Security Games with Limited Surveillance: The Optimal Stopping Problem and Approximation

1. INTRODUCTION

Stackelberg security games have been used in several deployed applications for allocating limited resources in order to protect critical infrastructure including LAX Airport, US Coast Guard, and the Federal Air Marshals Service [5, 9, 6, 18, 3, 14, 2, 19]. A Stackelberg security game models an interaction between a defender and an attacker [8]. The defender first commits to a security policy (which may be randomized), and the attacker conducts surveillance to learn the defender's policy before launching an attack. A solution to the game yields an optimal randomized strategy for the defender, based on the assumption that the attacker will observe this strategy and respond optimally. Software decision aids based on Stackelberg games have been successfully implemented in several real-world domains, including Los Angeles International Airport (LAX) [14], United States Federal Air Marshals Service (FAMS) [19], United States Transportation Security Agency (TSA) [15], and the United States Coast Guard [2, 16].

Terrorists conduct surveillance to select potential targets and gain strong situational awareness of targets' vulnerabilities and security operations [17]. Most existing work on security games, including deployed applications, assumes that the attacker is able to observe the defender's strategy perfectly. This assumption is a useful first-level approximation, but it is clearly simplistic. In reality, the attacker may have more limited observation capabilities since surveillance is costly and delays an attack. Attackers may also wish to reduce the number of observations due to the risk of being detected by security forces during surveillance activities [17]. Therefore, *it is important to consider situations where attackers select targets based on a limited number of observations using explicit belief updates*.

While there has been some related work that relaxes the perfect observation assumption in security games, the proposed approaches have some fundamental drawbacks. Korzhyk et al. [10] only consider two extreme situations: perfect observation and no observation. Realistically, an attacker may have partial knowledge of the defender's strategy. RECON [23] takes into account possible observation errors by assuming that the attacker's observation is within some distance from the defender's real strategy, but does not address how these errors arise or explicitly model the process of forming beliefs based on limited observations. The COBRA algorithm [13] focuses on human perception of probability distributions by applying support theory [20] from psychology. Both RECON and COBRA require hand-tuned parameters to model observations errors, which we avoid in this paper. Yin et. al [24] prove the equivalence of Stackelberg equilibria and Nash equilibria for some classes of security games. In general, however, Stackelberg and Nash equilibria may differ in security games, and the optimal strategy in cases with limited surveillance may be different for both. There also has been some work on understanding the value of commitment for the leader in general Stackelberg games where observations are limited or costly [4, 11, 21]. An et. al [1] propose a model wherein an attacker forms or updates a belief based on observed limited number of actions, and chooses an optimal response. But the model assumes that the defender can perfectly estimates the number of observations the attacker will make, which is unrealistic in practice.

In this paper, XXXX

2. STACKELBERG SECURITY GAMES

A Stackelberg security game [8] has two players, a defender who uses m identical resources to protect a set of targets $T = \{1, 2, \ldots, n\}$ $(m < n)$, and an attacker who selects a single target to attack. The defender has N pure strategies $\mathcal{A}$, each a coverage vector representing which m targets are covered. Our model can handle more general security settings in which there may exist scheduling constraints on the assignment of resources [7]. In that case, $\mathcal{A}$ represents feasible assignments. We write $A_i = 1$ if target i is covered in strategy $A \in \mathcal{A}$, and $A_i = 0$ otherwise. The defender can choose a randomized strategy $\mathbf{x}$, with $x_A \geq 0$ the probability of playing a strategy A. A defender strategy can be represented more compactly using a marginal coverage vector $\mathbf{c}(\mathbf{x}) = \langle c_i(\mathbf{x}) \rangle$ where $c_i(\mathbf{x}) = \sum_{A \in \mathcal{A}} x_A A_i$ is the probability that target i is covered by some defender resource [8]. The attacker's strategy is a vector $\mathbf{a} = \langle a_i \rangle$ where a_i is the probability of attacking target i. Since the attacker always has an optimal pure-strategy response, we restrict the attacker's strategies to pure strategies without loss of generality.

The payoffs for each player depend on which target is attacked and the probability that the target is covered. If the attacker attacks target i, there are two cases: If target i is covered, the defender receives a reward R_i^d and the attacker receives a penalty P_i^a. Otherwise, the payoffs for the defender and attacker are P_i^d and R_i^a, respectively. We assume that $R_i^d \geq P_i^d$ and $R_i^a \geq P_i^a$ in order to model that the defender would always prefer the attack to fail, while the attacker would prefer it to succeed. For a strategy profile $\langle \mathbf{c}, \mathbf{a} \rangle$, the expected utilities for both agents are given by:

$$U^d(\mathbf{c}, \mathbf{a}) = \sum_{i \in T} a_i U^d(\mathbf{c}, i), \text{ where } U^d(\mathbf{c}, i) = c_i R_i^d + (1 - c_i) P_i^d$$

$$U^a(\mathbf{c}, \mathbf{a}) = \sum_{i \in T} a_i U^a(\mathbf{c}, i), \text{ where } U^a(\mathbf{c}, i) = c_i P_i^a + (1 - c_i) R_i^a$$

In a Stackelberg game, the defender moves first, choosing $\mathbf{x}$, while the attacker observes $\mathbf{x}$ and plays an optimal response $\mathbf{a}$ to it. The standard solution concept is strong Stackelberg equilibrium (SSE) [22]. In an SSE, the defender chooses an optimal strategy $\mathbf{x}$, accounting for the attacker's best response $\mathbf{a}$, under the assumption that the attacker breaks ties in the defender's favor.

3. OPTIMAL STOPPING SECURITY GAMES

We propose to depart from the standard Stackelberg assumption that the attacker has full knowledge of $\mathbf{x}$ and the assumption that the defender knows how many observations the attacker will make. Instead, the Bayesian decision attacker who starts with a prior distribution over the defender's strategies sequentially decides whether to make more observations or to attack based on his posterior belief about the defender's strategies in order to maximize his expected payoff. We refer to our model as OPTS (OPtimal sTopping Security games). The sequence of moves in an OPTS is as follows.

1. First, the defender chooses a strategy. We assume that when choosing a strategy, the defender has knowledge of the attacker's prior beliefs about the defender's strategy and the number of observations the attacker will make.

2. Then, the attacker decides whether to make an observation or start to attack. After making one observation, the attacker updates his belief and decides to make another observation based on his posterior belief about the defender's strategy. The game ends when the attacker attacks his best target based on his posterior belief.

EXAMPLE 1. *We use the LAX airport as an example, based on the ARMOR application [14]. The police at LAX place m checkpoints on the entrance roads to LAX following a mixed strategy computed by ARMOR. The fact that attackers may engage in surveillance prior to an attack is based on real-world cases and feedback from security experts [17], and follows other Stackelberg models deployed in practice and justified elsewhere [12].*[1] *In practice, the attackers will make only a limited number of observations of how the checkpoints are placed before they launch an attack and the observation length may be based on reasoning over the benefit of making more observations and the surveillance cost. For instance, they might observe placements for several days, and then decide to launch an attack if they feel confident about the security strategy of the defender (e.g., he observes the same pattern all the time). Based on the past observations, an attacker may also decide to make more observations if he is not confident about his belief about the defending strategy. A single observation in this domain might involve the attacker driving around the different entrances to the airport in order to determine which ones are covered by checkpoints at any particular time, so each observation gives information about the full strategy of the defender.*[2]

[1]The model in this paper assumes a surveillance phase prior to any actual execution of an attack. In particular, we assume that executing an attack is sufficiently complex that it is prohibitively difficult to observe the pure strategy of the defender and immediately launch an attack against this pure strategy. This assumption is based on real-world cases and feedback from security experts [17], and follows other Stackelberg models deployed in practice and justified elsewhere [12]. One important factor in this is the difficulty of generating and executing complex conditional plans with limited resources.

