between the security level $J^{\sigma_{\lambda|n}}$ and the optimal game value v_λ.

First, we show one property of stationary strategies.

Lemma 4.1: Let $\sigma\in\Sigma$ be a stationary strategy of the informed player. The security level J^σ of σ satisfies $J^\sigma(p) = \mathbf{F}^{J^\sigma}_{\sigma(p)}(p)$, where $\mathbf{F}$ is defined as in (6).

Proof: Since player I's strategy is fixed to be σ, the discounted game Γ_λ becomes a discounted optimization problem, and hence satisfies Bellman's principle, i.e. $J^\sigma(p)$

$$= \min_{y\in\Delta(J)}(\lambda g(p,\sigma(p),y) + (1-\lambda)E_i\left(J^\sigma(p^+(p,\sigma(p),i))\right)$$
$$= \min_{y\in\Delta(J)}(\lambda g(p,\sigma(p),y) + (1-\lambda)T_{p,\sigma(p)}(J^\sigma)) = \mathbf{F}^{J^\sigma}_{\sigma(p)}(p)$$

∎

Now, let's study the difference between $J^{\sigma_{\lambda|n}}$ and v_λ.

Theorem 4.2: The sub-optimal policy $\sigma_{\lambda|n}$ defined in equation (8) guarantees a security level $J^{\sigma_{\lambda|n}}$ satisfying

$$\|v_\lambda-J^{\sigma_{\lambda|n}}\|_{\sup}\le\frac{2(1-\lambda)}{\lambda}\|v_\lambda-v_{\lambda|n}\|_{\sup}$$
$$=\frac{2(1-\lambda)^{n+1}}{\lambda}\|v_\lambda\|_{\sup}.$$

Proof: Lemma 4.1 indicates that $|v_\lambda(p)-J^{\sigma_{\lambda|n}}(p)| = |v_\lambda(p)-\mathbf{F}^{J^{\sigma_{\lambda|n}}}_{\sigma_{\lambda|n}(p)}(p)|$. Therefore, with functions $\mathbf{F}$ and $\mathbf{H}$ defined in (6) and (7), $|v_\lambda(p)-J^{\sigma_{\lambda|n}}(p)|\le|v_\lambda(p)-\mathbf{F}^{v_{\lambda|n}}_{\sigma_{\lambda|n}}(p)|+|\mathbf{F}^{v_{\lambda|n}}_{\sigma_{\lambda|n}}(p)-J^{\sigma_{\lambda|n}}(p)| = |\mathbf{H}^{v_\lambda}(p)-\mathbf{H}^{v_{\lambda|n}}(p)|+|\mathbf{F}^{v_{\lambda|n}}_{\sigma_{\lambda|n}}(p)-\mathbf{F}^{J^{\sigma_{\lambda|n}}}_{\sigma_{\lambda|n}(p)}(p)|$.

Take the supreme norm on both sides. With Lemma 3.2-3.3, $\|v_\lambda-J^{\sigma_{\lambda|n}}\|_{\sup}\le(1-\lambda)\left(\|v_\lambda-v_{\lambda|n}\|_{\sup}+\|v_{\lambda|n}-J^{\sigma_{\lambda|n}}\|_{\sup}\right)\le(1-$

$\lambda)\left(\|v_\lambda - v_{\lambda|n}\|_{\sup} + \|v_{\lambda|n} - v_\lambda\|_{\sup} + \|v_\lambda - J^{\sigma_{\lambda|n}}\|_{\sup}\right)$, which implies the result in this theorem. ∎

Theorem 4.2 guarantees that the game value induced by the stationary policy $\sigma_{\lambda|n}$ is close to the optimal game value v_λ, as long as the approximated game value $v_{\lambda|n}$ is close enough to the optimal game value v_λ.

V. LP FORMULATION OF $v_{\lambda|n}$

This section discusses how to efficiently compute the approximated game value $v_{\lambda|n}$ and the corresponding sub-optimal strategy $\sigma_{\lambda|n}$. The basic idea is based on [14].

