Stochastic Stackelberg security games

Deepanshu Vasal[1] [0000−0003−1089−8080]

[1]Department of Electrical and Computer Engineering,
University of Texas, Austin, Texas, 78704, USA
{dvasal} at utexas.edu

Abstract. In this paper, we consider a stochastic Stackelberg games where there is a defender (also called leader) who has to defend a target and an attacker (also called follower). The attacker has a private type that evolves as a controlled Markov process. The objective is to compute Stochastic Stackelberg equilibrium of the game where defender commits to a strategy. The attacker's strategy is the best response to defender strategy and defender's strategy is optimum given attacker plays best response. In general computing such equilibrium involves solving a fixed-point equation for the whole game. In this paper, we present an algorithm that computes such strategies by solving smaller fixed-point equations for each time t. This reduces the computational complexity of the problem from double exponential in time to linear in time. Based on this algorithm, we compute Stackelberg equilibrium of a security example.

1 Introduction

In the past decade, Stackelberg games have been used extensively in security of real world systems such as to protect ports, airports and wildlife [2,11,6,3]. In such games, there is a defender and an attacker, where defender commits to a strategy that is observable to the attacker. The attacker then plays a best response to attacker's strategy to maximize its utility. Knowing that the attacker will play a best response, the defender commits to and plays a strategy that maximizes its utility. Such pair of strategies is called a Stackelberg equilibrium. It is known that such strategies can provide higher utility to the defender than obtained in a Nash equilibrium of the game. Such games are currently in use by security agencies such as the US Coast Guard, the Federal Air Marshals Service, and the Los Angeles Airport Police [12]. Similar algorithms are used in wildlife protection in Uganda and Malaysia [16].

Most of the above real world applications of Stackelberg equilibrium are based on single-shot Bayesian game models. However, in many practical scenarios, the attacker and defender interact periodically, thus reducing the applicability of such models. Solving a dynamic stochastic Stackelberg game when the attacker has a private Markovian state is computationally challenging. This is because unlike other games, in such dynamic games of asymmetric information, there is coupling of players' strategies across time, rendering the complexity of finding equilibria of such games as double exponential in time. Recently, there has been results on sequential decomposition of certain classes of games of asymmetric information [15,13,14]. In repeated Stackelberg security games, there have been other approaches to mitigate this issue. For instance Kar et

al in [5] consider a repeated Stackelberg game and use a new human behavior model to study such games. Mareki et.al. in [7] study a Bayesian repeated Stackelberg game where they assume defenders are myopic, thus significantly simplifying the analysis of finding the equilibrium. Balcan et al in [1] consider a learning theoretic approach to study a repeated stackelberg game between attacker and defender where they use regret analysis to learn attacker's types, and show sub linear regret for both complete and partial information models.

In this paper, we provide a sequential decomposition algorithm for Stochastic dynamic Stackelberg games to compute equilibria with fully rational attacker and defender. Our algorithm reduces the complexity of finding Markovian equilibria of such games from double exponential to linear in time. Based on this algorithm, we study a security game where we numerically find its Stackelberg equilibria.

2 Model

We consider a stochastic Stackelberg game over a time horizon $[T] \stackrel{\triangle}{=} \{1, 2, \dots T\}$ with perfect recall as follows. Suppose there are two kinds of players: a leader and a follower. The leader has no private type, whereas follower privately observes a state $x_t \in \mathcal{X}$ at time t, where x_t evolves as a controlled Markov process in the following way,

$$P(x_t|a_{1:t-1}, x_{1:t-1}) = Q(x_t|a_{t-1}, x_{t-1}), \tag{1a}$$

where $a_t = (a_t^l, a_t^f)$ and Q are known kernels. Leader takes action $a_t^l \in \mathcal{A}^l$ at time t on observing $a_{1:t-1}$, which is common information among players, and the follower takes action $a_t^f \in \mathcal{A}^f$ at time t on observing $a_{1:t-1}$ and $x_{1:t}$ which it observes privately. The sets $\mathcal{A}^l, \mathcal{A}^f, \mathcal{X}$ are assumed to be finite. Let $\sigma^i = (\sigma_t^i)_{t\in[T]}$ be a probabilistic strategy of player $i \in \{l, f\}$ where $\sigma_t^l : \mathcal{A}^{t-1} \to \mathcal{P}(\mathcal{A}^l)$ such that the leader plays action A_t^l according to $A_t^l \sim \sigma_t^l(\cdot|a_{1:t-1})$. Similarly $\sigma_t^f : \mathcal{A}^{t-1} \times \mathcal{X}^t \to \mathcal{P}(\mathcal{A}^f)$ such that the follower plays action A_t^f according to $A_t^f \sim \sigma_t^f(\cdot|a_{1:t-1}, x_{1:t}^f)$. Let $\sigma \stackrel{\triangle}{=} (\sigma^i)_{i\in\{l,f\}}$ be a strategy profile of all players. At the end of interval t, the leader receives an instantaneous reward $R_t^l(x_t, a_t^l, a_t^f)$ and the follower receives an instantaneous reward $R_t^f(x_t, a_t^l, a_t^f)$. Suppose players discount their rewards by a discount factor $\delta \leq 1$.

3 Preliminaries

3.1 Stackelberg Equilibrium

The stackelberg equilibrium is defined for a game as follows. For a given strategy profile of the leader, σ^l, the follower maximizes its total discounted expected utility over finite horizon T,

$$\max_{\sigma^f} \mathbb{E}^{\sigma^l,\sigma^f} \left\{ \sum_{t=1}^{T} \delta^{t-1} R_t^f(X_t, A_t) \right\}. \tag{2}$$

Let $BR^f(\sigma^l)$ be the set of optimizing strategies of the follower given a strategy σ^l of the leader, i.e.

$$BR^f(\sigma^l) = \arg\max_{\sigma^f} \mathbb{E}^{\sigma^l,\sigma^f}\left\{\sum_{t=1}^{T} \delta^{t-1} R_t^f(X_t, A_t)\right\} \tag{3}$$

The leader finds its optimal strategy that maximizes its total expected discounted reward given that the follower will use its best response to it,

$$\tilde{\sigma}^l \in \max_{\sigma^l} \mathbb{E}^{\sigma^l, BR^f(\sigma^l)}\left\{\sum_{t=1}^{T} \delta^{t-1} R_t^l(X_t, A_t)\right\}, \tag{4}$$

Then $(\tilde{\sigma}_t^l, \tilde{\sigma}^f)$ constitute a Stackelberg equilibrium where $\tilde{\sigma}^f \in BR^f(\tilde{\sigma}^l)$.

3.2 Perfect Stackelberg equilibrium

In this paper, we will consider follower's equilibrium policies that only depend on its current state x_t i.e. at equilibrium, $a_t^f \sim \tilde{\sigma}_t(\cdot|a_{1:t-1}, x_t)$.[1]

For the game considered, we introduce a notion of Perfect Stackelberg Equilibrium (PSE), inspired by perfect Bayesian equilibrium [4] as follows.

Let $(\tilde{\sigma}, \mu)$ be a PSE of the game, where $\mu = (\mu_t)_{t\in[T]}$, and for any $t, a_{1:t-1}$, $\mu_t[a_{1:t-1}] \in \mathcal{P}(\mathcal{X})$ is the equilibrium belief on the current follower's state x_t, given the action history $a_{1:t-1}$, i.e. $\mu_t[a_{1:t-1}](x_t) = P^{\tilde{\sigma}}(x_t|a_{1:t-1})$. Then for all $t \in [T]$, $h_t^c = a_{1:t-1}$, $h_t^f = (a_{1:t-1}, x_{1:t})$, for any given σ^l, let $BR_t^f(\sigma^l)$ be defined as, $\forall h_t^f$

$$BR_t^f(\sigma^l) := \arg\max_{\sigma^f} \mathbb{E}^{\sigma^l,\sigma^f,\mu_t}\left\{\sum_{n=t}^{T} \delta^{n-t} R_n^l(X_n, A_n)|h_t^f\right\} \tag{5}$$

and

$$\tilde{\sigma}^l \in \max_{\sigma^l} \mathbb{E}^{\sigma^l, BR^f(\sigma^l),\mu_t}\left\{\sum_{n=t}^{T} \delta^{n-t} R_n^l(X_n, A_n)|h_t^c\right\}, \tag{6}$$

Then $(\tilde{\sigma}^l, \tilde{\sigma}^f)$ constitute a PSE of the game where $\tilde{\sigma}^f \in BR_t^f(\tilde{\sigma}^l)\ \forall\, t \in [T]$.

4 Common agent approach

We recall that the leader and the follower generate their actions at time t as follows, $a_t^l \sim \sigma_t^l(\cdot|a_{1:t-1})$ and $a_t^f \sim \sigma_t^f(\cdot|a_{1:t-1}, x_{1:t})$ An alternative way to view the problem is as follows. As is done in common information approach [8], at time t, a fictitious common agent observes the common information $a_{1:t-1}$ and generates prescription functions $\gamma_t = (\gamma_t^l, \gamma_t^f) = \psi_t[a_{1:t-1}]$. Player i uses its prescription function γ_t^i to

[1] Note, however, that for the purpose of equilibrium, the optimization will be performed in the space of all possible strategies that may depend on the entire history of state.

operate on its private information (if any) to produce its action a_t^i, i.e. $\gamma_t^l \in \mathcal{P}(\mathcal{A}^i)$ and $\gamma_t^f : \mathcal{X}^t \to \mathcal{P}(\mathcal{A}^i)$ and $a_t^l \sim \gamma_t^l(\cdot)$ and $a_t^f \sim \gamma_t^f(\cdot|x_{1:t})$. It is easy to see that for any σ policy profile of the players, there exists an equivalent ψ profile of the common agent (and vice versa) that generates the same control actions for every realization of the information of the players.