[2]An alternative model could be developed where the attacker picks one (or a few) targets to observe, and will therefore only learn about a part of the full pure strategy in each observation. This is an interesting direction for future work.

We assume that the attacker and the defender have common knowledge of the attacker's prior beliefs over the set of mixed strategies that the defender may execute. We also assume that the defender does not know the exact times when the attacker will conduct surveillance, and therefore cannot modify the mixed strategy during the observation sequence. This is realistic if the defender is operating in a steady state, and does not know when or where surveillance operations could take place for planning a specific attack.

In an OPTS, the attacker dynamically decides whether to stop surveillance after each observation based on his most updated belief. Assume that the attacker makes $\tau \geq 0$ observations $\sigma = \{\sigma^1, \ldots, \sigma^\tau\}$, each individual observation σ^i is one of the defender's pure strategies. The individual observations are drawn independently from the distribution representing the defender's mixed strategy. Such a sequence of observations σ can be compactly represented by an observation vector $\mathbf{o} = \langle o_A \rangle$ in which o_A is the number of times pure strategy A is observed. An observation vector $\mathbf{o}$ can represent $\frac{\tau!}{\prod_{A\in\mathcal{A}} o_A!}$ observation sequences. The observation vector space can be represented as $\mathcal{O} = \cup_{\tau\in\mathbb{Z}_{\geq 0}} \mathcal{O}_\tau$ where $\mathcal{O}_\tau = \{\mathbf{o} : o_A \in \{0, \ldots, \tau\}, \sum_{A\in\mathcal{A}} o_A = \tau\}$ is the observation vector space when the attacker makes exactly τ observations. (Without loss of generality, we treat the $\mathbf{o} = \langle o_A = 0 \rangle$ as the attacker's observation when the attacker decides not to conduct surveillance.)

Next we discuss how the attacker updates his belief given an observation vector. We assume that the attacker's belief is represented as Dirichlet distributions with support set $\mathcal{S} = \{\mathbf{x} : \sum_{A\in\mathcal{A}} x_A = 1, x_A \geq 0, \forall A \in \mathcal{A}\}$. A Dirichlet distribution $f(\mathbf{x})$ is characterized by a parameter vector $\alpha = \langle \alpha_A \rangle$ with $\alpha_A > -1$ for all $A \in \mathcal{A}$. It assigns probability $\beta \prod_{A\in\mathcal{A}} (x_A)^{\alpha_A}$ to the defender's mixed strategy $\mathbf{x}$, where $\beta = \frac{\Gamma(\sum_{A\in\mathcal{A}} \alpha_A + |\mathcal{A}|)}{\prod_{A\in\mathcal{A}} \Gamma(\alpha_A+1)}$ is a normalization constant expressed in terms of the gamma function Γ. The prior belief can be represented as follows:

$$f(\mathbf{x}) = \frac{\Gamma(\sum_{A\in\mathcal{A}} \alpha_A + |\mathcal{A}|)}{\prod_{A\in\mathcal{A}} \Gamma(\alpha_A + 1)} \prod_{A\in\mathcal{A}} (x_A)^{\alpha_A}$$

If the defender's mixed strategy is $\mathbf{x}$, the probability that the attacker will observe $\mathbf{o} \in \mathcal{O}_\tau$ is $f(\mathbf{o}|\mathbf{x}) = \frac{\tau!}{\prod_{A\in\mathcal{A}} o_A!} \prod_{A\in\mathcal{A}} (x_A)^{o_A}$. By applying Bayes' rule for observation $\mathbf{o}$, we can calculate the posterior distribution as:

$$f(\mathbf{x}|\mathbf{o}) = \frac{\Gamma(\sum_{A\in\mathcal{A}} \alpha_A + |\mathcal{A}| + \tau)}{\prod_{A\in\mathcal{A}} \Gamma(\alpha_A + o_A + 1)} \prod_{A\in\mathcal{A}} (x_A)^{\alpha_A + o_A}$$

Having observed $\mathbf{o}$, the attacker believes that the probability with which the defender chooses pure strategy A is

$$p(A|\mathbf{o}) = \int_{\mathcal{S}} x_A f(\mathbf{x}|\mathbf{o}) \mathrm{d}\mathbf{x} = \frac{\alpha_A + o_A + 1}{\sum_{A'\in\mathcal{A}} \alpha_{A'} + |\mathcal{A}| + \tau}.$$

The marginal coverage of target i according to the posterior belief $f(\mathbf{x}|\mathbf{o})$ is

$$c_i^{\mathbf{o}} = \sum_{A\in\mathcal{A}} A_i p(A|\mathbf{o}) = \frac{\sum_{A\in\mathcal{A}} A_i(\alpha_A + o_A + 1)}{\sum_{A\in A} \alpha_A + |\mathcal{A}| + \tau}.$$

In an OPTS, after making τ observations $\mathbf{o}$, the attacker can either attack his best target based on the posterior belief $f(\mathbf{x}|\mathbf{o})$ or wait to make another observation. If the attacker chooses to attack target i, he will gain an immediate expected utility $(1 - c_i^{\mathbf{o}})R_i^a + c_i^{\mathbf{o}} P_i^a$. If he chooses to make another observation, he has to "pay"

a fixed cost $\lambda > 0$, which represents the opportunity cost by delaying an attack and increasing the probability that the attackers are captured before an attack can be carried out. If the attacker makes τ observations before attacking, the attacker's final expected utility will be reduced by $\tau\lambda$.

4. ATTACKER'S OPTIMAL DECISION

This section discusses the attacker's optimal stopping decision making in an OPTS. Recall that at each time point $t \in \{1, 2, \dots\}$, the attacker can either choose to attack his best target or make another observation and decide when to attack in the future.

4.1 The Optimal Stopping Problem

One immediate observation is that the attacker's optimal stopping decision making is just based on his posterior belief after each observation since the attacker does not know the defender's real strategy $\mathbf{x}$ and he is updating his belief using the Bayes rule before making a decision. Therefore, we can first solve the attacker's optimal stopping problem without knowing the defender's real strategy $\mathbf{x}$, and then compute the defender's optimal strategy given the attacker's optimal stopping policy.

OBSERVATION 2. *The attacker's optimal stopping rule/policy is independent of the defender's real strategy* $\mathbf{x}$.