A. LP formulation of $v_{\lambda|1}$

Let us start from the first iteration. According to (5), the approximated value at the first iteration satisfies $v_{\lambda|1} = \max_{x_1\in\Delta(I)^{|S|}} \min_{y_1\in\Delta(J)} \lambda \sum_{s\in S} p^s x_1^{sT} G^s y_1$. Reference [15] showed that $v_{\lambda|1}$ can be computed through the following linear program.

$$\begin{aligned} &\max_{x_1,\ell_1} \ \lambda\ell_1 \\ s.t. \quad &\sum_{s\in S} p^s G^{sT} x_1^s \geq \ell_1 \mathbf{1} \\ &\mathbf{1}^T x_1^s = 1, \qquad \forall s\in S \\ &x_1^s \geq \mathbf{0}, \qquad \forall s\in S \end{aligned}$$

Let $z_1^s = p^s x_1^s$. The linear program above is changed to

$$\max_{z_1,\ell_1} \ \lambda\ell_1 \tag{10}$$

$$s.t. \quad \sum_{s\in S} G^{sT} z_1^s \geq \ell_1 \mathbf{1} \tag{11}$$

$$\mathbf{1}^T z_1^s = p^s, \qquad \forall s\in S \tag{12}$$

$$z_1^s \geq \mathbf{0}, \qquad \forall s\in S, \tag{13}$$

where $z_1 \in \mathbb{R}^{|I|\times|S|}$ is a matrix *strategy variable*, and ℓ_1 is a scalar *performance variable*. We call (11) the performance level constraint, (12) the a priori belief constraint, and (13) the non-negativeness constraint.

B. LP formulation of $v_{\lambda|n}$

By mathematical induction, the LP formulation of $v_{\lambda|n}$ takes a form similar to the one presented in (10-13).

Let $h_t^I = \{i_1, i_2, \cdots, i_{t-1}\} \in H_t^I$ be a history action sequence of player 1 at stage t. Define $\mathcal{P}(h_t^I) = \{i_1, i_2, \cdots, i_{t-2}\}$ as the parent of h_t^I, and $\mathcal{L}(h_t^I) = i_{t-1}$ as the leaf node of h_t^I. We have the following Theorem. Proof of Theorem 5.1 is presented in Appendix.

Theorem 5.1: Let $z_{t|h_t^I} \in \mathbb{R}^{|I|\times|S|}$ be the strategy variable and $\ell_{t|h_t^I} \in \mathbb{R}$ be the performance variable at stage t with history action sequence h_t^I. The approximated game value $v_{\lambda|n}(p)$ satisfies

$$v_{\lambda|n}(p) = \max_{\substack{z_{t|h_t^I}, \ell_{t|h_t^I}, \\ \forall t=1,\ldots,n, \\ \forall h_t^I \in H_t^I}} \sum_{t=1}^{n} \lambda(1-\lambda)^{t-1} \sum_{h_t^I\in H_t^I} \ell_{t|h_t^I} \tag{14}$$

$$s.t. \forall\ t = 1,\ldots,n, \forall\ h_t^I \in H_t^I$$

$$\sum_{s\in S} G^{sT} z^s_{t|h_t^I} \geq \ell_{t|h_t^I}\mathbf{1} \tag{15}$$

$$\mathbf{1}^T z^s_{t|h_t^I} = \left(Q^{\mathcal{L}(h_t^I)}(z^s_{t-1|\mathcal{P}(h_t^I)}(\mathcal{L}(h_t^I)))_{s\in S}\right)^s, \forall s\in S \tag{16}$$

$$z^s_{t|h_t^i} \geq \mathbf{0}, \forall s\in S \tag{17}$$

where $z^s_{t|h_t^I}$ indicates the sth column of $z_{t|h_t^I}$. The initial condition $z_{0|\mathcal{P}(h_1^I)}(\mathcal{L}(h_1^I)) = p$, and $Q^\emptyset$ is an identity matrix of proper dimension. Since $h_1^I = \emptyset$, $z_{1|h_1^I}$ is written as z_1. The corresponding sub-optimal strategy $x^s(p) = \frac{z_1^s}{\mathbf{1}^T z_1^s}$.

