Journal of the Operations Research Society of Japan
Vol. 62, No. 4, October 2019, pp. 152–168

ON DYNAMIC PATROLLING SECURITY GAMES

Akifumi Kira *Gunma University*
Naoyuki Kamiyama *Kyushu University*
Hirokazu Anai Hiroaki Iwashita Kotaro Ohori *Fujitsu Laboratories Ltd.*

(Received July 3, 2018; Revised June 12, 2019)

Abstract We consider Stackelberg patrolling security games in which a security guard and an intruder walk around a facility. In these games, at each timepoint, the guard earns a reward (intruder incurs a cost) depending on their locations at that time. The objective of the guard (resp., the intruder) is to patrol (intrude) the facility so that the total sum of rewards is maximized (minimized). We study three cases: In Case 1, the guard chooses a scheduled route first and then the intruder chooses a scheduled route after perfectly observing the guard's choice. In Case 2, the guard randomizes her scheduled routes and then intruder observes its probability distribution and also randomize his scheduled routes. In Case 3, the guard randomizes her scheduled routes as well, but the intruder sequentially observes the location of the guard and reroutes to reach one of his targets. We show that the intruder's best response problem in Cases 1 and 2 and Case 3 can be formulated as a shortest path problem and a Markov decision process, respectively. Moreover, the equilibrium problem in each case reduces to a polynomial-sized mixed integer linear programming, linear programming, and bilinear programming problem, respectively.

Keywords: Game theory, optimization, eye-catching patrol, time-expanded network, Markov decision process

1. Introduction

Over recent years, game-theoretic approaches have received considerable attention in terms of guarding cities and facilities against terror attacks. Usually, as security resources are limited, we cannot cover all possible security checkpoints at every moment. Game theory is thus used for appropriately randomizing when and where the resources are placed. Indeed, there exist many successful results that have been deployed in real-world domains such as Los Angeles International Airport [19], the Federal Air Marshals Service [24], the US Coast Guard [7, 23], and the Los Angeles Sheriff's department [28].

This paper considers rich Stackelberg patrolling games, related to Hozaki et al. [9], in which a security guard (the leader, female) and an intruder (the follower, male) walk around a facility represented as a time-expanded network. In these games, at each timepoint, the guard earns a visibility-based reward (the intruder incurs a cost) that depends on their locations at that time. The objective of the guard is to patrol the facility so that the total sum of rewards is maximized. In contrast, the objective of the intruder is to reach one of his targets so that the cost value is minimized.

We study three cases according to the intruder's capability to observe deployment patterns of the guard. In Case 1, the guard chooses a scheduled patrol route (i.e., a path on the time-expanded network). After perfectly observing or learning the guard's choice, the intruder chooses a scheduled intruding route. In other words, this case assume the most intelligent intruder. In Case 2, to prevent the intruder from learning, the guard randomizes her strategies. That is, the guard chooses a probability distribution of the scheduled patrol routes (this can be regarded as a flow with value 1 on the network). Here, we assume that,

when the intruder makes his decision, he can use information about the probability distribution chosen by the guard, but cannot learn which scheduled patrol route has been realized as a result of randomization. This is the basic setting of the recent Stackelberg security games. In Case 3, we deal with the situation where the intruder can observe the location of the guard sequentially. In this case, at each timepoint, the intruder can reroute to reach one of his targets safely, depending on the location of the guard. Therefore, it seems reasonable that the effect of randomization should be limited.

Significance of our models. Japanese police and security companies also share a basic concept that eye-catching patrol becomes a strong deterrent to crime. Namely, it is thought that terror attacks can be prevented beforehand by actively showing the presence of security guards. Therefore, our modeling aimed at optimizing the visibility-based utility is natural. On the other hand, when watching attackers dressed ordinary people, that is, before their attack is executed, it is not so easy to distinguish them as attackers. It depends on know-how of the guards patrolling. The important thing here is to ensure opportunities for the guards to sufficiently gaze at anyone entering by any route. Our games presented in this paper intend to make patrol routes to meet that demand.

Our contributions. Our first contribution is the modeling of three distinct cases. The players' strategies are clearly understood using the concept of time-expanded networks. Our second contribution is to identify solution methods for the intruder's best response. Given a patrol strategy, these can be used to estimate the potential loss in the worst case scenario. We show that the problems of finding a best response in Cases 1 and 2 can be formulated as a shortest path problem and that the problem of finding a best response in Case 3 can be formulated as a Markov decision process (MDP). Our third contribution is to present mathematical optimization formulations of the equilibrium problems. Finding a Stackelberg equilibrium in each case reduces to a mixed integer linear programming (MILP), linear programming (LP), and bilinear programming (BLP) problem, respectively, where the size of each optimization problem is polynomial in the size of the network.

2. Related Work

As an important class of attacker-defender Stackelberg games including network interdiction [10, 27], security resource allocation in networked physical domains such as urban road network has been extensively studied [11–13, 25]. This is close to our Case 2. However, as opposed to our model, the security resources (security guards) are not always mobile.

To protect a mobile target, Fang et al. [7] represent the defender's randomized scheduled routes compactly as flows on a time-extended network. In their game, the attacker's pure strategy is to select a target and the (discretized) time to attack. This leads to an LP formulation of min-max type for finding an equilibrium. In our Case 2, we introduce corresponding flow variables to represent the intruder's scheduled routes as well as those of the guard. Thus, another nontrivial step is required to derive a compact LP formulation.

There are many studies dealing with the situation where the guard, at each turn, chooses the next connected node to move to, allowing the intruder to observe the locations of the guard sequentially [1–3, 5]. However, in the literature as well as [7], the intruder can move toward one of his targets at any moment, regardless of the topology of the network and incur a cost only when he attacks it; in other words, the intruder does not incur a cost while he is moving to the target.

Hohzaki et al. [9] study a patrol problem in a building in which an intruder incurs a cost at each timepoint even while he is moving toward a target. Their dynamic programming

algorithm for obtaining a best response for the intruder can be applied to our Cases 1 and 2. It corresponds to our shortest path approach using the time-expanded network. However, as their game restricts the guard's strategies to given several scheduled routes, their approach cannot be applied to our equilibrium problems.

Finally, we note that the use of an MDP is not new in itself [6, 17, 18, 21], because this is a fundamental tool for handling dynamic environments. For example, Delle Fave et al [6] use a MDP to generate patrol policies with considering execution uncertainty. However, to the best of our knowledge, the model presented here has never been studied.

3. Problem Formulation

Suppose that the facility to be patrolled over a given time interval $\{0, 1, \ldots, \Theta\}$ is represented as a directed graph $\mathcal{G} = (\mathcal{U}, \mathcal{E})$ with super sources $s^{\mathrm{g}}, s^{\mathrm{i}}$ in $\mathcal{U}$ and super targets $t^{\mathrm{g}}, t^{\mathrm{i}}$ in $\mathcal{U}$, where $\mathcal{U}$ is a set of nodes and $\mathcal{E}$ is a set of arcs. The super source s^{g} for the security guard connects some (or possibly all) other nodes, and the super target t^{g} is connected by some (or possibly all) other nodes. The super source s^{i} for the intruder connects each source node (e.g., entrance to the building), and the super target t^{i} is connected by each target node. See Figure 1.