The attacker's optimal stopping decision making problem can be formulated as a Partially Observable Markov Decision Process (POMDP), which is described by a tuple $\langle S, AC, \Omega, TR, R, O, \mathbf{b}^0 \rangle$, where

- S is the state space which includes the defender's strategy space $\mathcal{S} = \{\mathbf{x} : \sum_{A \in \mathcal{A}} x_A = 1, x_A \geq 0, \forall A \in \mathcal{A}\}$ and an absorbing state ϕ which is reached when the attacker attacks a target.
- AC is the set of actions. At each state $\mathbf{x}$, the attacker's actions include $observe$ and $attack$ a target $i \in T$. At the absorbing state ϕ, the attacker has no action.
- Ω is the set of observations which includes the attacker's pure strategy space $\mathcal{A}$ and ζ, which can only observed at the absorbing state ϕ.
- $TP : S \times AC \times S \mapsto \mathbb{R}$ is the transition function, where $TP(s'|s, a)$ is the transition probability from s to s' if action a is executed. Since the defender's strategy is fixed in our case, the transition probabilities for states are trivial: $TP(\mathbf{x}'|\mathbf{x}, observe) = 1$ if $\mathbf{x}' = \mathbf{x}$ and $TP(\mathbf{x}'|\mathbf{x}, observe) = 0$ otherwise. In addition, $TP(\mathbf{x}'|\mathbf{x}, attack) = 0$ and $TP(\phi|\mathbf{x}, attack) = 1$.
- $R(s, a, s')$ is the reward that the attacker gains by taking a from s and reaching s'. Specifically, $R(\mathbf{x}, observe, \mathbf{x}') = -\lambda$ and $R(\mathbf{x}, attacki, \phi) = (1 - c_i(\mathbf{x}))R_i^a + c_i(\mathbf{x})P_i^a$.
- $O(A|\mathbf{x}, a)$ is the probability of receiving the observation A if the end state is $\mathbf{x}$ after a is taken. If $a = observe$, $O(A|\mathbf{x}, a) = f(A|\mathbf{x})$. If $a = attack$, the attacker observes state ζ.
- b^0 is the initial belief state which is the same as the attacker's prior belief $f(\mathbf{x})$ characterized by a parameter vector α.

The POMDP faced by the attacker is over a continuous multi-dimensional state space and has an infinite horizon, which makes it very difficult to be solved. However, the POMDP has special transition probabilities between states and we are able to show that the continuous state space POMDP is equivalent to an MDP with observation vectors as the states.

THEOREM 3. *The continuous state space POMDP is equivalent to an MDP with observation vectors as the states.*

PROOF. In general, a POMDP is equivalent to a belief-state MDP where each MDP state is a probability distribution (continuous belief state b) over the states of the original POMDP. In the following, we show that the POMDP $\langle S, AC, \Omega, TP, R, O, b^0 \rangle$ defined above is equivalent to an MDP $\langle B, AC, TM, R, b^0 \rangle$ where

- B is the set of belief states over the POMDP states which includes the observation vectors $\mathcal{O}$ and the absorbing state ϕ. An observation vector $\mathbf{o} \in \mathcal{O}$ represents a belief state b such that with the probability $b(\mathbf{x}) = f(\mathbf{x}|\mathbf{o})$, the defender's strategy is $\mathbf{x}$.
- AC is the same set of actions as for the original POMDP.
- $TM : B \times AC \times B \mapsto \mathbb{R}$ is the transition function. $TM(\mathbf{o}'|\mathbf{o}, observe) = p(A|\mathbf{o})$ if $\mathbf{o}' = \mathbf{o} \cup \{A\}$ for an $A \in \mathcal{A}$ and $TM(\mathbf{o}'|\mathbf{o}, observe) = 0$ otherwise. In addition, $TM(\mathbf{o}'|\mathbf{o}, attack) = 0$ and $TM(\phi|\mathbf{o}, attack) = 1$.
- $R(s, a, s')$ is the reward that the attacker gets by taking a from s and reaching s'. Specifically, $R(\mathbf{o}, observe, \mathbf{o}') = -\lambda$ and $R(\mathbf{o}, attacki, \phi) = (1 - c_i^{\mathbf{o}})R_i^a + c_i^{\mathbf{o}}P_i^a$.
- $\mathbf{b}^0$ is the initial state represented by the observation vector $\mathbf{o} = \langle o_A = 0 \rangle$.

We prove the equivalence of the POMDP and the MDP by showing that the belief state of the POMDP is the same as the state of the MDP after any sequence of observations (or actions) by induction.

Clearly, the initial belief state $f(\mathbf{x})$ of the POMDP is the same as the belief represented by the observation vector $\mathbf{o} = \langle o_A = 0 \rangle$. Suppose that the attacker has made τ observations $\sigma = \{\sigma^1, \dots, \sigma^\tau\}$ which can be compactly represented as an observation vector $\mathbf{o}$. We assume that the attacker's current belief state is b which is the same as the belief represented by the observation vector $\mathbf{o}$, i.e., $b(\mathbf{x}) = f(\mathbf{x}|\mathbf{o})$. If the attacker decides to attack, the attacker observes nothing and the attacker reaches the absorbing state ϕ in both models. If the attacker decides to make another observation and (without loss of generality) he observes pure strategy $A \in \mathcal{A}$, the attacker will reach state $\mathbf{o}' = \mathbf{o} \cup A$ and the attacker believes that the defender's strategy is $\mathbf{x}$ with probability $f(\mathbf{x}|\mathbf{o} \cup A)$ according to the MDP model. According to the POMDP model, the attacker believes that the defender's strategy is $\mathbf{x}$ with probability $b'(\mathbf{x})$ based on his action $a = observe$, observation A, and the last belief $b(\mathbf{x}) = f(\mathbf{x}|\mathbf{o})$. It follows that

$$
\begin{aligned}
b'(\mathbf{x}) &= p(\mathbf{x}|a, A, b) = p(A|a, \mathbf{x}, b) \cdot p(\mathbf{x}|a, b)/p(A|a, b) \\
&= p(A|\mathbf{x}) \cdot p(\mathbf{x}|a, b)/p(A|a, b) \\
&= x_A \cdot f(\mathbf{x}|\mathbf{o})/p(A|\mathbf{o}) \\
&= x_A \cdot \frac{\Gamma(\sum_{A' \in \mathcal{A}} \alpha_{A'} + |\mathcal{A}| + \tau)}{\prod_{A' \in \mathcal{A}} \Gamma(\alpha_{A'} + o_{A'} + 1)} \prod_{A' \in \mathcal{A}} (x_{A'})^{\alpha_{A'} + o_{A'}} \\
&\quad /(\frac{\alpha_A + o_A + 1}{\sum_{A' \in \mathcal{A}} \alpha_{A'} + |\mathcal{A}| + \tau}) \\
&= \frac{\Gamma(\sum_{A' \in \mathcal{A}} \alpha_{A'} + |\mathcal{A}| + \tau + 1)}{\prod_{A' \in \mathcal{A} \setminus \{A\}} \Gamma(\alpha_{A'} + o_{A'} + 1)\Gamma(\alpha_A + o_A + 2)} \cdot \\
&\quad (x_A)^{\alpha_A + o_A + 1} \cdot \prod_{A' \in \mathcal{A} \setminus \{A\}} (x_{A'})^{\alpha_{A'} + o_{A'}} \\
&= f(\mathbf{x}|\mathbf{o} \cup A).
\end{aligned}
$$

Furthermore, we can get a finite-state MDP if we could bound the maximum observation length. □

The MDP is in fact a directed acyclic graph (DAG) in consideration of its non-zero transition probabilities: There is an edge from state $\mathbf{o}$ to state $\mathbf{o}'$ if and only if $\mathbf{o}' = \mathbf{o} \cup \{A\}$ where $A \in \mathcal{A}$. Thus, an observation vector with observation length τ is only connected to $|\mathcal{A}|$ observations with observation length $\tau + 1$ if the attacker chooses to make an observation. The initial state represents the situation that the attacker makes no observation. There are infinitely many states since the attacker may make an infinite number of observations which could happen if the observation cost is very small.