Comparing the LP formulation (14-17) of $v_{\lambda|n}$ with the LP formulation (10-13) of $v_{\lambda|1}$, we find that they are in the similar form. The first constraint (15) is the performance level constraint at stage t with history action h_t^I. The second constraint (16) is the a priori belief constraint indicating how the strategy variable $z_{t|h_t^I}$ is related with the strategy variable $z_{t-1|\mathcal{P}(h_t^I)}$ at its parent node. The third constraint (17) is to make sure that the strategy variable is non-negative.

The strategy variable z and the performance variable ℓ are only defined for every possible history actions of player 1 instead of both players. It is because player 1's optimal strategy only depends on his own actions.

The size of the linear program (14) is *linear* with respect to the size $|S|$ of the state set and the size $|J|$ of player 2's action set, *polynomial* over the size $|I|$ of player 1's action set, and *exponential* over the number of iterations. It is easy to see that to compute $v_{\lambda|n}$, there are at most $|I|^{n-1}$ different history actions of player 1 at stage n. For each history action at each stage, we need an independent pair of variables $z \in \mathbb{R}^{|I|\times|S|}$ and $\ell \in \mathbb{R}$, and hence in total there are $(1 + |I| + \cdots + |I|^{n-1})(1 + |I||S|) = \mathbf{O}(|\mathbf{S}||\mathbf{I}^{\mathbf{n+1}})$ scalar variables. Meanwhile, for each history action at each stage, there are three sets of constraints. The performance level constraint (15) has $|J|$ inequalities, the a priori constraint (16) has $|S|$ equalities, and the non-negativeness constraint (17) has $|I||S|$ inequalities. In total, there are $(1 + |I| + \cdots + |I|^{n-1})(|J| + |S| + |I||S|) = \mathbf{O}(|\mathbf{J}||\mathbf{I}|^{\mathbf{n}} + |\mathbf{S}||\mathbf{I}|^{\mathbf{n+1}})$ constraints. Therefore, we say the size of the LP formulation (14) grows only linearly with respect to $|J|$ and $|S|$, polynomially with respect to $|I|$, and exponentially with respect to n.

VI. CASE STUDY: ASYMMETRIC INFORMATION STOCHASTIC INTRUSION DETECTION GAME

The stochastic intrusion detection game was initially introduced in [16]. An administrator is assigned to protect a system from hackers' attacks. If the administrator does low level (ll) maintenance, then the system is vulnerable with high probability. If the administrator does high level (hl) maintenance, the probability of new vulnerabilities emerging decreases. Describe the two states of the system as 'vulnerable'(v) and 'non-vulnerable' (nv). The transition matrix is given in Table VI. The attacker decides whether to launch an attack (a) or not (na) at each stage. The damage caused