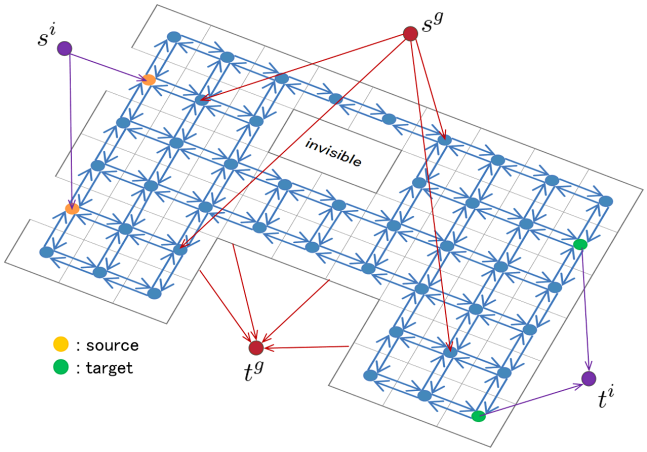

Figure 1: Facility represented as a directed graph $\mathcal{G} = (\mathcal{U}, \mathcal{E})$

A scheduled patrol route of the security guard is a pair of an s^{g}-t^{g} walk of length less than Θ and the timings of her moves on this path. A scheduled route of the intruder is a pair of an s^{i}-t^{i} walk of length less than Θ and the timings of his corresponding movements. Suppose that we are given a reward (or cost) function $r : \mathcal{U} \times \mathcal{U} \to \mathbb{R}_{\geq 0}$. If the guard is on $u \in \mathcal{U}$ and the intruder is on $v \in \mathcal{U}$ at some time, then the guard receives a reward (or the intruder incurs a cost) $r(u, v)$. We assume $r(u, v) = 0$ for all pairs (u, v) such that either u or v is in $\{s^{\mathrm{g}}, s^{\mathrm{i}}, t^{\mathrm{g}}, t^{\mathrm{i}}\}$. The objective of the guard (resp., the intruder) is to maximize (minimize) the total sum of the timepoint-wise rewards (costs). As a special case, we can use the degree of detection [9] defined by

$$r(u, v) := \frac{\delta(u, v)\alpha(v)}{\{d(u, v)\}^2}, \tag{3.1}$$

where $\delta(u, v) \in \{0\,(\text{invisible}), 1\,(\text{visible})\}$ is the visibility of the intruder to the guard, $d(u, v)$ is the distance between them and $\alpha(v) \in [0, 1]$ is the brightness at the intruder's position. This visibility-based utility models the reward that decreases quadratically with the distance

between the guard and intruder. We can also deal with a probability criterion. If we set the cost of the guard as

$$r(u,v) := \log(1-p_{uv}),$$

where $p_{uv} \in [0,1)$ represents the probability of detection depending on their locations, then the total sum of the cost represents the probability of the intruder arriving at his super target without being detected.

To model their strategies more clearly, we use the concept of time-expanded networks [8] (see Figure 2). For the given graph $\mathcal{G} = (\mathcal{U}, \mathcal{E})$, we generate the corresponding time-

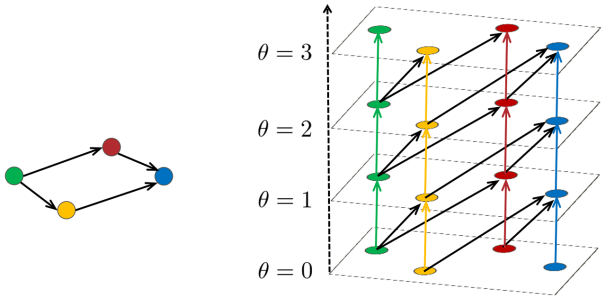

Figure 2: Time-expanded network $\mathcal{N}(\Theta) = (\mathcal{V}, \mathcal{A})$

expanded network $\mathcal{N}(\Theta) = (\mathcal{V}, \mathcal{A})$ with rewards $\hat{r} : \mathcal{V} \times \mathcal{V} \to \mathbb{R}_{\geq 0}$ in the following manner: For each $u \in \mathcal{U}$, we create $\Theta + 1$ copies $u_0, u_1, \ldots, u_\Theta$. Namely,

$$\mathcal{V} = \{u_\theta \,|\, u \in \mathcal{U},\ \theta = 0, 1, \ldots, \Theta\}.$$

For each $e = (u, v) \in \mathcal{E}$, there are Θ copies $e_0, e_1, \ldots, e_{\Theta-1}$, where e_θ connects u_θ to $v_{\theta+1}$. Moreover, for each $u \in \mathcal{U}$ and $\theta \in \{0, 1, \ldots, \Theta - 1\}$ there exists a holdover arc $(u_\theta, u_{\theta+1})$. Namely,

$$\begin{aligned}\mathcal{A} = &\{(u_\theta, v_{\theta+1}) \,|\, (u,v) \in \mathcal{E},\ \theta = 0, 1, \ldots, \Theta - 1\} \\ &\cup \{(v_\theta, v_{\theta+1}) \,|\, v \in \mathcal{U},\ \theta = 0, 1, \ldots, \Theta - 1\}.\end{aligned}$$

For every $u, v \in \mathcal{U}$ and every $\eta, \theta \in \{0, 1, \ldots, \Theta\}$, we set

$$\hat{r}(u_\eta, v_\theta) = \begin{cases} r(u,v) & \text{if } \eta = \theta, \\ 0 & \text{otherwise.} \end{cases} \tag{3.2}$$

For the sake of convenience, we overload the notation $\hat{r}$ to be used for any pair of arcs $(e, a) \in \mathcal{A} \times \mathcal{A}$ and let $\hat{r}(e, a) := \hat{r}(\partial^+ e, \partial^+ a)$, where, $\partial^+ a$ and $\partial^- a$ represent the initial vertex and the terminal vertex of arc a, respectively. We are now in position to describe our security games.

3.1. Case 1

Let us consider a leader-follower game. The leader is the security guard, and chooses $\omega \in \Omega^{\mathrm{g}}$ as her patrol path, where Ω^{g} represents the set of all s_0^{g}-t_Θ^{g} path on $\mathcal{N}(\Theta)$. The follower is the intruder. After perfectly observing or learning the security guard's choice, the intruder chooses $\omega' \in \Omega^{\mathrm{i}}$ as his intruding path, where Ω^{i} represents the set of all s_0^{i}-t_Θ^{i} path on $\mathcal{N}(\Theta)$. Thus, for any profile (ω, ω') in $\Omega^{\mathrm{g}} \times \Omega^{\mathrm{i}}$, the payoff $\psi_1(\omega, \omega')$ is defined as follows:

$$\psi_1(\omega, \omega') = \sum_{e \in \mathcal{A}} \sum_{a \in \mathcal{A}} \hat{r}(e,a) \mathbf{1}_\omega(e) \mathbf{1}_{\omega'}(a),$$

where

$$\mathbf{1}_\omega(a) = \begin{cases} 1 & \text{if } a \text{ is on } \omega \\ 0 & \text{otherwise,} \end{cases} \qquad \omega \in \Omega^{\mathrm{g}} \cup \Omega^{\mathrm{i}}, \quad a \in \mathcal{A}.$$

Hence, the Stackelberg equilibrium is written as follows:

$$\max_{\omega \in \Omega^{\mathrm{g}}} \min_{\omega' \in \Omega^{\mathrm{i}}} \psi_1(\omega, \omega'). \tag{3.3}$$

Figure 3 shows the game tree for this case.