If the attacker attacks his best target $\psi(\mathbf{o})$ at state with observation vector $\mathbf{o}$, he will gain an immediate utility[3]

$$W(\mathbf{o}) = U^a(\mathbf{o}) - \lambda \cdot \Delta(\mathbf{o})$$

where $U^a(\mathbf{o}) = c^{\mathbf{o}}_{\psi(\mathbf{o})}(P^a_{\psi(\mathbf{o})} - R^a_{\psi(\mathbf{o})}) + R^a_{\psi(\mathbf{o})}$ is the attacker's utility without considering observation cost, $c^{\mathbf{o}}_{\psi(\mathbf{o})}$ is the marginal coverage of target $\psi(\mathbf{o})$ according to the posterior belief $f(\mathbf{x}|\mathbf{o})$, and $\Delta(\mathbf{o}) = \sum_{A \in \mathcal{A}} o_A$ is the length of observation $\mathbf{o}$.

The attacker can also make another $\tau' > 0$ observations after he observes $\mathbf{o}$. If the attacker's expected utility by making more observations is lower than $W(\mathbf{o})$, he will just attack his best target $\psi(\mathbf{o})$. Formally, we define a value function $V(\mathbf{o})$ for each observation $\mathbf{o}$, which represents the attacker's expected utility when his observation is $\mathbf{o}$ and he follows the optimal policy. At each state, the attacker can either attack the best target $\psi(\mathbf{o})$ and gain a utility $W(\mathbf{o})$ or make one more observation and reach state $\mathbf{o}'$ with probability $p(A|\mathbf{o})$ where $\mathbf{o}' = \mathbf{o} \cup \{A\}$. Therefore, the optimal value function $V(\mathbf{o})$ satisfies the following dynamic programming recursion

$$V(\mathbf{o}) = \max\{W(\mathbf{o}), \sum_{A \in \mathcal{A}} p(A|\mathbf{o}) V(\mathbf{o} \cup \{A\})\}.$$

In other words, with observation $\mathbf{o}$, the attacker can either 1) attack his best target $\psi(\mathbf{o})$ and gain an immediate utility of $W(\mathbf{o})$ or 2) make another observation and gain an expected utility $\sum_{A \in \mathcal{A}} p(A|\mathbf{o})$ $V(\mathbf{o} \cup \{A\})$. Note that whether the attacker will continue to make observations after making the next observation depends on the attacker's next observed defender strategy.

LEMMA 4. *$V(\mathbf{o}) \le V_{\max}(\mathbf{o}) = R^a_{\max} - \lambda \cdot \Delta(\mathbf{o})$ where $R^a_{\max} = \max_{i \in T} R^a_i$ is the attacker's maximum reward.*

PROOF. It is very intuitive but the proof seems not obvious. □

Given the optimal value $V(\mathbf{o})$ for each state $\mathbf{o}$, we can decide the optimal policy (i.e., stopping rule) of the attacker as following: with observation $\mathbf{o}$, the attacker will make at least another observation if and only if $W(\mathbf{o}) < V(\mathbf{o})$. If the attacker decides to attack at one state, all its child states are not *reachable*. We define *an optimal observation graph* with states $\mathcal{O}^*$ which contains all the possible reachable observation vectors the attacker may make. The optimal observation graph can be constructed as follows. Initially, we set $\mathcal{O}^* = \emptyset$. First we add the initial state to $\mathcal{O}^*$. For each state $\mathbf{o} \in \mathcal{O}^*$ such that $W(\mathbf{o}) < \sum_{A \in \mathcal{A}} p(A|\mathbf{o}) V(\mathbf{o} \cup \{A\})$ (i.e., $W(\mathbf{o}) < V(\mathbf{o})$), we add state $\mathbf{o}' = \mathbf{o} \cup \{A\}$ to $\mathcal{O}^*$ for each $A \in \mathcal{A}$. This process continues until no states can be added to $\mathcal{O}^*$.

4.2 Analysis

In this section, we explore some general trends among the strategies and payoffs for the attacker. Assume that the attacker has made a very long sequence of observations. Making another observation cannot change the attacker's posterior belief much. Intuitively, the attacker's immediate expected gain by making another observation is monotone non-increasing. Furthermore, one would expect that if the immediate expected gain is less than the observation cost, the attacker will choose not to attack.

CONJECTURE 5. *For two different observations $\mathbf{o}$ and $\mathbf{o}'$ such that $o_A \le o'_A$ for each $A \in \mathcal{A}$ (i.e., the attacker may observe $\mathbf{o}'$ after observing $\mathbf{o}$). It follows that $\sum_{A \in \mathcal{A}} p(A|\mathbf{o}) U^a(\mathbf{o} \cup \{A\}) - U^a(\mathbf{o}) \ge \sum_{A \in \mathcal{A}} p(A|\mathbf{o}') U^a(\mathbf{o}' \cup \{A\}) - U^a(\mathbf{o}')$.*

In other words, Conjecture 5 says that the maximum immediate benefit of making another observation is monotone non-increasing. Intuitively, when the attacker has made enough observation, its belief will not change much by making another observation and thus, the benefit of making another observation may be negligible. Unfortunately, Conjecture 5 is not true, even in the following apparently much more restrictive class of games: the defender has only one resource $m = 1$, so that a pure strategy A consists precisely of protecting a single target i. The attacker's prior has $\alpha_A = 0$ for all strategies $A \in \mathcal{A}$. Furthermore, the game is zero-sum (so $P^d_i = -R^a_i$), and both players' utility is 0 when the attack fails (so $P^a_i = R^d_i = 0$). In these cases, the game is fully characterized by the target values to the attacker, which we simply write as $R_i := R^a_i$. Consider the zero-sum game with only two targets with values $R_1 = 100$ and $R_2 = 40$. First consider the observation $\mathbf{o} = \langle 0, 0 \rangle$. For the observation $\mathbf{o} = \langle 0, 0 \rangle$, it follows that $\psi(\mathbf{o}) = 1$ and the attacker will gain a utility of 50. If the attacker makes another observation, it will reach $\langle 1, 0 \rangle$ or $\langle 0, 1 \rangle$ with an equal probability. For both observations, the attacker will attack target 1. It thus follows that $\sum_{A \in \mathcal{A}} p(A|\mathbf{o}) U^a(\mathbf{o} \cup \{A\}) - U^a(\mathbf{o}) = 0$. Now consider observation $\mathbf{o}' = \langle 1, 0 \rangle$. If the attacker makes another observation, it will observe $\langle 2, 0 \rangle$ or $\langle 1, 1 \rangle$ and for the former observation, the attacker will attack target 2. We can easily compute that $\sum_{A \in \mathcal{A}} p(A|\mathbf{o}') U^a(\mathbf{o}' \cup \{A\}) - U^a(\mathbf{o}') = 175/24$.

In the above counter example, both observations has a very small length. One may expect that when conjecture is true when both observations have long observation lengths.

CONJECTURE 6. *There exists a τ such that for any two different observations $\mathbf{o}$ and $\mathbf{o}'$ satisfying the conditions $o_A \le o'_A$ for each $A \in \mathcal{A}$ and $\sum_{A \in \mathcal{A}} o_A \ge \tau$, it follows that $\sum_{A \in \mathcal{A}} p(A|\mathbf{o}) U^a(\mathbf{o} \cup \{A\}) - U^a(\mathbf{o}) \ge \sum_{A \in \mathcal{A}} p(A|\mathbf{o}') U^a(\mathbf{o}' \cup \{A\}) - U^a(\mathbf{o}')$.*

PROOF. To be completed. My intuition is that his is not true. Perhaps we can generate a counter example later. □

The above conjectures focus on the relative immediate benefit of making observations. The following observations build bounds on the value of making additional observation.