Here, we will consider Markovian common agent's policy as follows. We call a common agent's policy be of "type θ" if the common agent observes the common belief π_t derived from the common observation $a_{1:t-1}$, and generates prescription functions $\gamma_t = (\gamma_t^l, \gamma_t^f) = \theta_t[\pi_t]$, where π_t is a belief on the current state x_t defined as, $\pi_t(x_t) = P^\theta(x_t|a_{1:t-1})$. The follower uses these prescription function γ_t^f to operates on its current private type x_t to produce its action a_t^f, i.e. $\gamma_t^f : \mathcal{X} \to \mathcal{P}(\mathcal{A}^f)$ and $a_t^f \sim \gamma_t^f(\cdot|x_t)$. Similarly, the leader uses prescription function γ_t^l to produce its action a_t^l as $\gamma_t^l \in \mathcal{P}(\mathcal{A}^f)$ and $a_t^l \sim \gamma_t^f(\cdot)$.

In the next lemma we show that for any given θ policy, the belief states π_t can be updated recursively as follows. Let $\pi_1(x) := Q(x)$.

Lemma 1. *For any given policy of type θ, there exists update functions F, independent of θ, such that*

$$\pi_{t+1} = F(\pi_t, \gamma_t^f, a_t) \tag{7}$$

Proof. Please see Appendix A.

Definition 1. *We call a strategy profile σ Markov PSE (MPSE), if it is a PSE of type θ.*

In the next section, we design an algorithm to compute MPSE of the game.

5 Algorithm for MPSE computation

5.1 Backward Recursion

In this section, we define an equilibrium generating function $\theta = (\theta_t^i)_{i\in\{l,f\},t\in[T]}$, where $\theta_t : \mathcal{P}(\mathcal{X}) \to \{\mathcal{X} \to \mathcal{P}(\mathcal{A}^f)\} \times \mathcal{P}(\mathcal{A}^l)$ and a sequence of functions $(V_t^l, V_t^f)_{t\in\{1,2,\dots T+1\}}$, where $V_t^l : \mathcal{P}(\mathcal{X}) \to \mathbb{R}, V_t^f : \mathcal{P}(\mathcal{X}) \times \mathcal{X} \to \mathbb{R}$, in a backward recursive way, as follows.

1. Initialize $\forall \pi_{T+1} \in \mathcal{P}(\mathcal{X}), x_{T+1} \in \mathcal{X}$,

$$V_{T+1}^l(\pi_{T+1}) \stackrel{\triangle}{=} 0. \tag{8}$$

$$V_{T+1}^f(\pi_{T+1}, x_{T+1}) \stackrel{\triangle}{=} 0 \tag{9}$$

2. For $t = T, T-1, \dots 1,\ \forall \pi_t \in \mathcal{P}(\mathcal{X})$, let $\theta_t[\pi_t]$ be generated as follows. Set $\tilde{\gamma}_t = \theta_t[\pi_t]$, where $\tilde{\gamma}_t = (\tilde{\gamma}_t^l, \tilde{\gamma}_t^f)$ is the solution of the following fixed-point equation. For a given γ_t^l, define $BR(\gamma_t^l)$ as follows, $\forall x_t \in \mathcal{X}$,

$$BR_t^f(\gamma_t^l) = \Bigg\{ \tilde{\gamma}_t^f : \tilde{\gamma}_t^f \in \arg\max_{\gamma_t^f(\cdot|x_t)} \mathbb{E}^{\gamma_t^f(\cdot|x_t)\gamma_t^l,\, \pi_t} \Big\{ R_t^f(x_t, A_t) + \delta V_{t+1}^f(F(\pi_t, \tilde{\gamma}_t^f, A_t), X_{t+1}) \big| x_t \Big\} \Bigg\}, \tag{10}$$

where expectation in (21) is with respect to random variables (A_t, X_{t+1}) through the measure $\gamma_t^f(a_t^f|x_t)\gamma_t^l(a_t^l)Q(x_{t+1}|x_t, a_t)$ and F is defined in (7).
Then let $(\tilde{\gamma}_t^l, \tilde{\gamma}_t^f)$ is a solution of the following fixed-point equation,

$$\tilde{\gamma}_t^f \in BR_t^f(\tilde{\gamma}_t^l) \tag{11}$$

and

$$\tilde{\gamma}_t^l \in \arg\max_{\gamma_t^l} \mathbb{E}^{BR_t^f(\gamma_t^l)\gamma_t^l, \pi_t}\left\{R_t^l(X_t, A_t) + \delta V_{t+1}^l(F(\pi_t, BR_t^f(\gamma_t^l), A_t))\right\} \tag{12}$$

where the above expectation is defined with respect to random variables (X_t, A_t) through the measure $\pi_t(x_t)\hat{\gamma}_t^f(a_t^f)\gamma_t^l(a_t^l)$, and $\hat{\gamma}_t \in BR_t^f(\gamma_t^l)$.
Let $(\tilde{\gamma}_t^l, \tilde{\gamma}_t^f)$ be a pair of solution of the above operation. Then set $\forall x_t \in \mathcal{X}$,

$$V_t^f(\pi_t, x_t) \stackrel{\triangle}{=} \mathbb{E}^{\tilde{\gamma}_t^f(\cdot|x_t)\tilde{\gamma}_t^l, \pi_t}\left\{R_t^f(x_t, A_t) + \delta V_{t+1}^f(F(\pi_t, \tilde{\gamma}_t^f, A_t), X_{t+1})\big|x_t\right\}. \tag{13}$$

and

$$V_t^l(\pi_t) \stackrel{\triangle}{=} \mathbb{E}^{\tilde{\gamma}_t^f\tilde{\gamma}_t^l, \pi_t}\left\{R_t^l(X_t, A_t) + \delta V_{t+1}^l(F(\pi_t, \tilde{\gamma}_t^f, A_t))\right\}. \tag{14}$$

5.2 Forward Recursion

Based on θ defined in the backward recursion above, we now construct a set of strategies $\tilde{\sigma}$ (through beliefs μ) in a forward recursive way as follows.

1. Initialize at time $t = 1$,

$$\mu_1[\phi](x_1) := Q(x_1). \tag{15}$$

2. For $t = 1, 2 \ldots T, \forall i = 1, 2, a_{1:t} \in \mathcal{H}_{t+1}^c, x_{1:t} \in \mathcal{X}^t$

$$\tilde{\sigma}_t^l(a_t^l|a_{1:t-1}) := \theta_t^l[\mu_t[h_t^c]](a_t^l) \tag{16}$$

$$\tilde{\sigma}_t^f(a_t^f|a_{1:t-1}, x_{1:t}) := \theta_t^f[\mu_t[h_t^c]](a_t^f|x_t) \tag{17}$$

$$\mu_{t+1}[h_{t+1}^c] := F(\mu_t[h_t^c], \theta_t^f[\mu_t[h_t^c]], a_t) \tag{18a}$$

where F is defined in (7).

Theorem 1. *A strategy profile $\tilde{\sigma}$, as constructed through backward/forward recursion algorithm above is an MPSE of the game*

Proof. We will prove this theorem in two parts. In Part 1 in Appendix B for the follower, we prove that $\tilde{\sigma}^f \in BR_t^f(\tilde{\sigma}^l)$ i.e. $\forall\, t \in [T], \forall \sigma^f, h_t^f = (a_{1:t-1}, x_{1:t})$

$$\mathbb{E}^{\tilde{\sigma}^l, \tilde{\sigma}^f, \mu_t}\left\{\sum_{n=t}^{T} \delta^{n-t} R_n^f(X_n, A_n)|a_{1:t-1}, x_{1:t}\right\} \geq$$
$$\mathbb{E}^{\tilde{\sigma}^l, \sigma^f, \mu_t}\left\{\sum_{n=t}^{T} \delta^{n-t} R_n^f(X_n, A_n)|a_{1:t-1}, x_{1:t}\right\}. \tag{19}$$

In Part 2 in Appendix D for the leader, we show that

$$\mathbb{E}^{\tilde{\sigma}^l,\tilde{\sigma}^f,\mu_t[a_{1:t-1}]}\left\{\sum_{n=t}^{T}\delta^{n-t}R_n^l(X_n,A_n)|a_{1:t-1}\right\}\geq$$
$$\mathbb{E}^{\sigma^l,BR^f(\sigma^l),\mu_t[a_{1:t-1}]}\left\{\sum_{n=t}^{T}\delta^{n-t}R_n^l(X_n,A_n)|a_{1:t-1}\right\}, \tag{20}$$

where $\tilde{\sigma}^f \in BR^f(\tilde{\sigma}^l)$, as shown in Part 1.

Combining both the parts prove the above result.