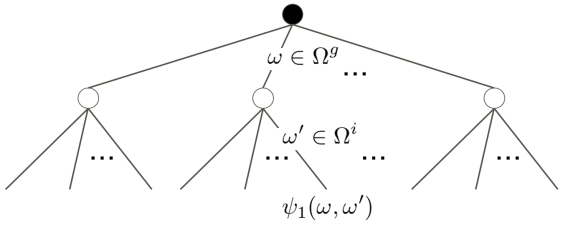

Figure 3: Game tree for Case 1

3.2. Case 2

In this case, the guard chooses a flow $f \in \mathcal{F}^{\mathrm{g}}$ as her randomized patrol path, where $\mathcal{F}^{\mathrm{g}}$ represents the set of all feasible s_0^{g}-t_Θ^{g} flow with value $|f| = 1$ on $\mathcal{N}(\Theta)$. Namely, $\mathcal{F}^{\mathrm{g}}$ is the set of all functions $f : \mathcal{A} \to \mathbb{R}$ satisfying

$$\sum_{a \in \delta^+ v} f(a) - \sum_{a \in \delta^- v} f(a) = \begin{cases} 1 & v = s_0^{\mathrm{g}} \\ 0 & v \in \mathcal{V} \setminus \{s_0^{\mathrm{g}}, t_\Theta^{\mathrm{g}}\} \\ -1 & v = t_\Theta^{\mathrm{g}}, \end{cases} \tag{3.4}$$

$$f(a) \geqq 0, \quad \forall a \in \mathcal{A}. \tag{3.5}$$

where $\delta^+ v$ and $\delta^- v$ denote the set of arcs $a \in \mathcal{A}$ leaving node v and entering node v, respectively. Notice that mixed strategies (or probability vectors) on Ω^g and flow strategies (i.e., $\mathcal{F}^{\mathrm{g}}$) are outcome-equivalent in our game as well as [26]. This outcome equivalence is also related to well known Kuhn's theorem [15] stating that mixed strategies and behavior strategies are outcome-equivalent in games with perfect recall. Similarly, let $\mathcal{F}^{\mathrm{i}}$, the strategy set of the intruder, be the set of all functions $h : \mathcal{A} \to \mathbb{R}$ satisfying

$$\sum_{a \in \delta^+ v} h(a) - \sum_{a \in \delta^- v} h(a) = \begin{cases} 1 & v = s_0^{\mathrm{i}} \\ 0 & v \in \mathcal{V} \setminus \{s_0^{\mathrm{i}}, t_\Theta^{\mathrm{i}}\} \\ -1 & v = t_\Theta^{\mathrm{i}}, \end{cases} \tag{3.6}$$

$$h(a) \geqq 0, \quad \forall a \in \mathcal{A}. \tag{3.7}$$

We note that, when the intruder makes his decision, he can use information about the flow f chosen by the guard, but he cannot learn which patrol path has been realized as a result of randomization. In this game, the objective value turns to the expected value of the total sum of rewards. Thus, for any profile $(f, h) \in \mathcal{F}^{\mathrm{g}} \times \mathcal{F}^{\mathrm{i}}$, the payoff $\psi_2(f, h)$ is given by

$$\psi_2(f, h) = \sum_{e \in \mathcal{A}} \sum_{a \in \mathcal{A}} \hat{r}(e, a) f(e) h(a).$$

Hence, the Stackelberg equilibrium is written as follows:

$$\max_{f\in\mathcal{F}^{\mathrm{g}}}\min_{h\in\mathcal{F}^{\mathrm{i}}}\psi_2(f,h). \tag{3.8}$$

Figure 4 shows the game tree for this case.

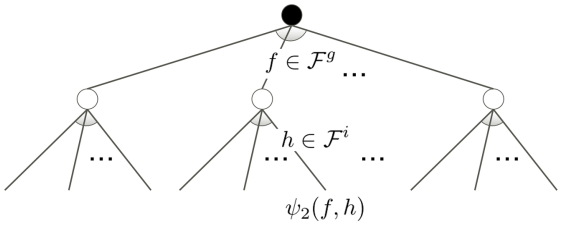

Figure 4: Game tree for Case 2

3.3. Case 3

In this case, at each time θ, the intruder can observe the position of the guard $u_\theta\in\mathcal{V}^{\mathrm{g}}_\theta$ (and his own position $v_\theta\in\mathcal{V}^{\mathrm{i}}_\theta$), where $\mathcal{V}^{\mathrm{g}}_\theta$ (resp., $\mathcal{V}^{\mathrm{i}}_\theta$), a subset of $\mathcal{V}$, denotes the set of copied nodes for time θ such that there exists a directed path to t^{g}_Θ (resp., t^{i}_Θ) in $\mathcal{N}(\Theta)$. We denote $\bigcup_{\theta=0}^{\Theta}\mathcal{V}^{\mathrm{g}}_\theta\times\mathcal{V}^{\mathrm{i}}_\theta$ by $\mathcal{X}$. The probability of the intruder passing through some arc $a\in\delta^+v_\theta$ may depend on u_θ. Therefore, the choice of the intruder is a function $m:\mathcal{A}\times\mathcal{X}\to[0,1]$ such that $m(\,\cdot\,|u,v)$ is a probability distribution on δ^+v for every (u,v) in $\mathcal{X}$. We denote all possible choices of the intruder by $\mathcal{M}$. The game tree for this case appears in Figure 5.

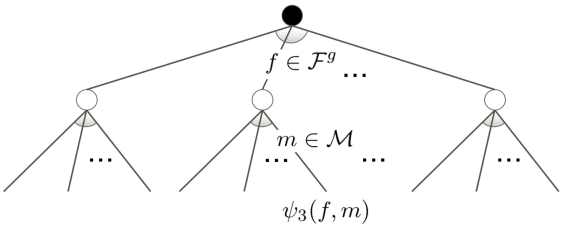

Figure 5: Game tree for Case 3

For any flow f in $\mathcal{F}^{\mathrm{g}}$, we define $\hat{f}:\mathcal{A}\to[0,1]$ in the following manner: For every $a\in\mathcal{A}$, if the initial vertex u of a (i.e., $u=\partial^+a$) satisfies $\sum_{e\in\delta^+u}f(e)>0$, then we set

$$\hat{f}(a)=\frac{f(a)}{\sum_{e\in\delta^+u}f(e)},$$

otherwise, $\hat{f}(a)$ is arbitrary, but for now we define $\hat{f}(a)=1/|\delta^+u|$. We note that $\hat{f}$ satisfies $\sum_{a\in\delta^+v}\hat{f}(a)=1$ for all v in $\mathcal{V}$. In game theory terms, the choices of the guard $\hat{f}\in\{\hat{f}\,|\,f\in\mathcal{F}^{\mathrm{g}}\}$ and the intruder $m\in\mathcal{M}$ correspond to behavior strategies for the extensive-form game with perfect recall, depicted in Figure 6, in which the players sequentially choose an arc to pass through next. Here, the guard has no information on the intruder's past decisions. On the other hand, the intruder has perfect information on the guard's past decisions.

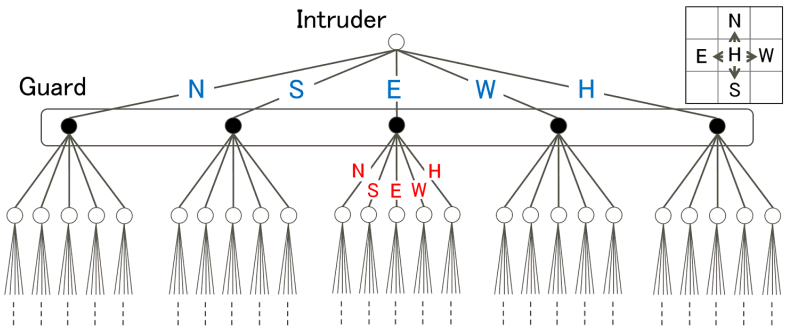

Figure 6: A part of the game tree for Case 3

For any $\omega = (u^0, e^0, u^1, e^1, \ldots, e^{\Theta-1}, u^\Theta)$ in Ω^{g} and any $\omega' = (v^0, a^0, v^1, a^1, \ldots, a^{\Theta-1}, v^\Theta)$ in Ω^{i}, the probability of realizing the pair of the paths (ω, ω') is given by

$$\mathrm{P}(\omega, \omega' \mid f, m) = \prod_{\theta=0}^{\Theta-1} \hat{f}(e^\theta) m(a^\theta \mid u^\theta, v^\theta).$$

Thus, for any $(f, m) \in \mathcal{F}^{\mathrm{g}} \times \mathcal{M}$, the pay-off $\psi_3(f, m)$ is defined as follows:

$$\psi_3(f, m) = \sum_{\omega \in \Omega^{\mathrm{g}}} \sum_{\omega' \in \Omega^{\mathrm{i}}} \psi_1(\omega, \omega') \mathrm{P}(\omega, \omega' \mid f, m).$$

Hence, the Stackelberg equilibrium is written as follows:

$$\max_{f \in \mathcal{F}^{\mathrm{g}}} \min_{m \in \mathcal{M}} \psi_3(f, m). \tag{3.9}$$

4. Strategy Evaluation

In this section, we show that the intruder's best response problems in Cases 1 and 2 and Case 3 can be formulated as a shortest path problem and a MDP, respectively. Given a patrol strategy, these formulations can be used to estimate the potential loss in the worst case scenario.