OBSERVATION 7. *Let $VI(\mathbf{o}) = \sum_{A \in \mathcal{A}} p(A|\mathbf{o}) U^a(\mathbf{o} \cup \{A\}) - U^a(\mathbf{o})$ be the value of information for observation $\mathbf{o}$. For any $\epsilon > 0$, there is a τ such that for all $\tau' > \tau$, $VI(\mathbf{o}') < \epsilon$ (where $\mathbf{o}'$ is an observation vector of length τ').*

PROOF. Without loss of generality, we assume that $\psi(\mathbf{o}) = k$, which implies that $U^a(c^{\mathbf{o}}_k, k) \ge U^a(c^{\mathbf{o}}_j, j)$ for other targets $j \in T \setminus \{k\}$, i.e., $c^{\mathbf{o}}_k(P^a_k - R^a_k) + R^a_k = \frac{\sum_{A \in \mathcal{A}} A_k(\alpha_A + o_A + 1)}{\sum_{A \in \mathcal{A}} \alpha_A + |\mathcal{A}| + \Delta(\mathbf{o})}(P^a_k - R^a_k) + R^a_k \ge \frac{\sum_{A \in \mathcal{A}} A_j(\alpha_A + o_A + 1)}{\sum_{A \in \mathcal{A}} \alpha_A + |\mathcal{A}| + \Delta(\mathbf{o})}(P^a_j - R^a_j) + R^a_j$. Let $\tau = \Delta(\mathbf{o})$. Let $\max_j(R^a_j - P^a_j) = M$. It follows that (assuming $\psi(\mathbf{o} \cup$

[3]Note that the expected utility is from the attacker's perspective and is based on his posterior belief. The *real* attacker utility also depends on the defender's strategy $\mathbf{x}$ which is unknown to the attacker.

$\{A\}) = j)$

$$
\begin{aligned}
VI(\mathbf{o}) &= \sum_{A\in\mathcal{A}} p(A|\mathbf{o})U^a(\mathbf{o}\cup\{A\}) - U^a(\mathbf{o}) \\
&= \sum_{A\in\mathcal{A}} p(A|\mathbf{o})\big(U^a(\mathbf{o}\cup\{A\}) - U^a(\mathbf{o})\big) \\
&\le \sum_{A\in\mathcal{A}} \frac{\alpha_A+o_A+1}{\sum_{A'\in\mathcal{A}}\alpha_{A'}+|\mathcal{A}|+\tau}\Big(\frac{\sum_{A'\in\mathcal{A}}A'_j(\alpha_{A'}+o_{A'}+1)}{\sum_{A'\in\mathcal{A}}\alpha_{A'}+|\mathcal{A}|+\tau+1} \\
&\quad (P_j^a-R_j^a)+R_j^a-U^a(\mathbf{o})\Big) \\
&\le \sum_{A\in\mathcal{A}} \frac{\alpha_A+o_A+1}{\sum_{A'\in\mathcal{A}}\alpha_{A'}+|\mathcal{A}|+\tau}\Big(\frac{\sum_{A'\in\mathcal{A}}A'_j(\alpha_{A'}+o_{A'}+1)}{\sum_{A'\in\mathcal{A}}\alpha_{A'}+|\mathcal{A}|+\tau+1} \\
&\quad (P_j^a-R_j^a)+R_j^a-\Big(\frac{\sum_{A'\in\mathcal{A}}A'_k(\alpha_{A'}+o_{A'}+1)}{\sum_{A'\in\mathcal{A}}\alpha_{A'}+|\mathcal{A}|+\tau}(P_k^a-R_k^a)+R_k^a\Big)\Big) \\
&\le \sum_{A\in\mathcal{A}} \frac{\alpha_A+o_A+1}{\sum_{A'\in\mathcal{A}}\alpha_{A'}+|\mathcal{A}|+\tau}\Big(\frac{\sum_{A'\in\mathcal{A}}A'_j(\alpha_{A'}+o_{A'}+1)}{\sum_{A'\in\mathcal{A}}\alpha_{A'}+|\mathcal{A}|+\tau+1} \\
&\quad (P_j^a-R_j^a)+R_j^a-\Big(\frac{\sum_{A'\in\mathcal{A}}A'_j(\alpha_{A'}+o_{A'}+1)}{\sum_{A'\in\mathcal{A}}\alpha_{A'}+|\mathcal{A}|+\tau}(P_j^a-R_j^a)+R_j^a\Big)\Big) \\
&\le \frac{(\sum_{A'\in\mathcal{A}}A'_j(\alpha_{A'}+o_{A'}+1))(R_j^a-P_j^a)}{(\sum_{A'\in\mathcal{A}}\alpha_{A'}+|\mathcal{A}'|+\tau+1)(\sum_{A'\in\mathcal{A}}\alpha_{A'}+|\mathcal{A}|+\tau)} \\
&\le \frac{M}{\sum_{A\in\mathcal{A}}\alpha_A+|\mathcal{A}|+\tau+1}.
\end{aligned}
$$

Therefore, for any $\epsilon>0$, there exists $\tau=\frac{M}{\epsilon}-\sum_{A\in\mathcal{A}}\alpha_A-|T|-1$ such that for all $\tau'>\tau$, $VI(\mathbf{o}')<\epsilon$ (where $\mathbf{o}'$ is an observation vector of length τ'). □

A stronger (but more useful) observation is as follows.

OBSERVATION 8. *Let $MVI(\mathbf{o})=\max_{A\in\mathcal{A}}U^a(\mathbf{o}\cup\{A\})-U^a(\mathbf{o})$. For any $\epsilon>0$, there is a τ such that for all $\tau'>\tau$, $MVI(\mathbf{o}')<\epsilon$ (where $\mathbf{o}'$ is an observation vector of length τ').*

PROOF. Without loss of generality, we assume that $\psi(\mathbf{o})=k$, which implies that $U^a(c_k^{\mathbf{o}},k)\ge U^a(c_j^{\mathbf{o}},j)$ for other targets $j\in T\setminus\{k\}$, i.e., $c_k^{\mathbf{o}}(P_k^a-R_k^a)+R_k^a=\frac{\sum_{A\in\mathcal{A}}A_k(\alpha_A+o_A+1)}{\sum_{A\in\mathcal{A}}\alpha_A+|\mathcal{A}|+\Delta(\mathbf{o})}(P_k^a-R_k^a)+R_k^a\ge\frac{\sum_{A\in\mathcal{A}}A_j(\alpha_A+o_A+1)}{\sum_{A\in\mathcal{A}}\alpha_A+|\mathcal{A}|+\Delta(\mathbf{o})}(P_j^a-R_j^a)+R_j^a$. Let $\tau=\Delta(\mathbf{o})$.