6 Infinite horizon

The above results can be extended to infinite horizon case when the reward functions R^l, R^f are time homogenous and are absolutely bounded, and $\delta < 1$. In the following, due to space constraints, we just state the algorithm without proof.

6.1 Backward Recursion

Define an equilibrium generating function $\theta = (\theta^i)_{i\in\{l,f\}}$, where $\theta : \mathcal{P}(\mathcal{X}) \to \left\{\mathcal{X} \to \mathcal{P}(\mathcal{A}^f)\right\}\times \mathcal{P}(\mathcal{A}^l)$ and a vector of functions (V^l, V^f), where $V^l : \mathcal{P}(\mathcal{X}) \to \mathbb{R}, V^f : \mathcal{P}(\mathcal{X})\times\mathcal{X} \to \mathbb{R}$ through the following one-shot fixed-point equation.

Let $\theta[\pi]$ be generated as follows. Set $\tilde{\gamma} = \theta[\pi]$, where $\tilde{\gamma} = (\tilde{\gamma}^l, \tilde{\gamma}^f)$ is the solution of the following fixed-point equation. For a given γ^l, define $BR(\gamma^l)$ as follows, $\forall x \in \mathcal{X}$,

$$BR^f(\gamma^l) = \Bigg\{\tilde{\gamma}^f : \tilde{\gamma}^f \in \arg\max_{\gamma^f(\cdot|x)} \mathbb{E}^{\gamma^f(\cdot|x)\gamma^l,\pi}\left\{R^f(x,A) + \delta V^f(F(\pi,\tilde{\gamma}^f,A),X')\big|x\right\}\Bigg\}, \tag{21}$$

where expectation in (21) is with respect to random variables (A, X') through the measure $\gamma^f(a^f|x)\gamma^l(a^l)Q(x'|x,a)$ and F is defined in (7).

Then let $(\tilde{\gamma}^l, \tilde{\gamma}^f, V^l, V^f)$ be a solution of the following fixed-point equation,

$$\tilde{\gamma}^f \in BR^f(\tilde{\gamma}^l) \tag{22}$$

and

$$\tilde{\gamma}^l \in \arg\max_{\gamma^l} \mathbb{E}^{BR^f(\gamma^l)\gamma^l,\pi}\left\{R^l(X,A) + \delta V^l(F(\pi, BR^f(\gamma^l),A))\right\} \tag{23}$$

where the above expectation is defined with respect to random variables (X, A, X') through the measure $\pi(x)\hat{\gamma}^f(a^f)\gamma^l(a^l)Q(x'|x,a)$, and $\hat{\gamma} \in BR^f(\gamma^l)$. And $\forall x \in \mathcal{X}$,

$$V^f(\pi,x) = \mathbb{E}^{\tilde{\gamma}^f(\cdot|x)\tilde{\gamma}^l,\pi}\left\{R^f(x,A) + \delta V^f(F(\pi,\tilde{\gamma}^f,A),X')\big|x\right\}. \tag{24}$$

and

$$V^l(\pi) = \mathbb{E}^{\tilde{\gamma}^f\tilde{\gamma}^l,\pi}\left\{R^l(X,A) + \delta V^l(F(\pi,\tilde{\gamma}^f,A))\big|x\right\}. \tag{25}$$

6.2 Forward Recursion

Based on θ defined above, we now construct a set of strategies $\tilde{\sigma}$ (through beliefs μ) in a forward recursive way as follows.

1. Initialize at time $t = 1$,

$$\mu_1[\phi](x_1) := Q(x_1). \tag{26}$$

2. For $t = 1, 2 \ldots, \forall i = 1, 2, a_{1:t} \in \mathcal{H}^c_{t+1}, x_{1:t} \in \mathcal{X}^t$

$$\tilde{\sigma}^l_t(a^l_t|a_{1:t-1}) := \theta^l[\mu_t[h^c_t]](a^l_t) \tag{27}$$

$$\tilde{\sigma}^f_t(a^f_t|a_{1:t-1}, x_{1:t}) := \theta^f[\mu_t[h^c_t]](a^f_t|x_t) \tag{28}$$

$$\mu_{t+1}[h^c_{t+1}] := F(\mu_t[h^c_t], \theta^f[\mu_t[h^c_t]], a_t) \tag{29}$$

where F is defined in (7).

6.3 Existence of ϵ-equilibrium

In this section, we prove existence of a $\epsilon-$solution of fixed-point equations (11),(12), defined below.

Lemma 2. *For all π_t,γ^l_t, $BR^f_t(\gamma^l_t)$ as defined in* (21) *is non-empty.*

Proof. Please see Appendix F.

Theorem 2. *There exists a $\tilde{\gamma}^l_t$ that is an $\epsilon-$maximizer of* (12)*,*

Proof. Since $\mathcal{X}, \mathcal{A}$ are finite, $R(x, a)$ is absolutely bounded, say by a constant M. Then V_{t+1} is absolutely bounded by $(T-t-1)M$. Since (12) involves optimizing a bounded (not necessarily continuous) function over a compact set, there always exists a $\tilde{\gamma}^l_t$ that is an $\epsilon-$maximizer of (12).

7 Complexity

In general, computing a Stackelberg equilibrium involves solving a fixed-point equation in the space of strategies of both the players for all histories of the game i.e. of the form $\sigma = f(\sigma)$ where f is appropriately defined from (4). For any time t, since $\sigma^f_t : \mathcal{A}^{t-1} \times \mathcal{X}^t \to \mathcal{P}(\mathcal{A}^f)$ and $\sigma^l_t : \mathcal{A}^{t-1} \to \mathcal{P}(\mathcal{A}^l)$, there exist $|\mathcal{P}(\mathcal{A}^l)|^{|\mathcal{A}|^{t-1} \times |\mathcal{X}|^t}$ number of possible strategies of the leader and $|\mathcal{P}(\mathcal{A})|^{|\mathcal{A}|^{t-1}}$ number of strategies of the follower. Since the complexity at the last time t dominates, solving a stackelberg equilibrium reduces to solving a fixed-point equation in the space of $|\mathcal{P}(\mathcal{A}^l)|^{|\mathcal{A}|^{T-1} \times |\mathcal{X}|^T} \times |\mathcal{P}(\mathcal{A}^l)|^{|\mathcal{A}|^{T-1}}$ number of strategies.

In our algorithm, each time t involves solving a fixed-point equation (11),(12), for every π_t, where $\pi_t \in \mathcal{P}(\mathcal{X})$. Thus computing a Stackelberg equilibrium involves solving $T|\mathcal{P}(\mathcal{X})|$ "smaller" fixed-point equations (11) (12). Therefore, our algorithm reduces the computational dependence on T from double exponential to linear. The complexity of solving each smaller fixed-point equation depends on the specific model parameters and is an important direction for future research.

8 Security Example

In this section, we consider a repeated Stackelberg game as a security example. We assume that $\mathcal{X} = \mathcal{A}^l = \mathcal{A}^f = \{0, 1\}$ and type of the defender is static i.e. $Q(x_{t+1}|x_t, a_t) = \mathbb{1}(x_{t+1} = x_t)$. We assume $\delta = 0.6$. Let $p^l = \gamma^l(1)$, $p^{f,0} = \gamma^f(1|0)$ and $p^{f,1} = \gamma^f(1|1)$ and the rewards of the players are given in Table I below.

Table 1

X=0	Attacker $A1$	Attacker $A2$
Defender $D1$	$(2, 1)$	$(4, 0)$
Defender $D2$	$(1, 0)$	$(3, 2)$

X=1	Attacker $A1$	Attacker $A2$
Defender $D1$	$(3, 2)$	$(2, 0)$
Defender $D2$	$(0, 1)$	$(1, 1)$

The equilibrium strategies and value functions are provided in Figures 1-6.

9 Conclusion

In this paper, we study a general leader/defender, follower/attacker security game where the attacker has a private type what evolves a controlled Markov process. We present a novel dynamic programing like methodology to sequentially decompose the problem of computing Markov perfect Stackelberg equilibrium for these games. Based on this algorithm we study a repeated security game where we numerically compute the equilibrium policies. We believe this result can be extended to the case when the leader also has a private type, which we propose as future work.