4.1. Shortest path approach to Cases 1 and 2

Suppose that the security guard's strategy $f \in \mathcal{F}^{\mathrm{g}}$ is fixed (Notice that a path is a special case of a flow). Then the problem of finding the best response (or one of the best responses) for the intruder is the following minimization problem:

$$\min_{h \in \mathcal{F}^{\mathrm{i}}} \psi_2(f, h) = \min_{h \in \mathcal{F}^{\mathrm{i}}} \sum_{a \in \mathcal{A}} \ell_f(a) h(a),$$

where

$$\ell_f(a) := \sum_{e \in \mathcal{A}} \hat{r}(e, a) f(e), \quad a \in \mathcal{A}. \tag{4.1}$$

This is the minimum-cost flow problem on the time-expanded network $\mathcal{N}(\Theta) = (\mathcal{V}, \mathcal{A})$ with costs $\ell_f : \mathcal{A} \to \mathbb{R}_{\geq 0}$, uncapacitated arcs, and supplies/demands $b : \mathcal{V} \to \mathbb{R}$, where

$$b(v) = \begin{cases} 1 & \text{if } v = s_0^{\mathrm{i}}, \\ -1 & \text{if } v = t_\Theta^{\mathrm{i}}, \\ 0 & \text{otherwise.} \end{cases}$$

It is well known that this setting of the minimum-cost flow problem has a solution $h^* \in \mathcal{F}^{\mathrm{i}}$ such that $h^*(a) \in \{0,1\}$ for every $a \in \mathcal{A}$. Therefore, our desired problem reduces to the problem of finding the shortest path from s_0^{i} to t_Θ^{i} on $\mathcal{N}(\Theta)$ with the lengths (costs) $\ell_f : \mathcal{A} \to \mathbb{R}_{\geq 0}$. Hence, we can obtain the intruder's best response using Dijkstra's Algorithm.

4.2. MDP formulation for Case 3

Suppose that a patrol strategy of the guard $f \in \mathcal{F}^{\mathrm{g}}$ is given. We would like to find the intruder's best response (or one of the best responses) to f . This problem reduces to the finite MDP $\mathfrak{D} = \big(\mathcal{X},\, (\mathcal{A}^{\mathrm{i}}, \mathcal{A}^{\mathrm{i}}(\cdot)),\, \hat{r},\, p\big)$ whose components are defined as follows:

1. $\mathcal{X}$ is the **state space** given by $\mathcal{X} = \bigcup_{\theta=0}^{\Theta} \mathcal{V}_\theta^{\mathrm{g}} \times \mathcal{V}_\theta^{\mathrm{i}}$. A state $x = (x^{\mathrm{g}}, x^{\mathrm{i}}) \in \mathcal{X}$ is made up of two components, where x^{g} (resp., x^{i}) $\in \mathcal{V}$ represents the location the guard (the intruder) is currently. Let $\hat{\mathcal{X}}$ be the set of all states except the terminal state, namely, $\hat{\mathcal{X}} = \mathcal{X} \setminus \{(t_\Theta^{\mathrm{g}}, t_\Theta^{\mathrm{i}})\}$.
2. $\mathcal{A}^{\mathrm{i}}$, the set of all arcs $a \in \mathcal{A}$ such that there exists a directed path from $\partial^+ a$ to t_Θ^{i} on $\mathcal{N}(\Theta)$, is the **action space**. For every $x \in \hat{\mathcal{X}}$, $\mathcal{A}^{\mathrm{i}}(x) \subset \mathcal{A}^{\mathrm{i}}$ represents the set of all feasible actions in state x, and it is given by
$$\mathcal{A}^{\mathrm{i}}(x) = \delta^+ x^{\mathrm{i}} \cap \mathcal{A}^{\mathrm{i}}, \quad x = (x^{\mathrm{g}}, x^{\mathrm{i}}) \in \hat{\mathcal{X}}.$$
For convenience, we also let $\mathcal{A}^{\mathrm{g}}$ be the set of all arcs $a \in \mathcal{A}$ such that there exists a directed path from $\partial^+ a$ to t_Θ^{g} on $\mathcal{N}(\Theta)$, and we define $\mathcal{A}^{\mathrm{g}}(x)$ as follows:
$$\mathcal{A}^{\mathrm{g}}(x) = \delta^+ x^{\mathrm{g}} \cap \mathcal{A}^{\mathrm{g}}, \quad x = (x^{\mathrm{g}}, x^{\mathrm{i}}) \in \hat{\mathcal{X}}.$$
3. $\hat{r} : \mathcal{X} \to \mathbb{R}_{\geq 0}$ is the **cost function** defined by (3.2).
4. $p = \{p(\cdot\,|x,a)\}$ is a **Markov transition law**. For every $x = (x^{\mathrm{g}}, x^{\mathrm{i}}) \in \hat{\mathcal{X}}$ and every $a \in \mathcal{A}^{\mathrm{i}}$ such that $a \in \mathcal{A}^{\mathrm{i}}(x)$, and for every $y = (y^{\mathrm{g}}, y^{\mathrm{i}}) \in \mathcal{X}$, we set
$$\begin{aligned} p(y|x,a) &= p(y^{\mathrm{g}}, y^{\mathrm{i}} | x^{\mathrm{g}}, x^{\mathrm{i}}, a) \\ &= \begin{cases} \hat{f}(x^{\mathrm{g}}, y^{\mathrm{g}}) & \text{if } (x^{\mathrm{g}}, y^{\mathrm{g}}) \in \mathcal{A}^{\mathrm{g}},\ y^{\mathrm{i}} = \partial^- a, \\ 0 & \text{otherwise.} \end{cases} \end{aligned}$$
If the process is in state x and action a is chosen, then the process goes to the next state y according to conditional transition probabilities $p(y\,|x,a)$.

Figure 7 illustrates the finite MDP $\mathfrak{D}$. In this figure, x_θ represents the state at time θ ($\theta = 1, 2, \ldots, \Theta$), and a_θ represents the action chosen at time θ ($\theta = 1, 2, \ldots, \Theta - 1$).

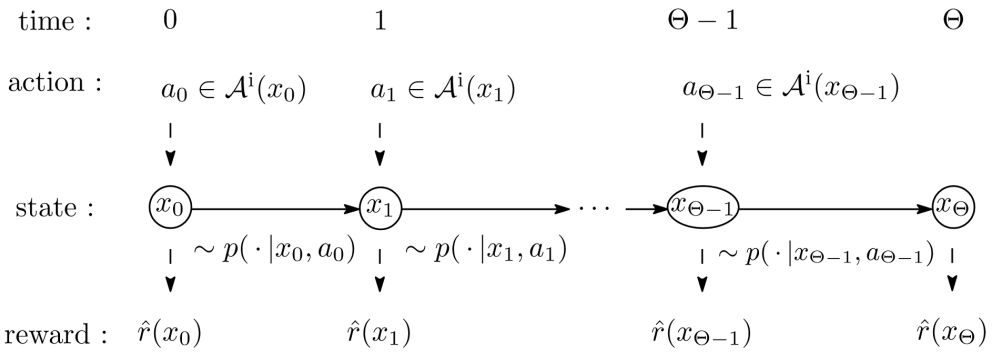

Figure 7: Finite Markov decision process $\mathfrak{D}$

Definition 4.1 (Markov policy)**.** *A mapping $\pi : \hat{\mathcal{X}} \to \mathcal{A}^{\mathrm{i}}$ is called a* (deterministic) *Markov policy if $\pi(x) \in \mathcal{A}^{\mathrm{i}}(x)$ for all $x \in \hat{\mathcal{X}}$. We denote the set of all Markov policies by Π.*

