Sequential Shortest Path Interdiction with Incomplete Information and Limited Feedback

Jing Yang
Department of Industrial Engineering, University of Pittsburgh, Pittsburgh, PA 15261, USA
jiy75@pitt.edu,

Juan S. Borrero
School of Industrial Engineering & Management, Oklahoma State University, Stillwater, OK 74078, USA
juan.s.borrero@okstate.edu,

Oleg A. Prokopyev
Department of Industrial Engineering, University of Pittsburgh, Pittsburgh, PA 15261, USA
droleg@pitt.edu,

Denis Sauré
Department of Industrial Enginnering, Universidad de Chile, Santiago, Chile
dsaure@dii.uchile.cl

We study sequential shortest path interdiction, where in each period an interdictor with incomplete knowledge of the arc costs blocks at most k arcs, and an evader with complete knowledge about the costs traverses a shortest path between two fixed nodes in the interdicted network. In each period, the interdictor, who aims at maximizing the evader's cumulative cost over a finite time horizon, and whose initial knowledge is limited to valid lower and upper bounds on the costs, observes only the total cost of the path traversed by the evader, but not the path itself. This *limited information feedback* is then used by the interdictor to refine her knowledge of the network's costs, which should lead to better decisions. Different interdiction decisions lead to different responses by the evader, and thus to different feedback. Focusing on minimizing the number of periods it takes a policy to recover a *full information* interdiction decision (that taken by an interdictor with complete knowledge about costs), we show that a class of greedy interdiction policies requires, in the worst case, an exponential number of periods to converge. Nonetheless, we show that, under less stringent modes of feedback, convergence in polynomial time is possible. In particular, we consider different versions of imperfect randomized feedback that allow establishing polynomial expected convergence bounds. Finally, we also discuss a generalization of our approach for the case of a strategic evader, who does not necessarily follow a shortest path in each period.

Key words: network interdiction, shortest path, learning, incomplete information, limited feedback

1. Introduction

Background and Motivation. In the shortest path interdiction problem (SPI), an *interdictor* blocks a subset of arcs on a network with the objective of maximizing the length (total cost) of the path chosen by an *evader*, who in turn selects such a path so as to minimize its length (Smith et al. 2013). Interdiction actions are limited by a budgetary constraint, typically expressed in terms of the (weighted) number of arcs that can be interdicted simultaneously, and the evader is assumed to select paths between two fixed and known nodes in the interdicted network. SPI arises naturally in various application areas (Smith and Song 2019), such as defense of critical infrastructure, infectious disease, and hazardous materials transportation control, as well as counter-terrorism, where two adversarial players compete in zero-sum games (see below for the details on an application to smuggling interdiction).

The traditional single-period, *full-information* version of SPI, where the interdictor has complete knowledge of the network's structure and costs, has been studied extensively in the past. Fulkerson and Harding (1977) model a variant of SPI where arcs can be partially blocked, via linear programming, and Israeli and Wood (2002) develop a mixed-integer programming (MIP) formulation of SPI, and develop decomposition algorithms for its solution. When the interdictor's budgetary constraint is expressed in terms of the number of arcs that can be blocked, SPI is also known as the *k-most-vital-arcs* problem, see, e.g., Malik et al. (1989), Corley and Sha (1982) and Ball et al. (1989), where k denotes the maximum number of arcs blocked. More recently, Morton et al. (2007) study SPI in the context of nuclear material smuggling where the evader's origin-destination pair is random, and the interdictor focuses on maximizing the expected length of the chosen path.

Various extensions of the traditional setting have been studied. For example, Sefair and Smith (2016) consider a setting where the interdictor selects her[1] actions as the evader traverses through a path, and, as a response, the evader can alter such path in an adaptive fashion. Also, Song and Shen (2016) study *risk-averse* SPI, where the network's costs are stochastic and the interdictor focuses on maximizing the probability that the length of the path chosen is above some threshold.