A

Proof.

$$\pi_{t+1}(x_{t+1}) = P^{\theta}(x_{t+1}|a_{1:t}) \tag{30}$$

$$= \frac{\sum_{x_t} P^{\theta}(x_t, a_t, x_{t+1}|a_{1:t-1})}{\sum_{x_t} P^{\theta}(x_t, a_t|a_{1:t-1})} \tag{31}$$

$$= \frac{\sum_{x_t} \pi_t(x_t)\gamma_t^f(a_t^f|x_t)\gamma_t^l(a_t^l)Q(x_{t+1}|x_t, a_t)}{\sum_{x_t} \pi_t(x_t)\gamma_t^f(a_t^f|x_t)\gamma_t^l(a_t^l)} \tag{32}$$

$$= \frac{\sum_{x_t} \pi_t(x_t)\gamma_t^f(a_t^f|x_t)Q(x_{t+1}|x_t, a_t)}{\sum_{x_t} \pi_t(x_t)\gamma_t^f(a_t^f|x_t)} \tag{33}$$

where $\gamma_t^l(a_t^l)$ cancels out in the numerator and the denominator. Thus,

$$\pi_{t+1} = F(\pi_t, \gamma_t^f, a_t) \tag{34}$$

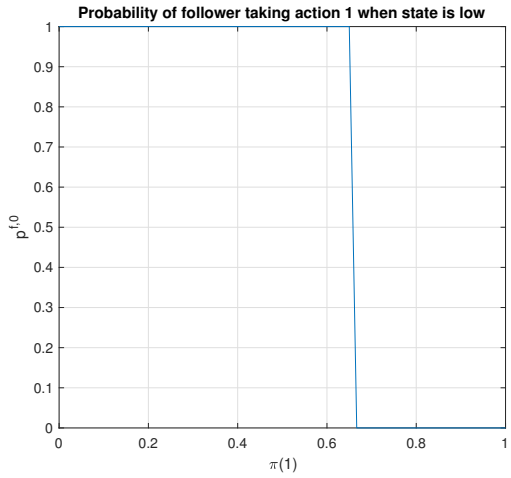

Fig. 1: Probability of follower taking action 1 when its state is low

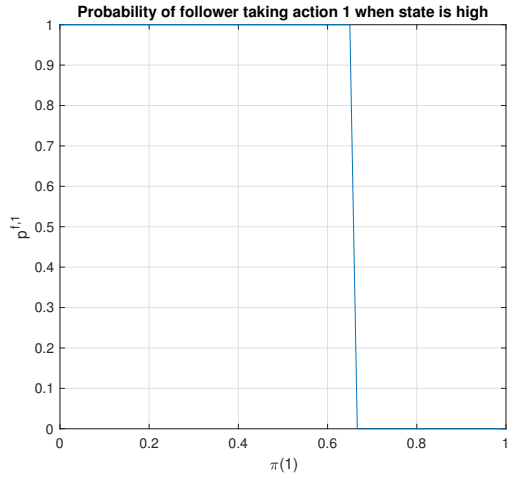

Fig. 2: Probability of follower taking action 1 when its state is high

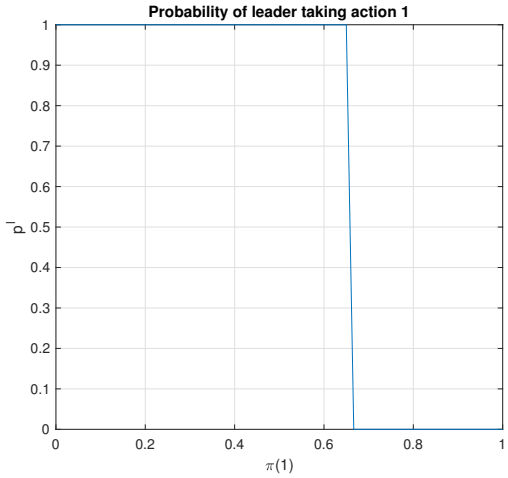

Fig. 3: Probability of leader taking action 1

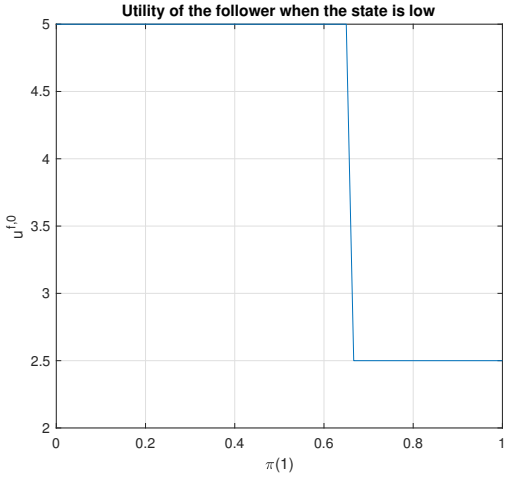

Fig. 4: Utility of follower when its state is low

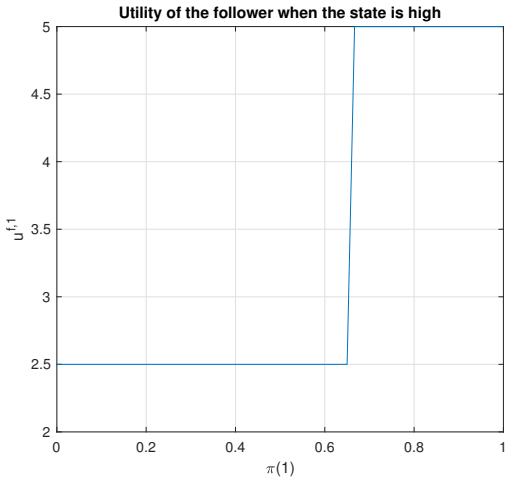

Fig. 5: Utility of follower when its state is high

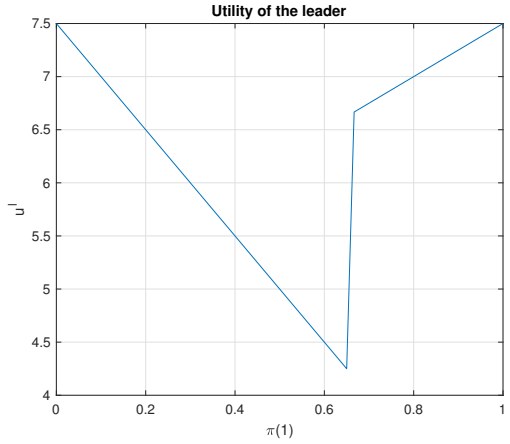

Fig. 6: Utility of leader

B Part 1: Follower

Proof. In the following theorem, we will prove that $\tilde{\sigma}^f \in BR_t^f(\tilde{\sigma}^l)\ \forall\ t \in [T]$, i.e. $\forall \sigma^f, h_t^f = (a_{1:t-1}, x_{1:t})$

$$\mathbb{E}^{\tilde{\sigma}^l, \tilde{\sigma}^f, \mu_t} \left\{ \sum_{n=t}^{T} \delta^{n-t} R_n^f(X_n, A_n) | a_{1:t-1}, x_{1:t} \right\} \geq$$
$$\mathbb{E}^{\tilde{\sigma}^l, \sigma^f, \mu_t} \left\{ \sum_{n=t}^{T} \delta^{n-t} R_n^f(X_n, A_n) | a_{1:t-1}, x_{1:t} \right\} \tag{35}$$

We prove the above result using induction and from results in Lemma 3, 4 and 5 proved in Appendix C.

For base case at $t = T$, $\forall (a_{1:T-1}, x_{1:T}) \in \mathcal{H}_T^f, \sigma^f$

$$\mathbb{E}^{\tilde{\sigma}_T^f \tilde{\sigma}_T^l, \mu_T[a_{1:T-1}]} \left\{ R_T^f(X_T, A_T) \big| a_{1:T-1}, x_{1:T} \right\}$$
$$= V_T^f(\mu_t[a_{1:T-1}], x_T) \tag{36a}$$
$$\geq \mathbb{E}^{\sigma_T^f \tilde{\sigma}_T^l, \mu_T[a_{1:T-1}]} \left\{ R_T^f(X_T, A_T) \big| a_{1:T-1}, x_{1:T} \right\}. \tag{36b}$$

where (36a) follows from Lemma 5 and (36b) follows from Lemma 3 in Appendix C. Let the induction hypothesis be that for $t+1$, $\forall (a_{1:t}, x_{1:t+1}) \in \mathcal{H}_{t+1}^f, \sigma^f$,

$$\mathbb{E}^{\tilde{\sigma}_{t+1:T}^f \tilde{\sigma}_{t+1:T}^l, \mu_{t+1}[a_{1:t}]} \left\{ \sum_{n=t+1}^{T} R_n^f(X_n, A_n) \big| a_{1:t}, x_{1:t+1} \right\}$$
$$\geq \mathbb{E}^{\sigma_{t+1:T}^f \tilde{\sigma}_{t+1:T}^l, \mu_{t+1}[a_{1:t}]} \left\{ \sum_{n=t+1}^{T} R_n^f(X_n, A_n) \big| a_{1:t}, x_{1:t+1} \right\} \tag{37a}$$