Suppose that a Markov policy π is employed by the intruder. In this case, the MDP commencing from each state x_0 can be regarded as a Markov chain. In other words, if we let X_n be the state after n step transitions from the initial state x_0, then $\{X_n\}$ is the Markov chain satisfying

$$\mathrm{P}^{\pi}(X_{n+1} = y \,|\, X_n = x) = p(y \,|\, x, \pi(x)), \quad x, y \in \mathcal{X},$$

where P^{π} represents the conditional probability given that the policy π is employed. We denote the total expected cost by $v(x;\pi)$. Namely,

$$v(x;\pi) := \mathrm{E}^{\pi}\left[\sum_{\theta=\theta(x)}^{\Theta} \hat{r}(X_\theta) \,\middle|\, X_{\theta(x)} = x \right], \ x \in \mathcal{X}, \ \pi \in \Pi,$$

where E^{π} represents the conditional expectation given that the policy π is employed, and $\theta(x)$ is an integer such that $x \in \mathcal{V}^{\mathrm{g}}_{\theta(x)} \times \mathcal{V}^{\mathrm{i}}_{\theta(x)}$. Our goal is to find the optimal policy (or one of the optimal policies) π^* such that

$$v(x;\pi^*) \leq v(x;\pi), \quad \forall x \in \mathcal{X}, \quad \forall \pi \in \Pi,$$

and the optimal value function $v : \mathcal{X} \to \mathbb{R}$ satisfying

$$v(x) = v(x, \pi^*), \quad x \in \mathcal{X}.$$

The following results directly follow from the basic theory of dynamic programming for finite MDPs (e.g., see [4, 20]). Hence, we can obtain the best response for the intruder and the optimal value $v(s^{\mathrm{g}}_0, t^{\mathrm{i}}_0)$ (i.e., the expected total reward in worst case) by solving the Bellman equation.

Theorem 4.1 (Bellman equation)**.**

$$\begin{aligned} &v(t^{\mathrm{g}}_{\Theta}, t^{\mathrm{i}}_{\Theta}) = 0, \\ &v(x) = \hat{r}(x) + \min_{a \in \mathcal{A}^{\mathrm{i}}(x)} \sum_{e \in \mathcal{A}^{\mathrm{g}}(x)} v(\partial^- e, \partial^- a) \hat{f}(e), \ x \in \hat{\mathcal{X}}. \end{aligned}$$

Theorem 4.2. *Let π^* be the Markov policy which, when the process is in state x, selects the action (or an action) minimizing the summation in the Bellman equation:*

$$\pi^*(x) \in \arg\min_{a \in \mathcal{A}^{\mathrm{i}}(x)} \sum_{e \in \mathcal{A}^{\mathrm{g}}(x)} v(\partial^- e, \partial^- a) \hat{f}(e), \quad x \in \hat{\mathcal{X}}.$$

Then π^* is optimal.

5. Equilibrium Problem

In this section, we show that the equilibrium problem in each case reduces to a MILP, LP, and BLP problem, respectively, such that the size of each optimization problem is polynomial in the size of the time-expanded network.

5.1. Compact LP formulation for Case 2

We note again that the inner minimization problem for the intruder is the well known minimum-cost flow problem. It has the LP problem form:

$$\mathrm{P}(f) \left| \begin{array}{ll} \mathrm{Min} & \displaystyle\sum_{a \in \mathcal{A}} \ell_f(a) h(a) \\ \mathrm{s.t.} & (3.6), (3.7), \end{array} \right.$$

and its dual form:

$$\mathrm{D}(f) \left| \begin{array}{ll} \text{Max} & \rho(t^{\mathrm{i}}_{\Theta}) - \rho(s^{\mathrm{i}}_0) \\ \text{s.t.} & \rho(\partial^- a) - \rho(\partial^+ a) \leqq \ell_f(a), \quad \forall a \in \mathcal{A}. \end{array} \right.$$

Using the dual form, our desired equilibrium problem can be expressed by the following LP problem as follows:

$$\left| \begin{array}{ll} \text{Max} & \text{objective function of } \mathrm{D}(f) \\ \text{s.t.} & \text{constraints of } \mathrm{D}(f) \\ & (3.4), (3.5), (4.1). \end{array} \right.$$

This problem has at most $|\mathcal{V}| + 2|\mathcal{A}|$ variables and $|\mathcal{V}| + 2|\mathcal{A}|$ constraints. Therefore we can obtain a Stackelberg equilibrium using a general-purpose LP solver. Our approach is essentially the same as the compact LP formulation for the polyhedral zero-sum game [14].

5.2. Compact MILP formulation for Case 1

The LP formulation for Case 2 becomes a MILP formulation for Case 1 by replacing the nonnegative constraint $f(a) \geqq 0$ with the binary constraint $f(a) \in \{0, 1\}$ for every $a \in \mathcal{A}$. To see this is true, notice that the difference between Cases 1 and 2 is that the strategies in Case 2 are flows, but those in Case 1 are restricted to paths (i.e., binary integer flows). However, we can equivalently replace the inner minimization problem for the intruder in (3.3) with that in (3.8), because the coefficients of the constraints in the minimum-cost flow problem form a totally unimodular matrix [22]. Furthermore, we can equivalently replace it with the dual problem. Hence we obtain the result.

5.3. BLP formulation for Case 3

We first relax the Bellman equation to inequalities:

$$v(t^{\mathrm{g}}_{\Theta}, t^{\mathrm{i}}_{\Theta}) \leq 0, \tag{5.1}$$

$$v(x) \leq \hat{r}(x) + \sum_{e \in \mathcal{A}^{\mathrm{g}}(x)} v(\partial^- e, \partial^- a) \hat{f}(e), \ a \in \mathcal{A}^{\mathrm{i}}(x), \ x \in \hat{\mathcal{X}}. \tag{5.2}$$

It is known that solving the Bellman equation is equivalent to the following LP problem [16].

$$\mathrm{M}(\hat{f}) \left| \begin{array}{ll} \text{Max} & \displaystyle\sum_{x \in \mathcal{X}} v(x) \\ \text{s.t.} & (5.1), (5.2). \end{array} \right.$$

Using this fact, the equilibrium problem can be expressed by the following BLP problem:

$$\left| \begin{array}{ll} \text{Max} & \text{objective function of } \mathrm{M}(\hat{f}) \\ \text{s.t.} & \text{constraints of } \mathrm{M}(\hat{f}) \\ & \displaystyle\sum_{a \in \delta^+ v \cap \mathcal{A}^{\mathrm{g}}} \hat{f}(a) = 1, \quad \forall v \in \bigcup_{\theta=0}^{\Theta-1} \mathcal{V}^{\mathrm{g}}_{\theta} \\ & \hat{f}(a) \geqq 0, \quad \forall a \in \mathcal{A}^{\mathrm{g}}, \end{array} \right.$$

where the value of the game is equal to the value of $v(s^{\mathrm{g}}_0, s^{\mathrm{i}}_0)$ at every solution in this problem. Notice that, as shown in (3.9), the outer maximization of the equilibrium problem is taken with respect to f (i.e., flows), but is replaced by $\hat{f}$ (i.e., behavior strategies) in the BLP formulation. This validity follows from the outcome-equivalence of f and $\hat{f}$.

6. Computational Experiments

In this section, we present computational results for the evaluation of strategies in Cases 2 and 3 (Case 1 is a special case of Case 2) and the results of computing the Stackelberg equilibrium in Case 2.