In the work above, the agents' interaction is limited to a single period. In contrast, we focus our attention on settings where the interdictor and the evader interact sequentially over time. This setting is motivated in part by applications in smuggling interdiction, where an interdictor (e.g., a U.S. law-enforcement or military task force) has to periodically reallocate resources (e.g., ships, helicopters, drones) to maximize the probability of detecting and capturing smugglers, minimize the flow of illegal materials, among others; see, e.g., the discussion and references in Gift (2010).

In such settings, the interdictor typically is not initially aware of all possible options that evaders may have at their disposal (e.g., smuggling routes). Moreover, the interdictor becomes aware of said options only when the evader makes use of them, and even then, the information collected on said options might only be partial (e.g., only a portion of a smuggling route might be revealed). Thus, it can be argued that learning plays a key role in practical settings, where the interdictor might (partially) observe the evader's actions (e.g., by interpreting satellite images, or by obtaining data from informants), but might not act upon them immediately (i.e., in the same period).

[1] In the remainder of the paper, we refer to the interdictor and the evader as she and he, respectively.

Sequential settings have been studied in the context of attacker-defender and defender-attacker problems using game-theoretical approaches, see Hausken and Zhuang (2011), Xu and Zhuang (2016), Zhuang et al. (2010) and the references therein. For example, Hausken and Zhuang (2011) consider how the government should balance resource allocation between attempting to downgrade a terrorist's resources and defending against a terrorist attack in a multi-period attacker-defender game. Zhuang et al. (2010) study multi-period attacker-defender games, where the defender can be deceptive, while the attacker has incomplete information but may have learning capabilities.

In terms of the availability of information, the class of problems we consider can be viewed as network interdiction with asymmetric information, as the decision-makers do not "have the same perception of their problem data" (Smith and Song 2019). For example, Bayrak and Bailey (2008) study the bilevel problem where the interdictor's and the evader's arc costs are different. Salmerón (2012) considers a setting, where the interdictor can be deceptive. In contrast, in our setting, arc costs coincide for both decision-makers and all interdiction actions are known to the evader; however, only the evader has full information about the underlying network, while the interdictor's initial information is limited, as also emphasized in our discussion above.

The key feature that separates our study from most of the extant literature in SPI is the interdictor's online learning ability to adapt as new information is collected by observing the evader's actions in multiple time periods. Indeed, the comprehensive survey by Smith and Song (2019) identifies only two other studies that involve some form of learning. The first study, Zheng and Castañón (2012), focuses on information collection (e.g., by sensor placement) when the interdictor and evader interact only once. The second study, Borrero et al. (2016), considers a setting with incomplete knowledge and learning where the evader and the interdictor interact repeatedly over time.

The setting of Borrero et al. (2016) is close to ours in that the agents interact sequentially over time, the evader has complete knowledge of the network, and the interdictor has incomplete information about the network's structure and costs. Borrero et al. (2016) assume that in each period the interdictor observes the full path used by the evader, as well as the costs of all arcs included in the path. There, performance is measured in terms of a policy's *time-stability*, which is defined as the number of periods until the interdictor's actions coincide with those taken by an interdictor with full prior knowledge of the network's structure and costs. In Borrero et al. (2016), the authors propose a class of *greedy* and *pessimistic* policies, where in each period the interdictor (greedily) implements a solution to the k-most vital arcs problem[2] in the observed network, under the worst-case realizations for the evader (pessimistic) of the currently unknown costs. In our work we adopt such a setting (including the performance criterion) with one major distinction: we first

[2] A set of k-most vital arcs in graph G consists of (at most) k arcs whose removal from G results in the greatest increase of the length (total cost) of the shortest path between two specified nodes.

consider the setting of *standard* feedback, where only the length of the chosen path is revealed to the interdictor; then we consider the setting of *imperfect* feedback, where only some arcs in the chosen path might be revealed with certain probability. In addition, we also depart from the setting above by extending our analysis to settings where the evader might act strategically and not necessarily choose a shortest path in every time period.