Then $\forall (a_{1:t-1}, x_{1:t}) \in \mathcal{H}_T^f, \sigma^f$, we have

$$\mathbb{E}^{\tilde{\sigma}_{t:T}^f \tilde{\sigma}_{t:T}^l, \mu_t[a_{1:t-1}]} \left\{ \sum_{n=t}^{T} R_n^f(X_n, A_n) \big| a_{1:t-1}, x_{1:t} \right\}$$
$$= V_t^f(\mu_t[a_{1:t-1}], x_t) \tag{38a}$$
$$\geq \mathbb{E}^{\sigma_t^f \tilde{\sigma}_t^l, \mu_t[a_{1:t-1}]} \left\{ R_t^f(X_t, A_t) + V_{t+1}^f(\mu_{t+1}[a_{1:t-1} A_t], X_{t+1}) \big| a_{1:t-1}, x_{1:t} \right\} \tag{38b}$$
$$= \mathbb{E}^{\sigma_t^f \tilde{\sigma}_t^l, \mu_t[a_{1:t-1}]} \Big\{ R_t^f(X_t, A_t) +$$
$$\mathbb{E}^{\tilde{\sigma}_{t+1:T}^f \tilde{\sigma}_{t+1:T}^l, \mu_{t+1}[a_{1:t-1}, A_t]} \left\{ \sum_{n=t+1}^{T} R_n^f(X_n, A_n) \ \big| a_{1:t-1}, A_t, x_{1:t}, X_{t+1} \right\} \big| a_{1:t-1}, x_{1:t} \Big\} \tag{38c}$$
$$\tag{38d}$$

$$\geq \mathbb{E}^{\sigma_t^f \tilde{\sigma}_t^l,\, \mu_t[a_{1:t-1}]} \Big\{ R_t^f(X_t, A_t) + \mathbb{E}^{\sigma_{t+1:T}^f \tilde{\sigma}_{t+1:T}^l \mu_{t+1}[a_{1:t-1}, A_t]} \left\{ \sum_{n=t+1}^{T} R_n^f(X_n, A_n) \, \middle| a_{1:t-1}, A_t, x_{1:t}, X_{t+1} \right\} \middle| a_{1:t-1}, x_{1:t} \right\} \tag{38e}$$

$$= \mathbb{E}^{\sigma_t^f \tilde{\sigma}_t^l,\, \mu_t[a_{1:t-1}]} \Big\{ R_t^f(X_t, A_t) + \mathbb{E}^{\sigma_{t:T}^f \tilde{\sigma}_{t:T}^l \mu_t[a_{1:t-1}]} \left\{ \sum_{n=t+1}^{T} R_n^f(X_n, A_n) \, \middle| a_{1:t-1}, A_t, x_{1:t}, X_{t+1} \right\} \middle| a_{1:t-1}, x_{1:t} \right\} \tag{38f}$$

$$= \mathbb{E}^{\sigma_{t:T}^f \tilde{\sigma}_{t:T}^l \, \mu_t[a_{1:t-1}]} \left\{ \sum_{n=t}^{T} R_n^f(X_n, A_n) \middle| a_{1:t-1}, x_{1:t} \right\}, \tag{38g}$$

where (38a) follows from Lemma 5, (38b) follows from Lemma 3, (38c) follows from Lemma 5, (38e) follows from induction hypothesis in (37a) and (38f) follows from Lemma 4.

C

Lemma 3. $\forall t \in [T], (a_{1:t-1}, x_{1:t}) \in \mathcal{H}_t^f, \sigma_t^f$

$$V_t^f(\mu_t[a_{1:t-1}], x_t) \geq \mathbb{E}^{\sigma_t^f \tilde{\sigma}_t^l,\, \mu_t[a_{1:t-1}]} \Big\{ R_t^f(X_t, A_t) + V_{t+1}^f(F(\mu_t[a_{1:t-1}], \tilde{\sigma}_t^f(\cdot|a_{1:t-1}, \cdot), A_t), X_{t+1}) \big| a_{1:t-1}, x_{1:t} \Big\} \tag{39}$$

Proof. We prove this lemma by contradiction. Suppose the claim is not true for t. This implies $\exists i, \hat{\sigma}_t^f, \hat{a}_{1:t-1}, \hat{x}_{1:t}$ such that

$$\mathbb{E}^{\hat{\sigma}_t^f \tilde{\sigma}_t^l,\, \mu_t[\hat{a}_{1:t-1}]} \Big\{ R_t^f(X_t, A_t) + V_{t+1}^f(F(\mu_t[\hat{a}_{1:t-1}], \tilde{\sigma}_t(\cdot|\hat{a}_{1:t-1}, \cdot), A_t), X_{t+1}) \big| \hat{a}_{1:t-1}, \hat{x}_{1:t} \Big\}$$
$$> V_t^f(\mu_t[\hat{a}_{1:t-1}], \hat{x}_t). \tag{40}$$

We will show that this leads to a contradiction.

Construct $\hat{\gamma}_t^f(a_t^f|x_t) = \begin{cases} \hat{\sigma}_t^f(a_t^f|\hat{a}_{1:t-1}, \hat{x}_{1:t}) & x_t = \hat{x}_t \\ \text{arbitrary} & \text{otherwise.} \end{cases}$

Then for $\hat{a}_{1:t-1}, \hat{x}_{1:t}$, we have

$$V_t^f(\mu_t[\hat{a}_{1:t-1}], \hat{x}_t) \tag{41a}$$
$$= \max_{\gamma_t^f(\cdot|\hat{x}_t)} \mathbb{E}^{\gamma_t^f(\cdot|\hat{x}_t) \tilde{\sigma}_t^l,\, \mu_t[\hat{a}_{1:t-1}]} \Big\{ R_t^f(\hat{x}_t, a_t) + V_{t+1}^f(F(\mu_t[\hat{a}_{1:t-1}], \sigma_t(\cdot|\hat{a}_{1:t-1}, \cdot), A_t), X_{t+1}) \big| \hat{x}_t \Big\}, \tag{41b}$$

$$\geq \mathbb{E}^{\hat{\gamma}_t^f(\cdot|\hat{x}_t)\tilde{\sigma}_t^l,\,\mu_t[\hat{a}_{1:t-1}]}\Big\{R_t^f(X_t,A_t)+$$

$$V_{t+1}^f(F(\mu_t[\hat{a}_{1:t-1}],\sigma_t(\cdot|\hat{a}_{1:t-1},\cdot),A_t),X_{t+1})\big|\hat{x}_{1:t}\Big\} \tag{41c}$$

$$= \mathbb{E}^{\hat{\sigma}_t^f\tilde{\sigma}_t^l,\mu_t[\hat{a}_{1:t-1}]}\Big\{R_t^f(\hat{x}_t,a_t)+$$

$$V_{t+1}^f(F(\mu_t[\hat{a}_{1:t-1}],\sigma_t(\cdot|\hat{a}_{1:t-1},\cdot),A_t),X_{t+1})\big|\hat{a}_{1:t-1},\hat{x}_{1:t}\Big\} \tag{41d}$$

$$> V_t^f(\mu_t[\hat{a}_{1:t-1}],\hat{x}_t) \tag{41e}$$

where (41b) follows from definition of V_t^f in (24), (41d) follows from definition of $\hat{\gamma}_t^f$ and (41e) follows from (40). However this leads to a contradiction.

Lemma 4. *$\forall t\in[T],(a_{1:t},x_{1:t+1})\in\mathcal{H}_{t+1}^f$ and σ_t^f*

$$\mathbb{E}^{\sigma_{t:T}^f\tilde{\sigma}_{t:T}^l,\,\mu_t[a_{1:t-1}]}\left\{\sum_{n=t+1}^{T}R_t^f(X_n,A_n)\big|a_{1:t},x_{1:t+1}\right\}=$$

$$\mathbb{E}^{\sigma_{t+1:T}^f\tilde{\sigma}_{t+1:T}^l,\,\mu_{t+1}[a_{1:t}]}\left\{\sum_{n=t+1}^{T}R_n^f(X_n,A_n)\big|a_{1:t},x_{1:t+1}\right\}. \tag{42}$$

Proof. Since the above expectations involve random variables $A_{t+1:T},X_{t+2:T}$, we consider, $P^{\sigma_{t:T}^f\tilde{\sigma}_{t:T}^l,\,\mu_t[a_{1:t-1}]}(a_{t+1:T},x_{t+2:T}\big|a_{1:t},x_{1:t+1})$.

$$P^{\sigma_{t:T}^f\tilde{\sigma}_{t:T}^l,\,\mu_t[a_{1:t-1}]}(a_{t+1:T},x_{t+2:T}\big|a_{1:t},x_{1:t+1}) \tag{43a}$$

$$=\tilde{\sigma}_{t+1}^f(a_{t+1}^f|a_{1:t},x_{t+1})\tilde{\sigma}_{t+1}^l(a_{t+1}^l|a_{1:t})Q(x_{t+2}|x_{t+1},a_{t+1})$$

$$P^{\sigma_{t:T}^f\tilde{\sigma}_{t:T}^l,\,\mu_t[a_{1:t-1}]}(a_{t+2:T},x_{t+3:T}\big|a_{1:t+1},x_{1:t+2}) \tag{43b}$$

By definition, $\tilde{\sigma}_t^f,\tilde{\sigma}_t^l$ depend on $a_{1:t}$ through the common equilibrium belief $\mu_{t+1}[a_{1:t}]$, as defined in (28). Moreover, the probability on $(a_{t+2:T},x_{t+3:T})$ given $a_{1:t+1},x_{t:t+2}$ depend on $a_{1:t+1},x_{t:t+2},\mu_{t+1}[a_{1:t}]$ through $\sigma_{t+2:T}^f\tilde{\sigma}_{t+2:T}^l$.