TABLE I

TRANSITION MATRICES OF STOCHASTIC INTRUSION DETECTION GAME

	nv	v
nv	0.9	0.8
v	0.1	0.2

Q^{hl}

	nv	v
nv	0.1	0.2
v	0.9	0.8

Q^{ll}

TABLE II

PAYOFF MATRICES OF STOCHASTIC INTRUSION DETECTION GAME

	a	na
hl	5	-10
ll	-1	0

G^{nv}

	a	na
hl	3	-11
ll	-2	0

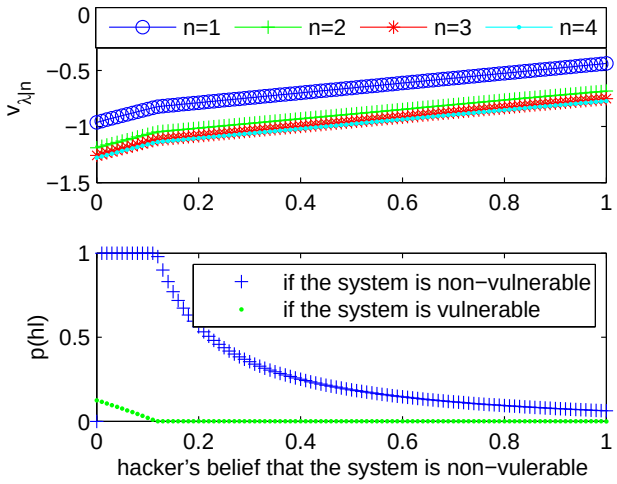
G^{v}

by an attack is always worse if the system is vulnerable, which is reflected by the payoff matrices in Table VI. The administrator's objective is to maximize his payoff while the hacker wants to minimize it. In the original problem in [16], the attacker is assumed to know whether the system is vulnerable, while in our case, the system state (vulnerable or not) is only known by the administrator, and the attacker only has an a priori belief p_0 of whether the system is vulnerable.

The asymmetric information stochastic intrusion detection problem is formulated as a discounted asymmetric game with the discount constant $\lambda = 0.7$. According to Theorem 5.1, a sub-optimal strategy $x(p) \in \Delta(I)^{|S|}$ is computed based on the approximated game value $v_{\lambda|4}$ at the 4th iteration. The approximated game value from the 1st to the 4th iteration is presented in the top plot of Figure 1 with the x-axis indicating hacker's belief that the system is non-vulnerable. We see that the game value is larger when the system is less vulnerable. In the two extreme cases, i.e. vulnerable for sure and non-vulnerable for sure, the game values are -1.278 and -0.7726, respectively. The sub-optimal strategy is given in the bottom plot Figure 1, where crosses and dots indicate the probability to do high level maintenance if the system is non-vulnerable and vulnerable, respectively. We notice that when the hacker believes that the system is less vulnerable, the administrator decreases the probability to do high level maintenance. This is because the hacker prefers to not attack when the system is less vulnerable, and the corresponding strategy for the administrator is to increase the probability to do low level maintenance.

The sub-optimal strategy is used in a 10-stage asymmetric information stochastic intrusion detection game for 5000 times with initial belief $p_0 = [0.5\ 0.5]$. We first analyze the expected game value of the experiment, and then run the experiment to see whether the outcome satisfies the expectation. Theorem 3.4 implies that $\|v_\lambda - v_{\lambda|4}\|_{\sup} \leq (1-0.7)^4 \|v_\lambda\|_{\sup}$, which indicates that $\|v_\lambda\|_{\sup} \leq \frac{1}{1-0.3^4}\|v_{\lambda|4}\|_{\sup} = 1.2884$. Theorem 3.4 also shows that $|v_\lambda(p_0) - v_{\lambda|4}(p_0)| \leq \|v_\lambda - v_{\lambda|4}\|_{\sup} \leq (1-0.7)^4\|v_\lambda\|_{\sup} = 0.0104$. Since $v_{\lambda|4}(p_0) = -0.9781$, we have $v_\lambda(p_0) \in [-0.9885, -0.9677]$. Theorem 4.2 says that the security level $J^{\sigma_{\lambda|4}}$ satisfies $|J^{\sigma_{\lambda|4}}(p_0) - v_\lambda(p_0)| \leq \|J^{\sigma_{\lambda|4}} - v_\lambda\|_{\sup} \leq \frac{2\times(1-0.7)^5}{0.7}\|v_\lambda\|_{\sup} = 0.0089$, and hence $J^{\sigma_{\lambda|4}}(p_0) \in [-0.9974, -0.9588]$. When running the 10-stage asymmetric game for 5000 times, at each stage, we assume the hacker uses a strategy that minimizes the payoff given the administrator's strategy is $\sigma_{\lambda|4}$. The average

Fig. 1. Approximated game value and sub-optimal strategy



security level guaranteed by the administrator is -0.9926 lying within the expected range $[-0.9974, -0.9588]$, which demonstrates the main results, Theorem 3.4, 4.2, and 5.1.