Incomplete Information and Limited Feedback. Borrero et al. (2019) extend the aforementioned framework to a general class of max-min bilevel linear mixed-integer optimization problems that model the interactions between an upper-level *leader* (interdictor) and a lower-level *follower* (evader). Along the way, it formalizes the notion of *feedback*, i.e., the information revealed to the interdictor by the evader's actions in each period. We use such a notion here.

Specifically, in the context of SPI, Borrero et al. (2019) define feedback as *standard* if in each period the interdictor learns the total cost incurred by the evader; standard feedback is called *response-perfect* if, in addition, the path chosen by the evader is revealed to the interdictor as well, and *value-perfect* if the cost of each arc on said path is also revealed. In this regard, Borrero et al. (2019) generalize the greedy and pessimistic policies of Borrero et al. (2016) (which assume feedback is both response- *and* value-perfect) to be *greedy* and *robust*, under the assumption that standard feedback is either value- *or* response-perfect. In practice, the notion of feedback being either response- or value-perfect is rather strong. For example, in the context of smuggling interdiction, it implies that the interdictor observes the details of the smuggler's route, along with the itemized costs (per arc). In practice, only partial information might be obtained from interrogating smugglers if caught, and limited resources (e.g., satellite images) might reveal the passage of smugglers only on a limited set of passage points.

The goal of this paper is to relax the rather stringent assumptions about feedback in sequential SPI. Specifically, we consider settings with standard feedback, where the interdictor observes only the total cost incurred by the evader in each period but neither the arcs used, nor their costs. In addition, we introduce the notions of *response-imperfect* and *value-imperfect* feedback: under the former notion, the interdictor learns only a subset of the arcs in the path chosen by the evader with some probability; under the latter notion, the interdictor learns the costs of arcs on a further subset of arcs, also with some probability. (See Section 2.2 for formal definitions.)

Contribution. The main contribution made by this paper consists of relaxing the assumption of perfect feedback in the context of sequential SPI. In doing so, we generalize the greedy and pessimistic policies of Borrero et al. (2016). Because the term "pessimistic" has already a known connotation in bilevel optimization terminology (see, e.g., Sinha et al. (2018)) we use the term "greedy and robust", as in Borrero et al. (2019), and propose a family of *greedy, robust* and *non-repetitive* (GRN) policies. As in Borrero et al. (2019), GRN policies are greedy and robust in that they implement a solution to the k-most vital arc problem in the observed network, by assuming the worst-case cost realization for the evader. In addition, the policies make sure not to repeat interdiction solutions implemented in previous periods if their observed costs do not match the interdictor's beliefs. This requirement has the effect of inducing exploration of alternative solutions. Not surprisingly, under standard feedback, GRN policies are guaranteed to converge to the full-information solution. However, we show that these policies have exponential time-stability in the worst case. Considering this, we introduce the notion of *imperfect* feedback, a compromise between perfect and standard feedback, and show that under such a feedback, time-stability for GRN policies admits polynomial expected convergence bounds.

Our second contribution follows from noting that exact computation of GRN policies is hard in general, even in settings where feedback allows for tractable (polyhedral) representation of the interdictor's knowledge. Hence, we provide an approximation to GRN policies, which we show: preserves theoretical convergence guarantees; is exact for a particular type of uncertainty representation; and can be computed by solving an MIP formulation using off-the-shelf solvers.

An additional noteworthy contribution made by the paper is the extension of the analysis to settings where the evader does not necessarily respond by choosing to traverse a shortest path in the interdicted network, and might instead react, for example, strategically. To do so, we generalize the concept of time-stability, so as to account for the time periods in which the evader effectively takes advantage of the interdictor's initial uncertainty to increase her regret (and discards periods in which the evader's actions are *clearly* sub-optimal; see the details in Section 6).