Thus,

$$P^{\sigma_{t:T}^f\tilde{\sigma}_{t:T}^l,\,\mu_t[a_{1:t-1}]}(a_{t+1:T},x_{t+2:T}\big|a_{1:t},x_{1:t+1}) \tag{43c}$$

$$=P^{\sigma_{t+1:T}^f\tilde{\sigma}_{t+1:T}^l,\,\mu_{t+1}[a_{1:t}]}(a_{t+1:T},x_{t+2:T}\big|a_{1:t},x_{1:t+1}) \tag{43d}$$

Lemma 5. *$\forall t\in[T],(a_{1:t-1},x_{1:t})\in\mathcal{H}_t^f$*

$$V_t^f(\mu_t[a_{1:t-1}],x_t)=\mathbb{E}^{\tilde{\sigma}_{t:T}^f\tilde{\sigma}_{t:T}^l,\mu_t[a_{1:t-1}]}\left\{\sum_{n=t}^{T}R_n^f(X_n,A_n)\big|a_{1:t-1},x_{1:t}\right\}. \tag{44}$$

Proof. We prove the lemma by induction. For $t=T$,

$$\mathbb{E}^{\tilde{\sigma}_T^f\tilde{\sigma}_T^l,\,\mu_T[a_{1:T-1}]}\Big\{R_T^f(X_T,A_T)\big|a_{1:T-1},x_{1:T}\Big\}$$

$$=\sum_{x_T^l a_T}R_T^f(x_T,a_T)\tilde{\sigma}_T^f(a_T^f|a_{1:T-1},x_T)\tilde{\sigma}_T^l(a_T^l|a_{1:T-1}) \tag{45a}$$

$$=V_T^f(\mu_T[a_{1:T-1}],x_T), \tag{45b}$$

where (45b) follows from the definition of V_t^f in (24) and the definition of $\tilde{\sigma}_T$ in the forward recursion in (28).

Suppose the claim is true for $t+1$, i.e. $\forall t \in [T], (a_{1:t}, x_{1:t+1}) \in \mathcal{H}_{t+1}^f$

$$V_{t+1}^f(\mu_{t+1}[a_{1:t}], x_{t+1}) = \mathbb{E}^{\tilde{\sigma}_{t+1:T}^f \tilde{\sigma}_{t+1:T}^l, \mu_{t+1}[a_{1:t}]} \left\{ \sum_{n=t+1}^{T} R_n^f(X_n, A_n) \big| a_{1:t}, x_{1:t+1} \right\}. \tag{46}$$

Then $\forall t \in [T], (a_{1:t-1}, x_{1:t}) \in \mathcal{H}_t^f$, we have

$$\begin{aligned}
&\mathbb{E}^{\tilde{\sigma}_{t:T}^f \tilde{\sigma}_{t:T}^l, \mu_t[a_{1:t-1}]} \left\{ \sum_{n=t}^{T} R_n^f(X_n, A_n) \big| a_{1:t-1}, x_{1:t} \right\} \\
&= \mathbb{E}^{\tilde{\sigma}_{t:T}^f \tilde{\sigma}_{t:T}^l, \mu_t[a_{1:t-1}]} \Big\{ R_t^f(X_t, A_t) + \\
&\quad \mathbb{E}^{\tilde{\sigma}_{t:T}^f \tilde{\sigma}_{t:T}^l, \mu_t[a_{1:t-1}]} \left\{ \sum_{n=t+1}^{T} R_n^f(X_n, A_n) \, \big| a_{1:t-1}, A_t, x_{1:t}, X_{t+1} \right\} \big| a_{1:t-1}, x_{1:t} \Big\}
\end{aligned} \tag{47a}$$

$$\begin{aligned}
&= \mathbb{E}^{\tilde{\sigma}_{t:T}^f \tilde{\sigma}_{t:T}^l, \mu_t[a_{1:t-1}]} \Big\{ R_t^f(X_t, A_t) + \\
&\quad \mathbb{E}^{\tilde{\sigma}_{t+1:T}^f \tilde{\sigma}_{t+1:T}^l, \mu_{t+1}[a_{1:t-1}, A_t]} \left\{ \sum_{n=t+1}^{T} R_n^f(X_n, A_n) \, \big| a_{1:t-1}, A_t, x_{1:t}, X_{t+1} \right\} \big| a_{1:t-1}, x_{1:t} \Big\}
\end{aligned} \tag{47b}$$

$$= \mathbb{E}^{\tilde{\sigma}_{t:T}^f \tilde{\sigma}_{t:T}^l, \mu_t[a_{1:t-1}]} \left\{ R_t^f(X_t, A_t) + V_{t+1}^f(\mu_{t+1}[a_{1:t-1} A_t], X_{t+1}) \big| a_{1:t-1}, x_{1:t} \right\} \tag{47c}$$

$$= \mathbb{E}^{\tilde{\sigma}_T^f \tilde{\sigma}_T^l, \mu_t[a_{1:t-1}]} \left\{ R_t^f(X_t, A_t) + V_{t+1}^f(\mu_{t+1}[a_{1:t-1} A_t], X_{t+1}) \big| a_{1:t-1}, x_{1:t} \right\} \tag{47d}$$

$$= V_t^f(\mu_t[a_{1:t-1}], x_t), \tag{47e}$$

where (47b) follows from Lemma 4 in Appendix C, (47c) follows from the induction hypothesis in (46), (47d) follows because the random variables involved in expectation, X_t^l, A_t, X_{t+1} do not depend on $\tilde{\sigma}_{t+1:T}^f \tilde{\sigma}_{t+1:T}^l$ and (47e) follows from the definition of $\tilde{\sigma}_t$ in the forward recursion in (28), the definition of μ_{t+1} in (29) and the definition of V_t^f in (24).

D Part 2: Leader

In the following we will show that

$$\mathbb{E}^{\tilde{\sigma}^l, \tilde{\sigma}^f, \mu_t} \left\{ \sum_{n=t}^{T} \delta^{n-t} R_n^l(X_n, A_n) | h_t^c \right\} \geq \mathbb{E}^{\sigma^l, BR^f(\sigma^l), \mu_t} \left\{ \sum_{n=t}^{T} \delta^{n-t} R_n^l(X_n, A_n) | h_t^c \right\}, \tag{48}$$

where $\tilde{\sigma}^f \in BR^f(\tilde{\sigma}^l)$, as shown in Part 1.

Proof. We prove the above result using induction and from results in Lemma 6, 7 and 8 proved in Appendix E.

For base case at $t = T$, $\forall (a_{1:T-1}) \in \mathcal{H}_T^l, \sigma^l$

$$\begin{aligned}
&\mathbb{E}^{\tilde{\sigma}_T^f \tilde{\sigma}_T^l,\, \mu_T[a_{1:T-1}]} \left\{ R_T^f(X_T, A_T) \big| a_{1:T-1} \right\} \\
&= V_T^f(\mu_t[a_{1:T-1}]) && \text{(49a)} \\
&\geq \mathbb{E}^{BR^f(\sigma^l)\sigma_T^l,\, \mu_T[a_{1:T-1}]} \left\{ R_T^f(X_T, A_T) \big| a_{1:T-1} \right\}. && \text{(49b)}
\end{aligned}$$

where (49a) follows from Lemma 8 and (49b) follows from Lemma 6 in Appendix E. Let the induction hypothesis be that for $t+1$, $\forall (a_{1:t}) \in \mathcal{H}_{t+1}^l, \sigma^l$,

$$\begin{aligned}
&\mathbb{E}^{\tilde{\sigma}_{t+1:T}^f \tilde{\sigma}_{t+1:T}^l,\, \mu_{t+1}[a_{1:t}]} \left\{ \sum_{n=t+1}^{T} R_n^f(X_n, A_n) \big| a_{1:t} \right\} \\
&\geq \mathbb{E}^{BR^f(\sigma^l)\sigma_{t+1:T}^l,\, \mu_{t+1}[a_{1:t}]} \left\{ \sum_{n=t+1}^{T} R_n^f(X_n, A_n) \big| a_{1:t} \right\} && \text{(50a)}
\end{aligned}$$

Then $\forall (a_{1:t-1}) \in \mathcal{H}_T^l, \sigma^l$, we have

$$\begin{aligned}
&\mathbb{E}^{\tilde{\sigma}_{t:T}^f \tilde{\sigma}_{t:T}^l,\, \mu_t[a_{1:t-1}]} \left\{ \sum_{n=t}^{T} R_n^l(X_n, A_n) \big| a_{1:t-1} \right\} \\
&= V_t^l(\mu_t[a_{1:t-1}]) && \text{(51a)} \\
&\geq \mathbb{E}^{BR^f(\sigma^l)\sigma_t^l,\, \mu_t[a_{1:t-1}]} \left\{ R_t^f(X_t, A_t) + V_{t+1}^f(\mu_{t+1}[a_{1:t-1} A_t]) \big| a_{1:t-1} \right\} && \text{(51b)} \\
&= \mathbb{E}^{BR^f(\sigma^l)\sigma_t^l,\, \mu_t[a_{1:t-1}]} \Big\{ R_t^f(X_t, A_t) + \\
&\qquad \mathbb{E}^{\tilde{\sigma}_{t+1:T}^f \tilde{\sigma}_{t+1:T}^l,\, \mu_{t+1}[a_{1:t-1}, A_t]} \left\{ \sum_{n=t+1}^{T} R_n^l(X_n, A_n) \, \big| a_{1:t-1}, A_t \big\} \big| a_{1:t-1} \right\} && \text{(51c)} \\
&\geq \mathbb{E}^{BR^f(\sigma^l)\sigma_t^l,\, \mu_t[a_{1:t-1}]} \Big\{ R_t^f(X_t, A_t) + \\
&\qquad \mathbb{E}^{BR^f(\sigma^l)\sigma_{t+1:T}^l \mu_{t+1}[a_{1:t-1}, A_t]} \left\{ \sum_{n=t+1}^{T} R_n^l(X_n, A_n) \, \big| a_{1:t-1}, A_t \big\} \big| a_{1:t-1} \right\} && \text{(51d)} \\
&= \mathbb{E}^{BR^f(\sigma^l)\sigma_t^l,\, \mu_t[a_{1:t-1}]} \Big\{ R_t^f(X_t, A_t) + \\
&\qquad \mathbb{E}^{BR^f(\sigma^l)\sigma_{t:T}^l \mu_t[a_{1:t-1}]} \left\{ \sum_{n=t+1}^{T} R_n^l(X_n, A_n) \, \big| a_{1:t-1}, A_t \big\} \big| a_{1:t-1} \right\} && \text{(51e)} \\
&= \mathbb{E}^{BR^f(\sigma^l)\sigma_{t:T}^l\, \mu_t[a_{1:t-1}]} \left\{ \sum_{n=t}^{T} R_n^l(X_n, A_n) \big| a_{1:t-1} \right\}, && \text{(51f)}
\end{aligned}$$

where (51a) follows from Lemma 8, (51b) follows from Lemma 6, (51c) follows from Lemma 8, (51d) follows from induction hypothesis in (50a) and (51e) follows from Lemma 7.