VII. CONCLUSION

This paper studies value iterations of discounted asymmetric information stochastic games, and shows that the approximated game values converges to the game values exponentially if the discount constant λ is in $(0, 1)$. The approximated game value is used to generate a sub-optimal strategy which guarantees a security level close to the true game value as long as the approximated game value is close enough to the true game value. Finally, an LP formulation is presented to compute the sub-optimal strategy. An asymmetric information stochastic intrusion detection game is used to demonstrate the main results.

APPENDIX

Lemma 1.1: For any finite positive integer n, any constant $\alpha \geq 0$, and any $p \in \Delta(S)$,

$$\alpha v_{\lambda|n}(p) = v_{\lambda|n}(\alpha p). \tag{18}$$

Proof: For $n = 1$, we have

$$v_{\lambda|1}(\alpha p) = \max_{x_1 \in \Delta(I)^{|S|}} \min_{y_1 \in \Delta(J)} \lambda g(\alpha p, x_1, y_1)$$
$$= \alpha \max_{x_1 \in \Delta(I)^{|S|}} \min_{y_1 \in \Delta(J)} \lambda g(p, x_1, y_1) = \alpha v_{\lambda|1}(p).$$

Now, let us assume equation (18) is true for $n-1$. From equation (1), we have $\bar{x}_{\alpha p, x_n}(i) = \alpha \bar{x}_{p,x_n}(i)$. From equation (2), we have $p^+(\alpha p, x_n, i) = p^+(p, x_n, i)$. Therefore, $T_{\alpha p, x_n}(v_{\lambda|n-1}) = \alpha T_{p,x_n}(v_{\lambda|n-1})$, and

$$\begin{aligned} & v_{\lambda|n}(\alpha p) \\ = & \max_{x_n \in \Delta(I)^{|S|}} \min_{y_n \in \Delta(J)} \big(\lambda g(\alpha p, x_n, y_n) \\ & + (1-\lambda) T_{\alpha p, x_n}(v_{\lambda|n-1})\big) \\ = & \max_{x_n \in \Delta(I)^{|S|}} \min_{y_n \in \Delta(J)} \big(\alpha \lambda g(p, x_n, y_n) \\ & + \alpha(1-\lambda) T_{p, x_n}(v_{\lambda|n-1})\big) = \alpha v_{\lambda|n}(p) \end{aligned}$$

■

Proof of Theorem 5.1 Equation (10-13) imply that Theorem 5.1 is true for $n = 1$. Assume that Theorem 5.1 is true for $n-1$. For the nth iteration, equation (5) indicates

$$v_{\lambda|n}(p) = \max_{x_1 \in \Delta(I)^{|S|}} \Bigg((1-\lambda) \sum_{i_1 \in I} v_{\lambda|n-1}(Q^{i_1}(p^s x_1^s(i_1))_{s\in S}) + \min_{y_1 \in \Delta(J)} \lambda g(p, x_1, y_1) \Bigg).$$

The 2nd equality holds because $v_{\lambda|n-1}(Q^{i_1}(p^s x_1^s(i_1))_{s\in S})$ is independent of y_1 and Lemma 1.1 indicates $T_{p,x_1}(v_{\lambda|n-1}) = v_{\lambda|n-1}(Q^{i_1}(p^s x_1^s(i_1))_{s\in S})$. Let us analyze this equation term by term.