The remainder of the paper is organized as follows. Section 2 outlines the mathematical model for sequential SPI under limited feedback, including our key assumptions, the formal definition of the feedback we consider, and the proposed GRN policies. In Section 3, we analyze convergence of time-stability for GRN policies under standard feedback. Section 4 analyzes GRN polices under imperfect feedback, and provides a polynomial upper bound on the expected time-stability under value-imperfect feedback. Section 5 presents our approximate GRN policies, and the MIP formulation for their computation; we also discuss special cases when our approximations coincide with GRN policies. In Section 6 we present an extension that addresses possible strategic behavior on the evader's behalf. Section 7 presents a set of computational experiments that illustrate the performance of the proposed policies. Finally, Section 8 presents concluding remarks and outlines directions for future research. All proofs and supporting material are relegated to the appendices.

2. Mathematical Model and Interdiction Policies

This section introduces our model for sequential SPI with incomplete information and limited feedback. First, we model the interaction between the interdictor and the evader and describe our key assumptions. Then we define different notions of feedback and introduce the GRN policies. Table 1 below summarizes the main notation used in the paper.

Table 1 Brief Summary of Key Notation

k	Maximum number of interdicted arcs in a period	$z_R^{t,*}$	Cost that the GRN interdictor expects to see in period t
$G(I)$	Subgraph resulting from removing the arcs in I	$\underline{z}_R(I^t)$	Approximate cost that a robust interdictor expects the evader to incur using I^t
$S(I)$	Set of all $1-n$ shortest paths in graph $G(I)$		
$z(I)$	Cost of the shortest $1-n$ path in $G(I)$	$\hat{\underline{z}}_R(I^t)$	Approximate cost that a robust non-repetitive interdictor expects to see using I^t
z^*	Optimal cost of the k-most vital arcs problem on G		
T	Time horizon	$\underline{z}_R^{t,*}$	Cost that the approximate GRN interdictor expects to see in period t
$\mathcal{C}^0$	Initial information about the cost vectors	$\tilde{z}_R(I^t)$	Cost that a robust non-repetitive interdictor expects to see using I^t under general evader
(ℓ_a, u_a)	Lower and upper bounds on cost of arc $a \in A$		
$\mathcal{C}^t$	Cost vectors consistent with information up to period t	$\tilde{z}_R^{t,*}$	Cost that the GRN interdictor expects to see in period t under general evader
P^t	Path chosen by the evader in period t		
P_r^t	Arcs learned in response-imperfect feedback	$R^{t,\pi}$	Regret of policy π until period t
P_v^t	Arcs learned in value-imperfect feedback	τ^π	Time-stability for policy π
I^t	Set of arcs blocked by the interdictor in period t	ξ^π	Earliest time when the cost expected by the interdictor equals the observed cost
$\mathcal{F}^{t,\pi}$	History up to period t under policy π		
$z^{t,\pi}$	Cost incurred by the evader given interdiction decision $I^{t,\pi}$	$\underline{\xi}^\pi$	ξ^π adjusted by the approximate interdictor
$z_R(I^t)$	Cost that a robust interdictor expects the evader to incur using I^t	$\tilde{R}^{t,\pi}$	Generalized regret (for general evader) of policy π until period t
$\hat{z}_R(I^t)$	Cost that a robust non-repetitive interdictor expects to see using I^t	$\tilde{\tau}^\pi$	Generalized time-stability (for general evader) for policy π

2.1. Problem Description

Preliminaries. Let $G := (N, A)$ be a directed network with node and arc sets N and A, respectively, and define $n = |N|$ and $m = |A|$. Also, let c_a denote the cost of traversing arc $a \in A$, and define $c := (c_a \colon a \in A)$. Assume for simplicity that nodes 1 and n are the evader's fixed source and destination nodes, respectively. For a set of arcs $I \subseteq A$, we define $G(I) := (N, A \setminus I)$ as the interdicted graph arising from G when the arcs in I are blocked. With this, we define $z(I)$ as the cost of the shortest $1-n$ path in the interdicted graph $G(I)$, i.e.,

$$z(I) := \min\left\{\sum_{a \in P} c_a \colon \quad P \text{ is an } 1-n \text{ path in } G(I)\right\}, \tag{1}$$

and $S(I)$ as the set of all shortest $1-n$ paths in $G(I)$, i.e.,

$$S(I) := \arg\min\left\{\sum_{a \in P} c_a \colon \; P \text{ is an } 1-n \text{ path in } G(I)\right\}.$$