6.1. Virtual facility model

To test our approaches in a practical setting, we generate graphs on a square lattice as facilities to be patrolled (see Figure 8). Nodes are arranged at normal cells of an $n \times n$ square grid. The first and last columns are the intruder's sources and targets, respectively, whereas the guard can start from any node at time 0. Obstacles that block the vision of the guard are placed at random, and are depicted as black-colored cells. We set three scenario variables: (i) spatial size, (ii) temporal size, and (iii) obstacle ratio. The spatial size represents the number of cells (nodes). The temporal size, that is Θ, represents the length of the time interval. The obstacle ratio determines the proportion of obstacles in the grid. We use the degree of detection, defined in (3.1), as the reward function. The brightness of all cells that are not obstacles is set to 1.

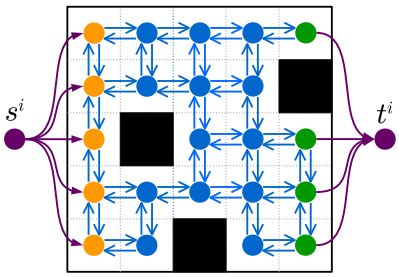

Figure 8: Graph on a square lattice with random obstacles

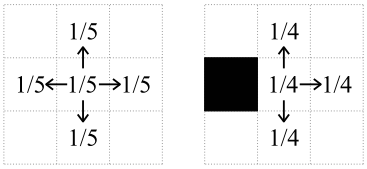

Figure 9: Move probabilities of the robotic patroller

To observe the computational behaviors of the intruder's best response, we consider a simple patrol robot following a random walk (see Figure 9.).

6.2. Strategy evaluation

We implemented the shortest path algorithm for Cases 1 and 2 and the algorithm solving the Bellman equation for Case 3 using C++. We executed them on a Linux Workstation with Intel® Xeon®-E3-1275 processor of 3.60GHz and 32GB memory installed.

First, we fixed the obstacle ratio to zero and the temporal size to 200. Then, we repeated the process of generating a facility and computing the intruder's best responses while varying the spatial size from 10×10 to 100×100 nodes. The results are shown in Figure 10. As expected, the payoff from the intruder's best response in Case 3 is generally much smaller than that in Case 2. This is because, at each timepoint, the intruder can reroute to reach one of his target safely. Although the computational time increases quadratically with the spatial size in both cases, it is found that our approaches are successfully work in a realistic time for problems with spatial sizes of up to 100×100. In the largest case, the MDP with $200 \times (100 \times 100)^2 = 20$ billion states is exactly solved.

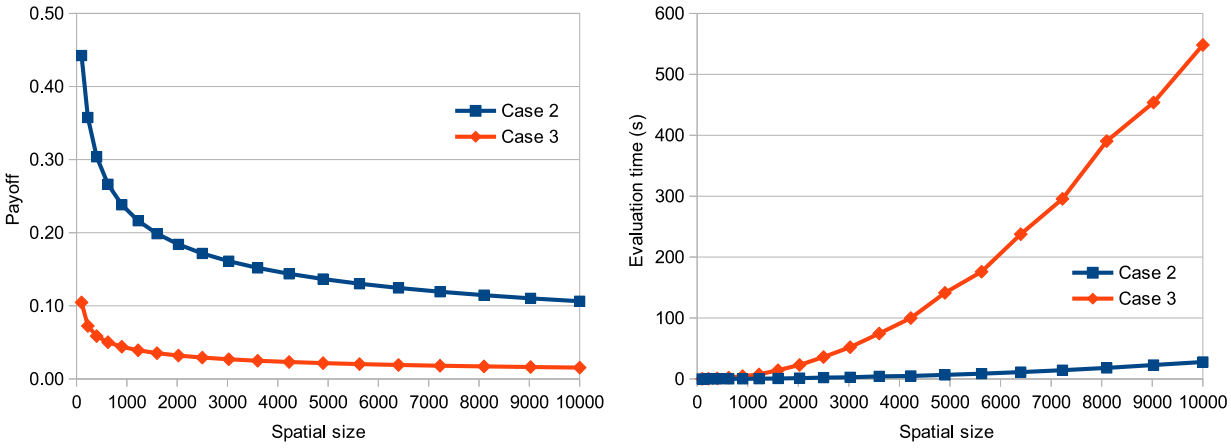

Figure 10: Effect of the spatial size

Second, we fixed the obstacle ratio to zero and the spatial size to 40×40 nodes. Then the computational experiments were carried out for various temporal size (see Figure 11). The results clearly show that the computational time increases linearly with the temporal size in both cases, but Case 3 is more sensitive to the size. We can also observe that the expected payoff for Case 3 becomes the same as the expected payoff for Case 2 when the temporal size matches the shortest distance between the sources and the targets, because the intruder loses his ability to reroute.

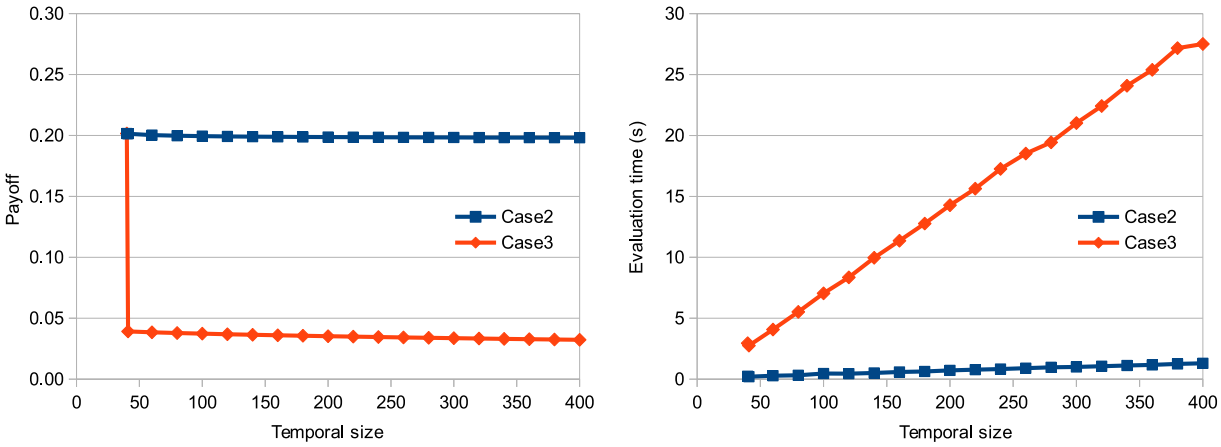

Figure 11: Effect of the temporal size

Finally, we fixed the number of time steps to 200 and the spatial size to 40×40 nodes. Then we carried out the experiments for various obstacle ratios. The results are shown in Figure 12.* We can observe that increasing the obstacle ratio first reduces the payoffs, especially in Case 3, the payoff goes to zero. This is because the intruder can use obstacles successfully to escape the patrol robot's time eyes. However, when the obstacle ratio exceeds a certain value, the payoffs turn to increase. This is because the obstacles reduced the number of available routes of the intruder.

*We computed the visibility $\delta(u, v)$ for all pair (u, v) of nodes in advance by determining whether or not there is an obstacle on the segment uv. This time is not included in the evaluation time, and is much shorter than the evaluation time.