Since Theorem 5.1 is true for the $n-1$th iteration, the first term $(1-\lambda)v_{\lambda|n-1}(Q^{i_1}(p^s x_1^s(i_1))_{s\in S})$ can be solved through the linear program below with the initial condition $z_{0|\mathcal{P}(h_1^I)}(\mathcal{L}(h_1^I)) = Q^{i_1}(p^s x_1^s(i_1))_{s\in S}$, and $Q^{\mathcal{L}(h_1^I)}$ an identity matrix.

$$(1-\lambda) \max_{\substack{z_{t'|h_{t'}^I}, \ell_{t'|h_{t'}^I}, \\ \forall t'=1,\dots,n-1, \\ \forall h_{t'}^I \in H_{t'}^I}} \sum_{t'=1}^{n-1} \lambda(1-\lambda)^{t'-1} \sum_{h_{t'}^I \in H_{t'}^I} \ell_{t'|h_{t'}^I}$$

$$s.t. \forall\ t' = 1, \dots, n-1, \forall\ h_{t'}^I \in H_{t'}^I$$

$$\sum_{s\in S} G^{sT} z^s_{t'|h_{t'}^I} \geq \ell_{t'|h_{t'}^I} \mathbf{1}$$

$$\mathbf{1}^T z^s_{t'|h_{t'}^I} = \left(Q^{\mathcal{L}(h_{t'}^I)} (z^s_{t'-1|\mathcal{P}(h_{t'}^I)}(\mathcal{L}(h_{t'}^I)))_{s\in S} \right)^s, \forall s \in S$$

$$z^s_{t'|h_{t'}^I} \geq \mathbf{0}, \forall s \in S$$

Let $t' = t-1$. The linear program above is written as

$$\max_{\substack{z_{t|h_t^I}, \ell_{t|h_t^I}, \\ \forall t=1,\dots,n-1, \\ \forall h_t^I \in H_t^I}} \sum_{t=2}^{n} \lambda(1-\lambda)^{t-1} \sum_{h_t^I \in H_t^I} \ell_{t|h_t^I}$$

$$s.t. \forall\ t = 2, \dots, n, \forall\ h_t^I \in H_t^I$$

$$\sum_{s\in S} G^{sT} z^s_{t|h_t^I} \geq \ell_{t|h_t^I} \mathbf{1}$$

$$\mathbf{1}^T z^s_{t|h_t^I} = \left(Q^{\mathcal{L}(h_t^I)} (z^s_{t-1|\mathcal{P}(h_t^I)}(\mathcal{L}(h_t^I)))_{s\in S} \right)^s, \forall s \in S$$

$$z^s_{t|h_t^I} \geq \mathbf{0}, \forall s \in S.$$

The initial condition is $z_{1|\mathcal{P}(h_2^I)}(\mathcal{L}(h_2^I)) = Q^{i_1}(p^s x_1^s(i_1))_{s\in S}$, and $Q^{\mathcal{L}(h_2^I)}$ an identity matrix. According to the second constraint of the linear program above, it is the same to say that the initial condition is $z_{1|\mathcal{P}(h_2^I)}(\mathcal{L}(h_2^I)) = (p^s x_1^s(i_1))_{s\in S}$, and $Q^{\mathcal{L}(h_2^I)} = Q^{i_1}$.

The second term $\min_{y_1\in\Delta(J)} \lambda g(p, x_1, y_1)$, according to Lemma III.2 of [14], has the value of

$$\max_{\ell_{1|h_1^I} \in \mathbb{R}} \lambda \ell_{1|h_1^I}$$

$$s.t. \sum_{s\in S} G^{sT} z^s_{1|h_1^I} \geq \ell_{1|h_1^I} \mathbf{1}$$

where $z^s_{1|h_1} = p^s x^s_{1|h_1^I}$.

Therefore, for the nth iteration, the game value $v_{\lambda|n}(p)$ is the value of

$$v_{\lambda|n}(p) = \max_{\substack{z_{t|h_t^I}, \ell_{t|h_t^I}, \\ \forall t=1,\dots,n, \\ \forall h_t^I \in H_t^I}} \sum_{t=1}^{n} \lambda(1-\lambda)^{t-1} \sum_{h_t^I \in H_t^I} \ell_{t|h_t^I} \quad (19)$$