We assume that the evader has complete knowledge about the graph, including its arc costs. For an example of settings where costs are uncertain to both the interdictor and the evader, see Song and Shen (2016). We also assume that the interdictor knows the graph G, and that she knows c only up to valid lower and upper bounds for its components. That is, she knows that $c \in \mathcal{C}^0$, where

$$\mathcal{C}^0 := \left\{(\hat{c}_1, \hat{c}_2, \ldots, \hat{c}_m) \in \mathbb{R}_+^m \colon \; \ell_a \le \hat{c}_a \le u_a, \, \forall a \in A\right\},$$

and where ℓ_a and u_a denote some finite lower and upper bounds of the cost for arc a, respectively. We refer to $\mathcal{C}^0$ as the initial information available to the interdictor, as it contains her initial knowledge about the network's cost vector.[3]

[3] Note that, unlike in Borrero et al. (2016), we assume that the interdictor initially knows all arcs in the network. The assumption is made without loss of generality, as one can always assume that the network is complete and lower/upper bounds for unknown arcs are set at zero/a (sufficiently) large constant.

In each period $t \in \mathcal{T} := \{0, 1, \ldots, T\}$, within a finite horizon of T periods, the following sequence of events takes place:

(*i*) For the duration of period t, the interdictor blocks the arcs in a set $I^t \subseteq A$ with $|I^t| \leq k$, where k denotes the interdiction budget.[4]

(*ii*) After observing the interdictor's action, the evader travels through a shortest path $P^t \in S(I^t)$, incurring on a cost $z^t \equiv z(I^t)$.

(*iii*) The interdictor obtains feedback $\mathcal{F}^t$ from the evader's actions (we define and discuss the notion of feedback in the next section).

Note that, following prior work, (*ii*) above assumes that the evader acts greedily in each period, thus preventing any strategic consideration on his behalf (see, e.g., Johnson et al. (2014) for a setting with an adaptive evader). We keep such an assumption for now, so as to streamline the exposition, and extend our analysis to more general settings in Section 6. Note that (*ii*) also assumes that the evader observes the interdictor's actions: as outlined in Borrero et al. (2016), we can interpret this assumption in the context of repeated interactions in a stochastic setting, where such monitoring might arise naturally from a learning process of trial-and-error by the evader.

Finally, we assume there are no $1-n$ cuts in G with k or fewer arcs, and hence, there is no trivial solution to the interdictor's problem; we also restrict our attention to policies with $I^0 = \emptyset$. These two assumptions are rather technical and made to simplify our analysis, and thus the exposition.

2.2. Feedback

We define the notion of feedback $\mathcal{F} := (\mathcal{F}^t : t \in \mathcal{T})$ as the sequence of information collected by the interdictor when observing the follower's evasion decisions in each period. We first consider the notion of standard feedback in Borrero et al. (2019).

Definition 1. [**Standard feedback.**] Feedback $\mathcal{F}$ is *standard* if for each period $t \in \mathcal{T}$ the interdictor observes the total cost incurred by the evader, z^t. ■

Standard feedback might arise, for example, in the context of smuggling interdiction, when the interdictor aims at maximizing (minimizing) the probability of detection (evasion). There, each arc cost can be interpreted as (minus the logarithm of) the probability of evasion at different arcs (see, e.g., details in Morton et al. (2007)), in which case standard feedback corresponds to observing the overall probability of evasion. While such a probability can not be observed directly, it can be inferred through repeated interactions between the evader and the interdictor by investigating various types of available data (e.g., prices in illegal markets, enforcement and punishment records

[4] This is, the maximum number of arcs that can be blocked in any period.