Let $\max_j(R_j^a-P_j^a)=M$. It follows that (assuming $\psi(\mathbf{o}\cup\{A\})=j$)

$$
\begin{aligned}
MVI(\mathbf{o}) &= \max_{A\in\mathcal{A}} U^a(\mathbf{o}\cup\{A\})-U^a(\mathbf{o}) \\
&= \max_{A\in\mathcal{A}} \big(U^a(\mathbf{o}\cup\{A\})-U^a(\mathbf{o})\big) \\
&\le \max_{A\in\mathcal{A}} \Big(\frac{\sum_{A'\in\mathcal{A}}A'_j(\alpha_{A'}+o_{A'}+1)}{\sum_{A'\in\mathcal{A}}\alpha_{A'}+|\mathcal{A}|+\tau+1}(P_j^a-R_j^a)+R_j^a \\
&\quad -U^a(\mathbf{o})\Big) \\
&\le \max_{A\in\mathcal{A}} \Big(\frac{\sum_{A'\in\mathcal{A}}A'_j(\alpha_{A'}+o_{A'}+1)}{\sum_{A'\in\mathcal{A}}\alpha_{A'}+|\mathcal{A}|+\tau+1}(P_j^a-R_j^a)+R_j^a \\
&\quad -\Big(\frac{\sum_{A'\in\mathcal{A}}A'_k(\alpha_{A'}+o_{A'}+1)}{\sum_{A'\in\mathcal{A}}\alpha_{A'}+|\mathcal{A}|+\tau}(P_k^a-R_k^a)+R_k^a\Big)\Big) \\
&\le \max_{A\in\mathcal{A}} \Big(\frac{\sum_{A'\in\mathcal{A}}A'_j(\alpha_{A'}+o_{A'}+1)}{\sum_{A'\in\mathcal{A}}\alpha_{A'}+|\mathcal{A}|+\tau+1}(P_j^a-R_j^a)+R_j^a \\
&\quad -\Big(\frac{\sum_{A'\in\mathcal{A}}A'_j(\alpha_{A'}+o_{A'}+1)}{\sum_{A'\in\mathcal{A}}\alpha_{A'}+|\mathcal{A}|+\tau}(P_j^a-R_j^a)+R_j^a\Big)\Big) \\
&\le \max_{A\in\mathcal{A}} \Big(\frac{(\sum_{A'\in\mathcal{A}}A'_j(\alpha_{A'}+o_{A'}+1))(R_j^a-P_j^a)}{(\sum_{A'\in\mathcal{A}}\alpha_{A'}+|\mathcal{A}'|+\tau+1)(\sum_{A'\in\mathcal{A}}\alpha_{A'}+|\mathcal{A}|+\tau)} \\
&\le \frac{M}{\sum_{A\in\mathcal{A}}\alpha_A+|\mathcal{A}|+\tau+1}
\end{aligned}
$$

Therefore, for any $\epsilon>0$, there is a $\tau=\frac{M}{\epsilon}-\sum_{A\in\mathcal{A}}\alpha_A-|T|-1$ such that for all $\tau'>\tau$, $MVI(\mathbf{o}')<\epsilon$ (where $\mathbf{o}'$ is an observation vector of length τ'). □

OBSERVATION 9. *If $\Delta(\mathbf{o})>\frac{M}{\epsilon}-\sum_{A\in\mathcal{A}}\alpha_A-|T|-1$, it follows that $V(\mathbf{o})=W(\mathbf{o})$, i.e., the attacker will not make another observation after observing .*

PROOF. We can prove it by contradiction by assuming that $V(\mathbf{o})>W(\mathbf{o})$. Since $\Delta(\mathbf{o})>\frac{M}{\epsilon}-\sum_{A\in\mathcal{A}}\alpha_A-|T|-1$, it follows that $\max_{A\in\mathcal{A}}U^a(\mathbf{o}\cup\{A\})-U^a(\mathbf{o})<\lambda$, i.e., $W(\mathbf{o}\cup\{A\})<W(\mathbf{o})$ for any $A\in\mathcal{A}$.

By assuming $V(\mathbf{o})>W(\mathbf{o})$, it follows that $V(\mathbf{o}\cup\{A\})>W(\mathbf{o}\cup\{A\})$ for an $A\in\mathcal{A}$, i.e., the attacker will choose to make another observation at at least one child state of observation $\mathbf{o}$. This can be proved by contradiction. If $V(\mathbf{o}\cup\{A\})=W(\mathbf{o}\cup\{A\})$ for each $A\in\mathcal{A}$, it follows that $V(\mathbf{o})=\max\big\{W(\mathbf{o}),\sum_{A\in\mathcal{A}}p(A|\mathbf{o})V(\mathbf{o}\cup\{A\})\big\}=\max\big\{W(\mathbf{o}),\sum_{A\in\mathcal{A}}p(A|\mathbf{o})W(\mathbf{o}\cup\{A\})\big\}\le\max\big\{W(\mathbf{o}),\sum_{A\in\mathcal{A}}p(A|\mathbf{o})W(\mathbf{o})\big\}=W(\mathbf{o})$, which is a contradiction.

Furthermore, we can prove that $V(\mathbf{o}\cup\{A\})\ge V(\mathbf{o})$ for an $A\in\mathcal{A}$ such that $V(\mathbf{o}\cup\{A\})>W(\mathbf{o}\cup\{A\})$. In fact, we just need to prove $V(\mathbf{o}\cup\{A\})\ge V(\mathbf{o})$ for an $A\in\mathcal{A}$ since if $V(\mathbf{o}\cup\{A\})\ge V(\mathbf{o})$, it follows that $V(\mathbf{o}\cup\{A\})>W(\mathbf{o}\cup\{A\})$ since $W(\mathbf{o}\cup\{A\})<W(\mathbf{o})$ and $W(\mathbf{o})<V(\mathbf{o})$ by assumption. Since $V(\mathbf{o})>W(\mathbf{o})$, we have $V(\mathbf{o})=\sum_{A\in\mathcal{A}}p(A|\mathbf{o})V(\mathbf{o}\cup\{A\})$ which indicates that $V(\mathbf{o}\cup\{A\})\ge V(\mathbf{o})$ for an $A\in\mathcal{A}$.

Given that $V(\mathbf{o}\cup\{A\})\ge W(\mathbf{o}\cup\{A\})$, we can find an $A'\in\mathcal{A}$ such that $V(\mathbf{o}\cup\{A\}\cup\{A'\})\ge W(\mathbf{o}\cup\{A\}\cup\{A'\})$ and $V(\mathbf{o}\cup\{A\}\cup\{A'\})\ge V(\mathbf{o}\cup\{A\})\ge V(\mathbf{o})$. By continuing this analysis, we can find an observation $\mathbf{o}'>\mathbf{o}$ such that $V(\mathbf{o}')>V_{\max}(\mathbf{o})=R_{\max}^a-\lambda\cdot\Delta(\mathbf{o})$, which is impossible. □

Observation 9 suggests that while computing the attacker's optimal policy, we could ignore states with observation vectors that have a length longer than $\psi=\frac{R_{\max}^a}{\lambda}-|T|-1$ since the attacker will always choose to attack before entering those states.

4.3 The Algorithm

The optimal stopping problem is solved once we have computed all the value functions for the MDP. One well-known approach for solving an MDP is value iteration. Since the MDP has special structures, the value iteration process is basically a backward induction process. To compute the value for observation $\mathbf{o}$, we first need to compute the value for each of its child state. While there are infinitely many states, fortunately Observation 9 suggests that we can ignore observation vectors that have a length longer than $\psi=\frac{R_{\max}^a}{\lambda}-|T|-1$ since the attacker will always choose to attack before entering those states. Therefore, we can start backward induction from observation vectors with length $\psi=\frac{R_{\max}^a}{\lambda}-|T|-1$.

However, the bound $\psi=\frac{R_{\max}^a}{\lambda}-|T|-1$ may not be tight in practice and we may compute the optimal observation graph even when we start backward induction from observation vectors with length that is far smaller than $\psi=\frac{R_{\max}^a}{\lambda}-|T|-1$. Here we presents an incremental algorithm for solving this optimal stopping problem.

The algorithm is based on the following two key insights. First, the attacker may only reach a small number of observations in his observation space. Intuitively, after making a large number of observations, making another observation cannot change the attacker's belief much, and thus, the attacker's gain by making another observation may be negligible as compared with the surveillance cost. Therefore, we can approximately compute the attacker's stopping rule by ignoring observations with longer observation lengths.