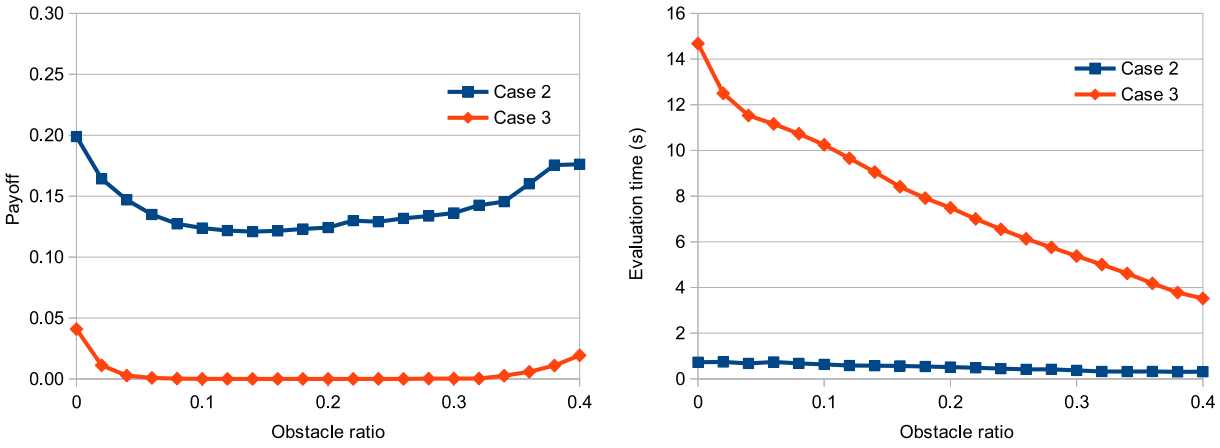

Figure 12: Effect of the obstacle ratio

6.3. Equilibrium strategies in Case 2

We fixed the obstacle ratio to zero and the temporal size to 100. Then, we repeated the process of generating a facility and computing the Stackelberg equilibrium for spatial sizes from 3×3 to 10×10 nodes. In this experiments, we used Gurobi Optimizer ver. 6.5.1 (Gurobi Optimization, Inc.) as the optimization engine to solve the LP problem formulated in Section 5. The results are depicted as line graphs (in green) in Figure 13. For comparison, the results of the strategy evaluation for the robotic patroller, obtained in Section 6.2, are depicted as a line graph (in blue). The computational time required to compute the Stackelberg equilibrium is not insignificant unfortunately. However, this solution enables us to consider how good the given strategy is by comparing with exact optimal strategies.

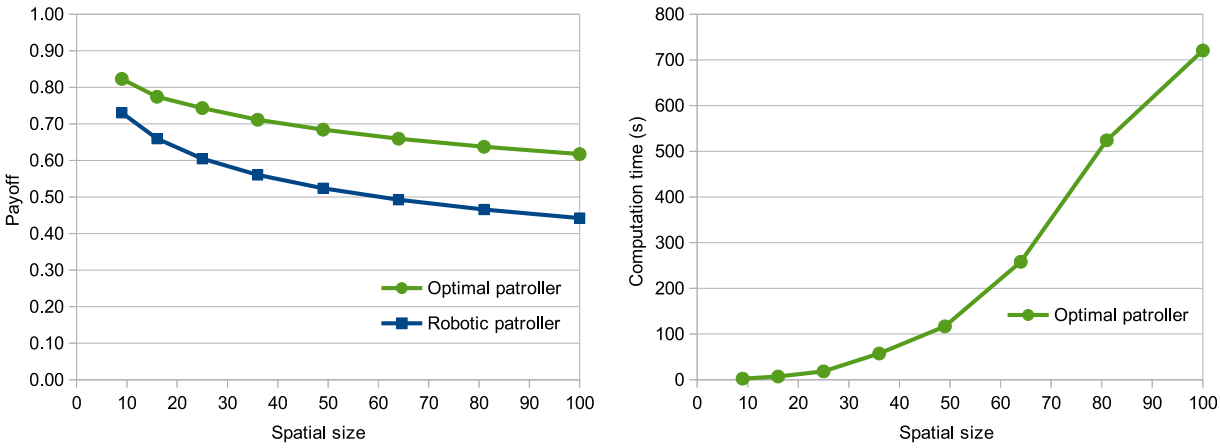

Figure 13: Effect of the spatial size

Second, we fixed the obstacle ratio to zero and the spatial size to 10×10 nodes. The computational experiments were carried out with various temporal sizes. The results are shown in Figure 14.

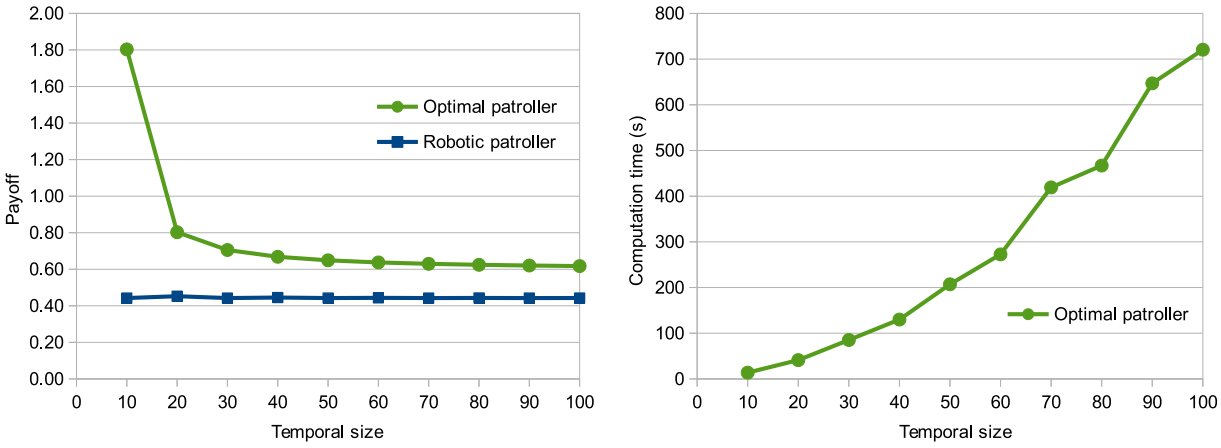

Figure 14: Effect of the temporal size

Finally, we fixed the number of time steps to 50 and the spatial size to 10×10 nodes. The computational experiments were carried out for various obstacle ratios. The results are shown in Figure 15. We can observe that the random walk strategy of the patrol robot becomes progressively worse, compared with the equilibrium strategies, as the obstacle ratio increases. In the literature, it is often said that using the simple random walk strategy is a suboptimal solution, that is, it is not so worse. However, we see that it depends on the structure of the facility to be patrolled.

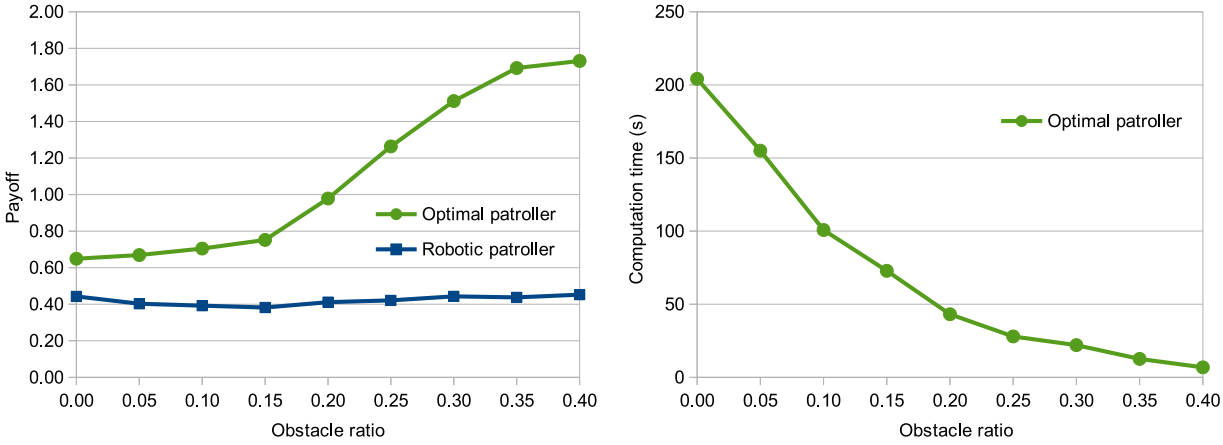

Figure 15: Effect of the obstacles ratio

7. Concluding Remarks

We have formulated rich dynamic patrolling games in which a security guard patrols a facility so as to maximize the total sum of rewards (degree of detection), whereas an intruder attempts to reach a target while minimizing this value. For all three cases, we have proposed the mathematical optimization formulation to compute the intruder's best response. Given a patrol strategy, these formulations can be used to estimate the loss in the worst case scenario. Moreover, for all three cases, we have proposed the mathematical optimization formulation for computing the equilibrium strategy. MILP formulation for Case 1 and BLP formulation for Case 3 have scalability issues. However, these formulations together with LP formulation for Case 2 are useful in clearly understanding the computational difficulties of

the problems. Constructing more effective solution methods is our future work. The authors believe that the findings from this study will be valuable for many real-world applications.