E

Lemma 6. $\forall t \in [T], (a_{1:t-1}) \in \mathcal{H}_t^l, \sigma_t^l$

$$V_t^f(\mu_t[a_{1:t-1}], x_t) \geq \mathbb{E}^{BR^f(\sigma_t^l)\sigma_t^l\, \mu_t[a_{1:t-1}]} \Big\{ R_t^f(X_t, A_t) + V_{t+1}^l(F(\mu_t[a_{1:t-1}], BR^f(\sigma_t^l)(\cdot|a_{1:t-1}, \cdot), A_t)) \Big| a_{1:t-1} \Big\} \tag{52}$$

Proof. We prove this lemma by contradiction. Suppose the claim is not true for t. This implies $\exists \hat{\sigma}_t^l, \hat{a}_{1:t-1}$ such that

$$\mathbb{E}^{BR^f(\hat{\sigma}_t^l)\hat{\sigma}_t^l,\, \mu_t[\hat{a}_{1:t-1}]} \Big\{ R_t^f(X_t, A_t) + V_{t+1}^f(F(\mu_t[\hat{a}_{1:t-1}], BR^f(\hat{\sigma}_t^l)(\cdot|\hat{a}_{1:t-1}, \cdot), A_t)) \Big| \hat{a}_{1:t-1} \Big\}$$
$$> V_t^f(\mu_t[\hat{a}_{1:t-1}]). \tag{53}$$

We will show that this leads to a contradiction.

Construct $\hat{\gamma}_t^l(a_t^l) = \hat{\sigma}_t^l(a_t^l|\hat{a}_{1:t-1})$

Then for $\hat{a}_{1:t-1}$, we have

$$V_t^l(\mu_t[\hat{a}_{1:t-1}]) \tag{54a}$$
$$= \max_{\gamma_t^l} \mathbb{E}^{\gamma_t^l BR^f(\gamma_t^l)\, \mu_t[\hat{a}_{1:t-1}]} \Big\{ R_t^l(X_t, A_t) + V_{t+1}^l(F(\mu_t[\hat{a}_{1:t-1}], BR^f(\gamma_t^l), A_t^f)) \Big| \hat{a}_{1:t-1} \Big\} \tag{54b}$$
$$\geq \mathbb{E}^{\hat{\gamma}_t^l(\cdot|\hat{x}_t) BR^f(\hat{\gamma}_t^l),\, \mu_t[\hat{a}_{1:t-1}]} \big\{ R_t^l(X_t, A_t) + V_{t+1}^l(F(\mu_t[\hat{a}_{1:t-1}], BR^f(\hat{\gamma}_t^l), A_t)) \big| \hat{a}_{1:t-1} \big\} \tag{54c}$$
$$= \mathbb{E}^{\hat{\sigma}_t^l BR^f(\hat{\sigma}_t^l)\mu_t[\hat{a}_{1:t-1}]} \big\{ R_t^l(X_t, A_t) + V_{t+1}^l(F(\mu_t[\hat{a}_{1:t-1}], BR^f(\hat{\sigma}_t^l)(\cdot|\hat{a}_{1:t-1}, \cdot), A_t)) \big| \hat{a}_{1:t-1} \big\} \tag{54d}$$
$$> V_t^l(\mu_t[\hat{a}_{1:t-1}]) \tag{54e}$$

where (54b) follows from definition of V_t^l in (24), (54d) follows from definition of $\hat{\gamma}_t^l$ and (54e) follows from (53). However this leads to a contradiction.

Lemma 7. $\forall t \in [T], (a_{1:t}) \in \mathcal{H}_{t+1}^l$ *and* σ_t^f

$$\mathbb{E}^{\sigma_{t:T}^l \tilde{\sigma}_{t:T}^l,\, \mu_t[a_{1:t-1}]} \left\{ \sum_{n=t+1}^{T} R_t^l(X_n, A_n) \Big| a_{1:t} \right\} =$$
$$\mathbb{E}^{\sigma_{t+1:T}^l \tilde{\sigma}_{t+1:T}^l,\, \mu_{t+1}[a_{1:t}]} \left\{ \sum_{n=t+1}^{T} R_n^l(X_n, A_n) \Big| a_{1:t} \right\}. \tag{55}$$

Proof. Since the above expectations involve random variables $A_{t+1:T}, X_{t+2:T}$, we consider, $P^{\sigma^f_{t:T}\tilde{\sigma}^l_{t:T},\, \mu_t[a_{1:t-1}]}(a_{t+1:T}, x_{t+2:T}\big|a_{1:t})$.

$$P^{\sigma^f_{t:T}\tilde{\sigma}^l_{t:T},\, \mu_t[a_{1:t-1}]}(a_{t+1:T}, x_{t+2:T}\big|a_{1:t}) \tag{56a}$$
$$= \sum_{x_{t+1}} \mu_{t+1}[a_{1:t}](x_{t+1})\tilde{\sigma}^f_{t+1}(a^f_{t+1}|a_{1:t}, x_{t+1})\tilde{\sigma}^l_{t+1}(a^l_{t+1}|a_{1:t})Q(x_{t+2}|x_{t+1}, a_{t+1})$$
$$P^{\sigma^f_{t:T}\tilde{\sigma}^l_{t:T},\, \mu_t[a_{1:t-1}]}(a_{t+2:T}, x_{t+3:T}\big|a_{1:t+1}, x_{t+2}) \tag{56b}$$

By definition, $\tilde{\sigma}^f_t, \tilde{\sigma}^l_t$ depend on $a_{1:t}$ through the common equilibrium belief $\mu_{t+1}[a_{1:t}]$, as defined in (28). Moreover, the probability on $(a_{t+2:T}, x_{t+3:T})$ given $a_{1:t+1}, x_{t:t+2}$ depend on $a_{1:t+1}, x_{t:t+2}, \mu_{t+1}[a_{1:t}]$ through $\sigma^f_{t+2:T}\tilde{\sigma}^l_{t+2:T}$.

Thus,

$$P^{\sigma^f_{t:T}\tilde{\sigma}^l_{t:T},\, \mu_t[a_{1:t-1}]}(a_{t+1:T}, x_{t+2:T}\big|a_{1:t}) \tag{56c}$$
$$= P^{\sigma^f_{t+1:T}\tilde{\sigma}^l_{t+1:T},\, \mu_{t+1}[a_{1:t}]}(a_{t+1:T}, x_{t+2:T}\big|a_{1:t}) \tag{56d}$$

Lemma 8. $\forall t \in [T], (a_{1:t-1}) \in \mathcal{H}^c_T$

$$V^l_t(\mu_t[a_{1:t-1}]) = \mathbb{E}^{\tilde{\sigma}^f_{t:T}\tilde{\sigma}^l_{t:T}, \mu_t[a_{1:t-1}]}\left\{\sum_{n=t}^{T} R^l_n(X_n, A_n)\big|a_{1:t-1}\right\}. \tag{57}$$

Proof. We prove the lemma by induction. For $t = T$,

$$\mathbb{E}^{\tilde{\sigma}^f_T\tilde{\sigma}^l_T,\, \mu_T[a_{1:T-1}]}\left\{R^l_T(X_T, A_T)\big|a_{1:T-1}\right\}$$
$$= \sum_{x_T, a_T} \mu_t[a_{1:t-1}](x_T)R^l_T(x_T, a_T)\tilde{\sigma}^f_T(a^l_T|a_{1:T-1}, x_T)\tilde{\sigma}^l_T(a^l_T|a_{1:T-1}) \tag{58a}$$
$$= V^l_T(\mu_T[a_{1:T-1}]), \tag{58b}$$

where (58b) follows from the definition of V^l_t in (24) and the definition of $\tilde{\sigma}_T$ in the forward recursion in (28).