Second, if an observation $\mathbf{o}$ is not reachable even if we increase the value $V(\mathbf{o})$, then the observation $\mathbf{o}$ is not reachable. Similarly, if an observation $\mathbf{o}$ is reachable even if we decrease the value $V(\mathbf{o})$, then the observation $\mathbf{o}$ is reachable.

Given that 1) $V(\mathbf{o}) = \max\{W(\mathbf{o}), \sum_{A\in\mathcal{A}} p(A|\mathbf{o})V(\mathbf{o}\cup\{A\})\}$, 2) $W(\mathbf{o}) = U^a(\mathbf{o}) - \lambda\cdot\Delta(\mathbf{o})$, and 3) $U^a_{\min} \le U^a(\mathbf{o}) \le U^a_{\max}$ where $U^a_{\max} = \max_{i\in T} R^a_i$ and $U^a_{\min} = \min_{i\in T} P^a_i$, it then follows that

$$W(\mathbf{o}) \le V(\mathbf{o}) \le U^a_{\max} - \lambda\cdot\Delta(\mathbf{o})$$

Our algorithm simply only considers observations with length no longer than τ. There are two questions here: how to set the value $V(\mathbf{o})$ for each observation $\mathbf{o}\in\mathcal{O}_\tau$ and how it will affect the solution quality, i.e., the value of each $\mathbf{o}'\in\mathcal{O}_{\tau'}$ such that $\tau' < \tau$. Since we know that $W(\mathbf{o}) \le V(\mathbf{o}) \le U^a_{\max} - \lambda\cdot\Delta(\mathbf{o})$, for the approximation approach we try the following two ways to set the value $V(\mathbf{o})$ for each observation $\mathbf{o}\in\mathcal{O}_\tau$: $V(\mathbf{o}) = W(\mathbf{o})$ or $V(\mathbf{o}) = U^a_{\max} - \lambda\cdot\Delta(\mathbf{o})$.

Let the optimal value of observation $\mathbf{o}$ be $V^\tau_{\min}(\mathbf{o})$ when we set $V(\mathbf{o}) = W(\mathbf{o})$ for each $\mathbf{o}\in\mathcal{O}_\tau$. We can compute the optimal value of observation $\mathbf{o}'\in\mathcal{O}_{<\tau}$ by applying backward induction as follows:

$$V^\tau_{\min}(\mathbf{o}) = \begin{cases} W(\mathbf{o}) & \text{if } \mathbf{o}\in\mathcal{O}_\tau \\ 0 & \text{if } \mathbf{o}\in\mathcal{O}_{>\tau} \\ \max\{W(\mathbf{o}), \sum_{A\in\mathcal{A}} p(A|\mathbf{o})V^\tau_{\min}(\mathbf{o}\cup\{A\})\} & \text{if } \mathbf{o}\in\mathcal{O}_{<\tau} \end{cases}$$

Similarly, we can set $V(\mathbf{o}') = U^a_{\max} - \lambda\cdot\Delta(\mathbf{o}')$ for each $\mathbf{o}'\in\mathcal{O}_\tau$ and let the optimal value of observation $\mathbf{o}$ be $V^\tau_{\max}(\mathbf{o})$ in this case. $V^\tau_{\max}(\mathbf{o})$ can be formulated as:

$$V^\tau_{\max}(\mathbf{o}) = \begin{cases} U^a_{\max} - \lambda\cdot\Delta(\mathbf{o}) & \text{if } \mathbf{o}\in\mathcal{O}_\tau \\ 0 & \text{if } \mathbf{o}\in\mathcal{O}_{>\tau} \\ \max\{W(\mathbf{o}), \sum_{A\in\mathcal{A}} p(A|\mathbf{o})V^\tau_{\max}(\mathbf{o}\cup\{A\})\} & \text{if } \mathbf{o}\in\mathcal{O}_{<\tau} \end{cases}$$

Let the optimal value for each observation $\mathbf{o}$ be $V^*(\mathbf{o})$. It is easy to see (by induction) that

$$V^\tau_{\min}(\mathbf{o}) \le V^*(\mathbf{o}) \le V^\tau_{\max}(\mathbf{o}).$$

PROPOSITION 10. *For any $\tau' > \tau$, it follows that $V^\tau_{\min}(\mathbf{o}) \le V^{\tau'}_{\min}(\mathbf{o})$ and $V^\tau_{\max}(\mathbf{o}) \ge V^{\tau'}_{\max}(\mathbf{o})$ for $\mathbf{o}\in\mathcal{O}_{<\tau}$.*

PROOF. Straightforward. □

In other words, the approximated value of each observation is closer to the optimal value when we start the backward induction process from a longer observation deadline.

Given the values $V^\tau_{\min}(\mathbf{o})$ for all observations $\mathbf{o}\in\mathcal{O}_{<\tau}$, we can form the optimal observation graph $G^\tau_{\min}$ for the attacker. Similarly, we can define the optimal observation graph $G^\tau_{\max}$ based on values $V^\tau_{\max}(\mathbf{o})$ for all observations $\mathbf{o}\in\mathcal{O}_{<\tau}$.

PROPOSITION 11. *It follows that $G^\tau_{\min}$ is a subgraph of $G^\tau_{\max}$. Furthermore, if $G^\tau_{\min} = G^\tau_{\max}$ and $V^\tau_{\min}(\mathbf{o}) = V^\tau_{\max}(\mathbf{o})$ for each state $\mathbf{o}$ on the approximate optimal observation graph $G^\tau_{\max}$, the approximate optimal observation graph $G^\tau_{\max}$ or $G^\tau_{\min}$ is the same as the optimal observation graph G^*.*

PROOF. Straightforward. □

Based on Propositions 10 and 11, we propose a search heuristic (Algorithm 1) to iteratively increase τ to find out the optimal observation graph G^*. The algorithm starts with a small τ and check whether the approximate optimal observation graph is optimal. If not, increase τ until τ reaches a threshold ψ. If ψ is large enough, the algorithm can compute the optimal stopping rule.

Algorithm 1: Compute Optimal Observation Graph

```
τ ← 1;
while τ < ψ do
    if G^τ_min = G^τ_max and V^τ_min(o) = V^τ_max(o) for each state o on
    the approximate optimal observation graph G^τ_min then
        return G^τ_min
    end
    else
        τ ← 2τ;
    end
end
```

5. COMPUTING THE DEFENDER'S OPTIMAL STRATEGY

After solving the attacker's optimal stopping problem, we obtain an optimal observation graph with states $\mathcal{O}^*$. Let the leaves of graph be $\mathcal{O}'$ which represent the set of observations for each of which the attacker will attack its best target.

We now introduce an exact (but nonconvex) mathematical program for computing the defender's optimal strategy $\mathbf{x}$, assuming that $\psi(\mathbf{o})$ is pre-computed for each observation $\mathbf{o}\in\mathcal{O}'$.