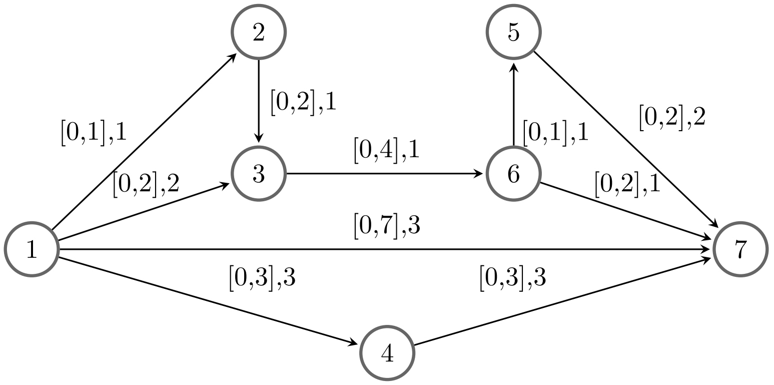


Figure 1 Network G used in Example 1. The labeling of the arcs is given by $[\ell_a, u_a], c_a$.

from the law-enforcement agencies); see examples of the related studies in Buehn and Eichler (2009), Gathmann (2008), Magliocca et al. (2019), Yürekli and Sayginsoy (2010) and the references therein.

Hereafter, unless otherwise noted, we assume that feedback is always standard. In the sequel, we argue that this form of feedback imposes a high toll in terms of time-stability convergence (as defined later in this section). Thus, we define two additional types of feedback with somewhat stronger assumptions on the amount of information collected.

Definition 2. [**Response-imperfect feedback.**] We say standard feedback $\mathcal{F}$ is *response-imperfect* if for each period $t \in \mathcal{T}$, the interdictor learns that the evader used arc $a \in P^t$ with some probability $p_r \in [0,1]$, independently across all arcs and periods. ■

Definition 3. [**Value-imperfect feedback.**] We say response-imperfect feedback $\mathcal{F}$ is *value-imperfect* if for each period $t \in \mathcal{T}$, the interdictor learns the cost of arc $a \in P^t$ with some probability $p_v \in [0,1]$, independently across all arcs and periods. ■

We assume that, for a given arc, the feedback above is nested, and let $P_r^t \subseteq P^t$ and $P_v^t \subseteq P_r^t$ denote the sets of arcs that the interdictor observes and learns their costs under response-imperfect and value-imperfect feedback, respectively. We note that while probabilities p_r and p_v are the same for all arcs, one could consider arc-dependent probabilities at the expense of having a more convoluted notation. Here, we opt to maintain simplicity of exposition and keep arc-independent probabilities. Also, note that if $p_r = 1$, then response-imperfect feedback reduces to response-perfect feedback in Borrero et al. (2019), and if $p_r = p_v = 1$, then value-imperfect feedback reduces to value-perfect feedback in Borrero et al. (2019). The next example illustrates the difference between these notions.

Example 1. Consider graph G depicted in Figure 1. Assume that $k = 2$ and $T = 2$, and suppose that the interdiction decisions are $I^0 = \emptyset$, $I^1 = \{(1,4),(1,7)\}$ and $I^2 = \{(1,7),(3,6)\}$. In such a case, the evader's decisions in each period are given by $P^0 = 1 \to 7$, $P^1 = 1 \to 3 \to 6 \to 7$ and $P^2 = 1 \to 4 \to 7$, with costs $z^0 = 3$, $z^1 = 4$ and $z^2 = 6$, respectively. The information collected by the interdictor from the evader's actions under different feedback types is as follows.