Although we assume that the guard and the intruder have the same maximum moving speed, our models can be naturally expanded to handle cases in which the maximum speeds of them are different by preparing two time-expanded networks (one for the guard and one for the intruder). By adding constraints to the guard's flow strategies (i.e., $\mathcal{F}^{\mathrm{g}}$), our model can handle cases where the security guard must visit some given checkpoints. In the future, we would like to verify the effectiveness of the proposed method while conducting further numerical experiments assuming such various situations.

Acknowledgments

The authors would like to thank the anonymous referees for helpful comments and suggestions on this manuscript. Akifumi Kira was supported in part by JSPS KAKENHI Grant Numbers 26730010 and 17K12644, Japan. Naoyuki Kamiyama was supported by JST, PRESTO Grant Number JPMJPR14E1, Japan.

References

[1] N. Agmon, S. Kraus, and G. A. Kaminka: Multi-robot perimeter patrol in adversarial settings. In *2008 IEEE International Conference on Robotics and Automation (ICRA)* (IEEE, 2008), 2339–2345.

[2] F. Amigoni, N. Gatti, and A. Ippedico: A game-theoretic approach to determining efficient patrolling strategies for mobile robots. In *2008 IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology (WI-IAT)*, **2** (IEEE, 2008), 500–503.

[3] N. Basilico, N. Gatti, and F. Amigoni: Leader-follower strategies for robotic patrolling in environments with arbitrary topologies. In *Proceedings of The 8th International Conference on Autonomous Agents and Multiagent Systems (AAMAS)*, **2** (IFAAMAS, 2009), 57–64.

[4] R. Bellman: *Dynamic Programming* (Princeton University Press, Princeton, NJ, 1957).

[5] B. Bošanskỳ, V. Lisỳ, M. Jakob, and M. Pěchouček: Computing time-dependent policies for patrolling games with mobile targets. In *Proceedings of The 10th International Conference on Autonomous Agents and Multiagent Systems (AAMAS)*, **3** (IFAAMAS, 2011), 989–996.

[6] F.M. Delle Fave, A.X. Jiang, Z. Yin, C. Zhang, M. Tambe, S. Kraus, and J.P. Sullivan: Game-theoretic security patrolling with dynamic execution uncertainty and a case study on a real transit system. *Journal of Artificial Intelligence Research*, **50-2** (2014), 321–367.

[7] F. Fang, A.X. Jiang, and M. Tambe: Optimal patrol strategy for protecting moving targets with multiple mobile resources. In *Proceedings of The 2013 International Conference on Autonomous Agents and Multiagent Systems (AAMAS)* (IFAAMAS, 2013), 957–964.

[8] L.R. Ford Jr and D.R. Fulkerson: *Flows in Networks* (Princeton university press, 2015).

[9] R. Hohzaki, S. Morita, and Y. Terashima: A patrol problem in a building by search theory. In *2013 IEEE Symposium on Computational Intelligence for Security and Defense Applications (CISDA)* (IEEE, 2013), 104–111.

[10] E. Israeli and R.K. Wood: Shortest-path network interdiction. *Networks*, **40-2** (2002), 97–111.

[11] H. Iwashita, K. Ohori, H. Anai, and A. Iwasaki: Simplifying urban network security games with cut-based graph contraction. In *Proceedings of The 2016 International Conference on Autonomous Agents and Multiagent Systems (AAMAS)* (IFAAMAS, 2016), 205–213.

[12] M. Jain, V. Conitzer, and M. Tambe: Security scheduling for real-world networks. In *Proceedings of The 2013 International Conference on Autonomous Agents and Multiagent Systems (AAMAS)* (IFAAMAS, 2013), 215–222.

[13] M. Jain, D. Korzhyk, O. Vaněk, V. Conitzer, M. Pěchouček, and M. Tambe: A double oracle algorithm for zero-sum security games on graphs. In *Proceedings of The 10th International Conference on Autonomous Agents and Multiagent Systems (AAMAS)*, **1** (IFAAMAS, 2011), 327–334.

[14] T. Kavitha, J. Mestre, and M. Nasre: Popular mixed matchings. *Theoretical Computer Science*, **412-24** (2011), 2679–2690.

[15] H.W. Kuhn: Extensive games and the problem of information, In H.W. Kuhn and A.W. Tucker (eds.): *Contributions to the Theory of Games*, **2** (*Annals of Mathematics Studies*, **28**) (Princeton University Press. Princeton, 1953), 193–216.

[16] A.S. Manne: Linear programming and sequential decisions. *Management Science*, **6-3** (1960), 259–267.

[17] D.G. Olson and G.P. Wright: Models for allocating police preventive patrol effort. *Journal of the Operational Research Society*, **26-4** (1975), 703–715.

[18] P. Paruchuri, M. Tambe, F. Ordóñez, and S. Kraus: Security in multiagent systems by policy randomization. In *Proceedings of The Fifth International Joint Conference on Autonomous Agents and Multiagent Systems (AAMAS)*, **3** (ACM, 2006), 273–280.

[19] J. Pita, M. Jain, J. Marecki, F. Ordóñez, C. Portway, M. Tambe, C. Western, P. Paruchuri, and S. Kraus: 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 (AAMAS): Industrial Track* (IFAAMAS, 2008), 125–132.

[20] M.L. Puterman: *Markov Decision Processes: Discrete Stochastic Dynamic Programming* (John Wiley & Sons, 2014).

[21] S. Ruan, C. Meirina, F. Yu, K.R. Pattipati, and R.L. Popp: Patrolling in a stochastic environment. Technical report (Defense Technical Information Center, 2005).

[22] A. Schrijver: *Theory of Linear and Integer Programming* (John Wiley & Sons, 1998).

[23] 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 *Proceedings of The 11th International Conference on Autonomous Agents and Multiagent Systems (AAMAS)*, **1** (IFAAMAS, 2011), 13–20.

[24] J. Tsai, C. Kiekintveld, F. Ordonez, M. Tambe, and S. Rathi: IRIS - a tool for strategic security allocation in transportation networks. In *Proceedings of The 8th International Conference on Autonomous Agents and Multiagent Systems (AAMAS): Industrial Track* (IFAAMAS, 2009), 881–886.

[25] J. Tsai, Z. Yin, J.-y. Kwak, D. Kempe, C. Kiekintveld, and M. Tambe: Urban security: Game-theoretic resource allocation in networked physical domains. In *Proceedings of The Twenty-Fourth AAAI Conference on Artificial Intelligence* (AAAI Press, 2010), 881–886.

[26] A. Washburn and K. Wood: Two-person zero-sum games for network interdiction. *Operations Research*, **43-2** (1995), 243–251.

[27] R.K. Wood: Deterministic network interdiction. *Mathematical and Computer Modelling*, **17-2** (1993), 1–18.

[28] Z. Yin, A.X. Jiang, M.P. Johnson, C. Kiekintveld, K. Leyton-Brown, T. Sandholm, M. Tambe, and J.P. Sullivan: TRUSTS: Scheduling randomized patrols for fare inspection in transit systems. In *Proceedings of The Twenty-Fourth Innovative Applications of Artificial Intelligence conference (IAAI)* (AAAI Press, 2012), 8 pages.

Akifumi Kira
Faculty of Social and Information Studies
Gunma University
4-2 Aramaki-machi, Maebashi
Gunma 371-8510, Japan
E-mail: a-kira@si.gunma-u.ac.jp