$$s.t. \forall\ t = 1, \dots, n, \forall\ h_t^I \in H_t^I$$

$$\sum_{s\in S} G^{sT} z^s_{t|h_t^I} \geq \ell_{t|h_t^I} \mathbf{1} \quad (20)$$

$$\mathbf{1}^T z^s_{t|h_t^I} = \left(Q^{\mathcal{L}(h_t^I)} (z^s_{t-1|\mathcal{P}(h_t^I)}(\mathcal{L}(h_t^I)))_{s\in S} \right)^s, \forall s \in S \quad (21)$$

$$z^s_{t|h_t^I} \geq \mathbf{0}, \forall s \in S \quad (22)$$

The initial condition $z_{0|\mathcal{P}(h_t^1)}(\mathcal{L}(h_1^I)) = p$, and $Q^{\emptyset}$ is an identity matrix of proper dimension. The behavior strategy $x_1^s = \frac{z_1^s}{\mathbf{1}^T z_1^s}$.

Therefore, Theorem 5.1 still holds for n-stage games, and hence complete the proof.

REFERENCES

[1] J. C. Harsanyi, "Games with Incomplete Information Played by "BAyesian" Players," *Part I,II, III, Management Science*, vol. 14, no. 5, pp. 159–182,320–334,486–502, 1967-8.

[2] R. Aumann and M. Maschler, "Repeated Games of Incomplete Information, The Zero-Sum Extensive Case," *Reports ST-143*, pp. 37–116, 1968.

[3] R. J. Aumann and M. Maschler, *Repeated games with incomplete information*. MIT press, 1995.

[4] D. Rosenberg, E. Solan, and N. Vieille, "Stochastic Games with A Single Controller and Incomplete Information," *SIAM Journal on Control and Optimization*, vol. 43, no. 1, pp. 86–110, 2004.

[5] J. Renault, "The Value of MArkov Chain Games with Lack of Information on One Side," *Mathematics of Operations Research*, vol. 31, no. 3, pp. 490–512, 2006.

[6] A. Gilpin and T. Sandholm, "Solving Two-Person Zero-Sum Repeated Games of Incomplete Information," in *Proceedings of the 7th international joint conference on Autonomous agents and multiagent systems-Volume 2*. International Foundation for Autonomous Agents and Multiagent Systems, 2008, pp. 903–910.

[7] T. Sandholm, "The State of Solving Large Incomplete-Information Games, and Application to Poker," *AI Magazine*, vol. 31, no. 4, pp. 13–32, 2010.

[8] D. Koller, N. Megiddo, and B. Von Stengel, "Efficient Computation of Equilibria for Extensive Two-Person Games," *Games and Economic Behavior*, vol. 14, no. 2, pp. 247–259, 1996.

[9] J. Shi and M. L. Littman, "Abstraction Methods for Game Theoretic Poker," in *Computers and Games*. Springer, 2001, pp. 333–345.

[10] A. Gilpin, S. Hoda, J. Pena, and T. Sandholm, "Gradient-Based Algorithms for Finding Nash Equilibria in Extensive Form Games," in *Internet and Network Economics*. Springer, 2007, pp. 57–69.

[11] M. Johanson, N. Bard, N. Burch, and M. Bowling, "Finding Optimal Abstract Strategies in Extensive-Form Games," in *AAAI*, 2012.

[12] B. De Meyer and D. Rosenberg, ""Cav u" and The Dual Game," *Mathematics of Operations Research*, vol. 24, no. 3, pp. 619–626, 1999.

[13] S. Sorin, *A First Course on Zero-Sum Repeated Games*. Springer Science & Business Media, 2002, vol. 37.

[14] L. Li and J. Shamma, "LP Formulation of Asymmetric Zero-Sum Stochastic Games," in *Decision and Control (CDC), 2014 IEEE Annual Conference on*. IEEE, 2014.

[15] J.-P. Ponssard and S. Sorin, "The LP Formulation of Finite Zero-Sum Games with Incomplete Information," *International Journal of Game Theory*, vol. 9, no. 2, pp. 99–105, 1980.

[16] T. Alpcan and T. Başar, *Network Security: A Decision and Game-theoretic Approach*. Cambridge University Press, 2010.