- Under value-perfect feedback, the interdictor observes the total cost and the arcs used by the evader along with their costs in each time period. That is, in period $t=0$ the interdictor observes P^0, $c_{(1,7)}$ and z^0; for $t=1$, the interdictor observes P^1, $c_{(1,3)}$, $c_{(3,6)}$, $c_{(6,7)}$, and z^1; and for $t=2$, the interdictor observes P^2, $c_{(1,4)}$, $c_{(4,7)}$ and z^2.
- Under response-perfect feedback, in each period the interdictor observes the total cost along with the arcs used by the evader, but not the individual arc's costs. That is, for $t=0$, the interdictor learns P^0 and z^0; for $t=1$, she observes that P^1 and z^1; and for $t=2$, the interdictor observes P^2 and z^2.
- Under standard feedback, the information revealed for the interdictor is limited to only the costs of the evasion paths in each period. That is, in periods $t=1$, $t=2$ and $t=3$, the interdictor observes z^0, z^1 and z^2, respectively, but not the actual evasion paths taken.
- Under response-imperfect feedback, for $t=0$, in addition to z^0, the interdictor observes that arc $(1,7)$ is contained in P^0 with probability p_r; for $t=1$, in addition to z^1, the interdictor may learn, for example, that $(1,3)$, $(3,6)$ and $(6,7)$ are part of P^1 (each with probability p_r).
- Under value-imperfect feedback, in addition to the costs of P^0, P^1, and P^2, the arc cost information of the evasion paths can be obtained by the interdictor. For example, for $t=0$, the interdictor might observe that $(1,7)$ is part of P^0 with probability p_r, and given that the interdictor observes $(1,7)$, she also learns $c_{(1,7)}$ with probability p_v. ■

Next, we define the notion of an interdiction policy, and present the class of greedy, robust and non-repetitive policies, which are the main focus of this paper.

2.3. GRN Policies

Preliminaries. An interdiction policy $\pi := (\pi^t \colon t \in \mathcal{T})$ is a deterministic sequence of set functions such that, for each $t \in \mathcal{T}$, $I^{t,\pi} := \pi^t(\mathcal{F}^{s,\pi} : s < t)$ represents the set of arcs blocked in period t, where $\mathcal{F}^{s,\pi}$ represents the feedback obtained under policy π in period $s \in \mathcal{T}$ (for notational convenience, we include the interdictor's actions within such a feedback).[5] For example, in the case of standard feedback, we have that

$$\mathcal{F}^{t,\pi} := \left\{ z(I^{t,\pi}), I^{t,\pi} \right\}, \quad t \in \mathcal{T}.$$

In order to measure the performance of a policy, we focus on minimizing the number of periods that the policy takes to implement solutions that coincide with those taken by an oracle with full-information on the network's costs. Hence, we define *time-stability* of policy π as:

$$\tau^\pi := \min\left\{ t \in \mathcal{T} \colon z^* = z(I^{s,\pi}), \text{ for all } s \geq t \ \right\},$$

[5] In the remainder of the paper whenever necessary we use the superscript π to denote the dependency on policy π.

where

$$z^* := \max\big\{z(I):\ I \subseteq A \text{ s.t. } |I| \leq k\big\}. \tag{2}$$

Observe that z^* is the maximum cost for the evader that can be induced by the interdictor. We use $z^{t,\pi}$ to denote the cost incurred by the evader under interdiction $I^{t,\pi}$ in policy π at time t. Note that because the evader solves for a shortest path problem in each time period, we have $z^{t,\pi} = z(I^{t,\pi})$.

Remark 1. In the online optimization literature, performance is typically measured in terms of a policy's *regret*, which is defined as the cumulative loss in cost incurred by the policy relative to that achieved by an *oracle* decision-maker with complete prior information about the underlying problem data, see Cesa-Bianchi and Lugosi (2006). In the network interdiction setting, regret of policy π until time t is given by $R^{t,\pi} := \sum_{s\leq t}(z^* - z^{s,\pi})$. The concept of time-stability is introduced in Borrero et al. (2016), where it is observed that an upper bound on time-stability implies an upper bound on regret, i.e., $R^{t,\pi} \leq \mu\tau^{\pi}$, where μ is an upper bound for $z^* - z^{s,\pi}$ for any $s \leq t$. ■