P1:

$$\max \quad \sum_{\mathbf{o}\in\mathcal{O}'} \frac{\Gamma(\mathbf{o})!}{\prod_{A\in\mathcal{A}} o_A!} \prod_{A\in\mathcal{A}} (x_A)^{o_A} d^{\mathbf{o}} \tag{1}$$

$$\mathbf{s.t.} \quad x_A \in [0,1] \qquad \forall A\in\mathcal{A} \tag{2}$$

$$\sum_{A\in\mathcal{A}} x_A = 1 \tag{3}$$

$$c_i = \sum_{A\in\mathcal{A}} x_A A_i \qquad \forall i\in T \tag{4}$$

$$d^{\mathbf{o}} = c_{\psi(\mathbf{o})} R^d_{\psi(\mathbf{o})} + (1 - c_{\psi(\mathbf{o})}) P^d_{\psi(\mathbf{o})} \qquad \forall \mathbf{o}\in\mathcal{O}_\tau \tag{5}$$

P1 defines the defender's optimal strategy by considering all possible $\mathbf{o}\in\mathcal{O}'$ and evaluating her expected utility for each observation. Equation (2) is the objective function which maximizes the defender's expected payoff $\sum_{\mathbf{o}\in\mathcal{O}'} f(\mathbf{o}|\mathbf{x})d^{\mathbf{o}}$ where $d^{\mathbf{o}}$ is the defender's expected utility when the attacker's observation is $\mathbf{o}$. Equations (3) and (4) define the legal strategy space for the defender. Equation (5) defines the marginal coverage for each target given the defender's strategy $\mathbf{x}$. Equation (6) defines the defender's expected payoff $d^{\mathbf{o}} = c_{\psi(\mathbf{o})} R^d_{\psi(\mathbf{o})} + (1 - c_{\psi(\mathbf{o})}) P^d_{\psi(\mathbf{o})}$ when the attacker attacks $\psi(\mathbf{o})$ for observation $\mathbf{o}$.

6. EVALUATION

For experiments, we will evaluate the runtime and quality of the algorithms with different parameter value combinations.

We can compare this algorithm with SGLS.

7. NEXT STEPS

Implement the algorithm and check whether it can scale up.

Another potential step is coming up with good approximation algorithms for solving the nonconvex optimization problem.

8. REFERENCES

[1] B. An, D. Kempe, C. Kiekintveld, E. Shieh, S. Singh, M. Tambe, and Y. Vorobeychik. Security games with limited surveillance. In *Proc. of the 26th Conference on Artificial Intelligence*, pages 1241–1248, 2012.

[2] B. An, J. Pita, E. Shieh, M. Tambe, C. Kiekintveld, and J. Marecki. GUARDS and PROTECT: Next generation applications of security games. *SIGECOM*, 10:31–34, March 2011.

[3] B. An, M. Tambe, F. Ordóñez, E. Shieh, and C. Kiekintveld. Refinement of strong Stackelberg equilibria in security games. In *Proc. of the 25th Conference on Artificial Intelligence*, pages 587–593, 2011.

[4] K. Bagwell. Commitment and observability in games. *Games and Economic Behavior*, 8:271–280, 1995.

[5] N. Basilico, N. Gatti, and F. Amigoni. Leader-follower strategies for robotic patrolling in environments with arbitrary topologies. In *Proc. of The 8th International Conference on Autonomous Agents and Multiagent Systems (AAMAS)*, pages 500–503, 2009.

[6] J. P. Dickerson, G. I. Simari, V. S. Subrahmanian, and S. Kraus. A graph-theoretic approach to protect static and moving targets from adversaries. In *Proc. of The 9th International Conference on Autonomous Agents and Multiagent Systems (AAMAS)*, pages 299–306, 2010.

[7] M. Jain, E. Kardes, C. Kiekintveld, F. Ordóñez, and M. Tambe. Security games with arbitrary schedules: A branch and price approach. In *Proc. of The 24th AAAI Conference on Artificial Intelligence*, pages 792–797, 2010.

[8] C. Kiekintveld, M. Jain, J. Tsai, J. Pita, M. Tambe, and F. Ordóñez. Computing optimal randomized resource allocations for massive security games. In *Proc. of The 8th International Conference on Autonomous Agents and Multiagent Systems (AAMAS)*, pages 689–696, 2009.

[9] D. Korzhyk, V. Conitzer, and R. Parr. Complexity of computing optimal Stackelberg strategies in security resource allocation games. In *Proc. of The 24th AAAI Conference on Artificial Intelligence*, pages 805–810, 2010.

[10] D. Korzhyk, V. Conitzer, and R. Parr. Solving Stackelberg games with uncertain observability. In *Proc. of The 10th International Conference on Autonomous Agents and Multiagent Systems (AAMAS)*, pages 1013–1020, 2011.

[11] J. Morgan and F. Vardy. The value of commitment in contests and tournaments when observation is costly. *Games and Economic Behavior*, 60(2):326–338, 2007.

[12] J. Pita, M. Jain, F. Ordóñez, M. Tambe, S. Kraus, and R. Magori-Cohen. Effective solutions for real-world Stackelberg games: When agents must deal with human uncertainties. In *Proc. of The 8th International Conference on Autonomous Agents and Multiagent Systems (AAMAS)*, pages 369–376, 2009.

[13] J. Pita, M. Jain, M. Tambe, F. Ordóñez, and S. Kraus. Robust solutions to Stackelberg games: Addressing bounded rationality and limited observations in human cognition. *Artificial Intelligence*, 174(15):1142–1171, 2010.

[14] J. Pita, M. Jain, C. Western, C. Portway, M. Tambe, F. Ordóñez, S. Kraus, and P. Parachuri. Deployed ARMOR protection: The application of a game-theoretic model for security at the Los Angeles International Airport. In *Proc. of The 7th International Conference on Autonomous Agents and Multiagent Systems (AAMAS)*, pages 125–132, 2008.

[15] J. Pita, M. Tambe, C. Kiekintveld, S. Cullen, and E. Steigerwald. GUARDS - game theoretic security allocation on a national scale. In *Proc. of The 10th International Conference on Autonomous Agents and Multiagent Systems (AAMAS)*, pages 37–44, 2011.

[16] E. Shieh, B. An, R. Yang, M. Tambe, C. Baldwin, J. DiRenzo, B. Maule, and G. Meyer. PROTECT: A deployed game theoretic system to protect the ports of the United States. In *Proc. of The 11th International Conference on Autonomous Agents and Multiagent Systems (AAMAS)*, 2012.

[17] E. Southers. *LAX - terror target: the history, the reason, the countermeasure*, chapter Security and Game Theory: Algorithms, Deployed Systems, Lessons Learned, pages 27–50. Cambridge University Press, 2011.

[18] M. Tambe. *Security and Game Theory: Algorithms, Deployed Systems, Lessons Learned*. Cambridge University Press, 2011.

[19] J. Tsai, S. Rathi, C. Kiekintveld, F. Ordóñez, and M. Tambe. IRIS: a tool for strategic security allocation in transportation networks. In *Proc. of The 8th International Conference on Autonomous Agents and Multiagent Systems (AAMAS)*, pages 37–44, 2009.

[20] A. Tversky and D. J. Koehler. Support thoery: A nonextensional representation of subjective probability. *Psychological Review*, 101:547–567, 1994.

[21] E. van Damme and S. Hurkens. Games with imperfectly observable commitment. *Games and Economic Behavior*, 21(1-2):282–308, 1997.

[22] B. von Stengel and S. Zamir. Leadership with commitment to mixed strategies. Technical Report LSE-CDAM-2004-01, CDAM Research Report, 2004.

[23] Z. Yin, M. Jain, M. Tambe, and F. Ordóñez. Risk-averse strategies for security games with execution and observational uncertainty. In *Proc. of The 25th AAAI Conference on Artificial Intelligence (AAAI)*, pages 758–763, 2011.

[24] Z. Yin, D. Korzhyk, C. Kiekintveld, V. Conitzer, , and M. Tambe. Stackelberg vs. nash in security games: interchangeability, equivalence, and uniqueness. In *Proc. of The 9th International Conference on Autonomous Agents and Multiagent Systems (AAMAS)*, pages 1139–1146, 2010.