Suppose the claim is true for $t+1$, i.e., $\forall t \in [T], (a_{1:t}) \in \mathcal{H}^c_{t+1}$

$$V^l_{t+1}(\mu_{t+1}[a_{1:t}]) = \mathbb{E}^{\tilde{\sigma}^f_{t+1:T}\tilde{\sigma}^l_{t+1:T},\, \mu_{t+1}[a_{1:t}]}\left\{\sum_{n=t+1}^{T} R^l_n(X_n, A_n)\big|a_{1:t}\right\}. \tag{59}$$

Then $\forall t \in [T], (a_{1:t-1}) \in \mathcal{H}^c_t$, we have

$$\mathbb{E}^{\tilde{\sigma}^f_{t:T}\tilde{\sigma}^l_{t:T},\, \mu_t[a_{1:t-1}]}\left\{\sum_{n=t}^{T} R^l_n(X_n, A_n)\big|a_{1:t-1}\right\}$$
$$= \mathbb{E}^{\tilde{\sigma}^f_{t:T}\tilde{\sigma}^l_{t:T},\, \mu_t[a_{1:t-1}]}\left\{R^l_t(X_t, A_t) + \right.$$
$$\left.\mathbb{E}^{\tilde{\sigma}^f_{t:T}\tilde{\sigma}^l_{t:T},\, \mu_t[a_{1:t-1}]}\left\{\sum_{n=t+1}^{T} R^l_n(X_n, A_n)\ \big|a_{1:t-1}, A_t\right\}\big|a_{1:t-1}\right\} \tag{60a}$$

$$= \mathbb{E}^{\tilde{\sigma}^f_{t:T}\tilde{\sigma}^l_{t:T},\,\mu_t[a_{1:t-1}]}\left\{R^l_t(X_t,A_t)+\right.$$
$$\left.\mathbb{E}^{\tilde{\sigma}^f_{t+1:T}\tilde{\sigma}^l_{t+1:T},\,\mu_{t+1}[a_{1:t-1},A_t]}\left\{\sum_{n=t+1}^{T} R^l_n(X_n,A_n)\,\big|a_{1:t-1},A_t\right\}\big|a_{1:t-1}\right\} \quad (60b)$$
$$= \mathbb{E}^{\tilde{\sigma}^f_{t:T}\tilde{\sigma}^l_{t:T},\,\mu_t[a_{1:t-1}]}\left\{R^l_t(X_t,A_t)+V^l_{t+1}(\mu_{t+1}[a_{1:t-1}A_t])\big|a_{1:t-1}\right\} \quad (60c)$$
$$= \mathbb{E}^{\tilde{\sigma}^f_T\tilde{\sigma}^l_T,\,\mu_t[a_{1:t-1}]}\left\{R^l_t(X_t,A_t)+V^l_{t+1}(\mu_{t+1}[a_{1:t-1}A_t])\big|a_{1:t-1}\right\} \quad (60d)$$
$$= V^l_t(\mu_t[a_{1:t-1}]), \quad (60e)$$

where (60b) follows from Lemma 7 in Appendix E, (60c) follows from the induction hypothesis in (59), (60d) follows because the random variables involved in expectation, X^l_t, A_t, X_{t+1} do not depend on $\tilde{\sigma}^f_{t+1:T}\tilde{\sigma}^l_{t+1:T}$ and (60e) follows from the definition of $\tilde{\sigma}_t$ in the forward recursion in (28), the definition of μ_{t+1} in (29) and the definition of V^l_t in (24).

F

Lemma 9. *For all π_t,γ^l_t, $BR^f_t(\gamma^l_t)$ as defined in* (21) *is non-empty.*

Proof.

$$BR^f_t(\gamma^l_t) = \left\{\hat{\gamma}^f_t : \forall x_t \in \mathcal{X}, \hat{\gamma}^f_t(\cdot|x_t) \in \arg\max_{\gamma^f_t(\cdot|x_t)} \right.$$
$$\left.\mathbb{E}^{\gamma^f_t(\cdot|x_t)\gamma^l_t,\,\pi_t}\left\{R^f_t(x_t,A_t)+\delta V^f_{t+1}(F(\pi_t,\hat{\gamma}^f_t,A_t),X_{t+1})\big|x_t\right\}\right\} \quad (61)$$

(i) Let $\mathcal{G}$ be the space of γ, where $\gamma : \mathcal{X} \to \mathcal{P}(\mathcal{A})$. Clearly $\mathcal{B}$ is convex since it is a space of probability measures, and compact which is implied by the finiteness of $\mathcal{X}, \mathcal{A}$.

(ii) Fix γ^l_t. For all $\hat{\gamma}^f_t$, define a correspondence $B : \mathcal{G} \rightrightarrows \mathcal{G}$ as follows

$$B(\bar{\gamma}^f_t) := \left\{\hat{\gamma}^f_t : \forall x_t \in \mathcal{X} : \hat{\gamma}^f_t(\cdot|x_t) \in \arg\max_{\gamma^f_t(\cdot|x_t)} \right.$$
$$\left.\sum_{a_t,x_{t+1}} \left[R^f_t(x_t,A_t)+\delta V^f_{t+1}(F(\pi_t,\bar{\gamma}^f_t,A_t),X_{t+1})\big|x_t\right]\gamma^f_t(a^f_t|x_t)\gamma^l_t(a^l_t)Q(x_{t+1}|x_t,a_t)\right\} \quad (62)$$

Then $B(\bar{\gamma}^f_t)$ is non-empty, closed and convex since the optimization in (62) is a linear program in variables $\gamma^f_t(\cdot|x_t)$, which lie in a compact space that is the probability simplex.

(iii) Since the optimization in (62) is linear and thus continuous in $\gamma^f_t(\cdot|x_t)$, therefore B has closed graph property (from Berge's Maximum theorem [9]). Thus using Kakutani's fixed-point theorem [10][Lemma 20.1], there exists a fixed-point of $B(\cdot)$ which also belongs to $BR^f(\gamma^l_t)$. Thus $BR^f(\gamma^l_t)$ is non-empty.

References

1. Balcan, M.F., Blum, A., Haghtalab, N., Procaccia, A.D.: Commitment without regrets: Online learning in stackelberg security games. In: Proceedings of the sixteenth ACM conference on economics and computation. pp. 61–78. ACM (2015)
2. Basilico, N., Gatti, N., Amigoni, F.: Leader-follower strategies for robotic patrolling in environments with arbitrary topologies. In: Proceedings of The 8th International Conference on Autonomous Agents and Multiagent Systems-Volume 1. pp. 57–64. International Foundation for Autonomous Agents and Multiagent Systems (2009)
3. Conitzer, V., Sandholm, T.: Computing the optimal strategy to commit to. In: Proceedings of the 7th ACM conference on Electronic commerce. pp. 82–90. ACM (2006)
4. Fudenberg, D., Tirole, J.: Perfect bayesian equilibrium and sequential equilibrium. journal of Economic Theory **53**(2), 236–260 (1991)
5. Kar, D., Fang, F., Delle Fave, F., Sintov, N., Tambe, M.: A game of thrones: when human behavior models compete in repeated stackelberg security games. In: Proceedings of the 2015 International Conference on Autonomous Agents and Multiagent Systems. pp. 1381–1390. International Foundation for Autonomous Agents and Multiagent Systems (2015)
6. Kiekintveld, C., Jain, M., Tsai, J., Pita, J., Ordóñez, F., Tambe, M.: Computing optimal randomized resource allocations for massive security games. In: Proceedings of The 8th International Conference on Autonomous Agents and Multiagent Systems-Volume 1. pp. 689–696. International Foundation for Autonomous Agents and Multiagent Systems (2009)
7. Marecki, J., Tesauro, G., Segal, R.: Playing repeated stackelberg games with unknown opponents. In: Proceedings of the 11th International Conference on Autonomous Agents and Multiagent Systems-Volume 2. pp. 821–828. International Foundation for Autonomous Agents and Multiagent Systems (2012)
8. Nayyar, A., Mahajan, A., Teneketzis, D.: Decentralized stochastic control with partial history sharing: A common information approach. Automatic Control, IEEE Transactions on **58**(7), 1644–1658 (2013)
9. Ok, E.A.: Real analysis with economic applications, vol. 10. Princeton University Press (2007)
10. Osborne, M.J., Rubinstein, A.: A Course in Game Theory, MIT Press Books, vol. 1. The MIT Press (1994)
11. Pita, J., Jain, M., Marecki, J., Ordóñez, F., Portway, C., Tambe, M., Western, C., Paruchuri, P., Kraus, S.: Deployed armor protection: the application of a game theoretic model for security at the los angeles international airport. In: Proceedings of the 7th international joint conference on Autonomous agents and multiagent systems: industrial track. pp. 125–132. International Foundation for Autonomous Agents and Multiagent Systems (2008)
12. Tambe, M.: Security and game theory: algorithms, deployed systems, lessons learned. Cambridge university press (2011)
13. Vasal, D., Anastasopoulos, A.: Decentralized Bayesian learning in dynamic games. In: Allerton Conference on Communication, Control, and Computing (2016), `https://arxiv.org/abs/1607.06847`
14. Vasal, D., Anastasopoulos, A.: Signaling equilibria of dynamic LQG games with asymmetric information. In: Conference on Decision and Control (2016)
15. Vasal, D., Sinha, A., Anastasopoulos, A.: A systematic process for evaluating structured perfect bayesian equilibria in dynamic games with asymmetric information. IEEE Transactions on Automatic Control (2018)
16. Yang, R., Ford, B., Tambe, M., Lemieux, A.: Adaptive resource allocation for wildlife protection against illegal poachers. In: Proceedings of the 2014 international conference on Autonomous agents and multi-agent systems. pp. 453–460. International Foundation for Autonomous Agents and Multi agent Systems (2014)