Information Update. For any $t \in \mathcal{T}$, set $\mathcal{C}^t$ denotes the interdictor's belief regarding the possible cost vectors (i.e., at t the interdictor knows that $c \in \mathcal{C}^t$). Starting from $\mathcal{C}^0$, the interdictor updates this belief set using feedback $\mathcal{F}^t$. For example, under standard feedback she could update $\mathcal{C}^t$ as:

$$\mathcal{C}^{t+1} = \mathcal{C}^t \cap \left\{\hat{c} \in \mathbb{R}^m_+ :\ \exists P \in S_{\hat{c}}(I^t) \text{ s.t. } \sum_{a\in P} \hat{c}_a = z(I^t)\right\}, \quad t \in \mathcal{T} \setminus \{T\}, \tag{3}$$

where, in a slight abuse of notation, $S_{\hat{c}}(I^t)$ refers to the set of shortest paths on the network $G(I^t)$ when costs are given by vector $\hat{c}$. Note that the update (3) implies that:

(i) For any $\hat{c} \in \mathcal{C}^t$, all the paths in the remaining graph have a cost of at least $z(I^t)$, i.e., $\sum_{a\in P} \hat{c}_a \geq z(I^t)$, for all $1-n$ paths $P \in G(I^t)$;

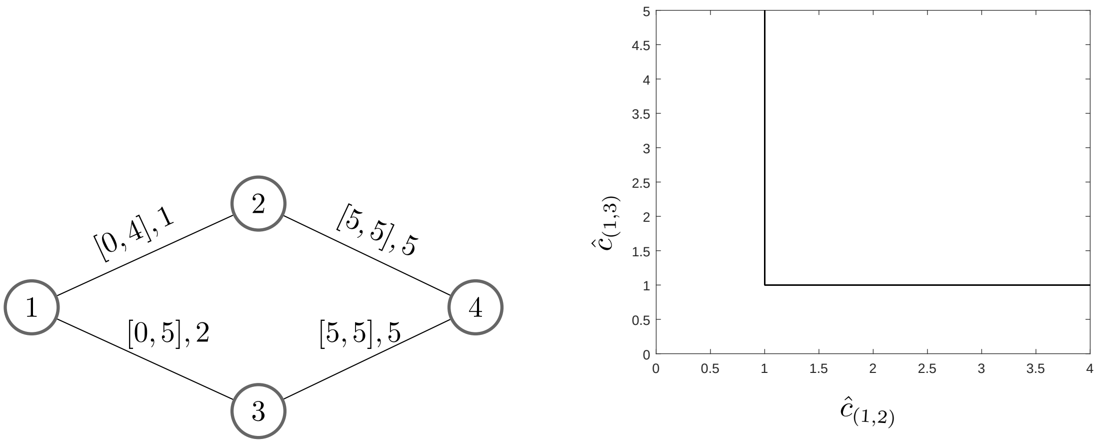
(ii) there is at least one $1-n$ path with cost $z(I^t)$.

While the update mechanism above is the "best" in settings with standard feedback (in the sense that it reduces $\mathcal{C}^t$ the most), we consider alternative mechanisms that are more tractable from an algorithmic point of view. The underlying reason for considering less efficient updates follows from the non-convex nature of the update (3), which we illustrate in the following example.

Example 2. Consider the instance in Figure 2(a). Let $k = 1$. Note that $\mathcal{C}^0 = [0,4] \times [0,5] \times [5,5] \times [5,5]$. Suppose that $I^0 = \emptyset$, so that $P^0 = 1 \to 2 \to 4$ and the interdictor observes $z^0 = 6$. Using update (3) results in $\mathcal{C}^1 = \mathcal{C}^0 \cap \{\hat{c}_{(1,2)} = 1 \text{ and } \hat{c}_{(1,3)} \geq 1\} \cup \{\hat{c}_{(1,2)} \geq 1 \text{ and } \hat{c}_{(1,3)} = 1\}$. Figure 2(b) depicts feasible values for $\hat{c}_{(1,2)}$ and $\hat{c}_{(1,3)}$ in $\mathcal{C}^1$ that form two line segments. Clearly, $\mathcal{C}^1$ is non-convex. ■

Unless otherwise specified, we refer to performance of policies with respect to a generic update mechanism. The latter is assumed to satisfy the following properties:

A1 : $c \in \mathcal{C}^t$ for all $t \in \mathcal{T}$.



(a) Network G used in Example 2

(b) Illustration of non-convexity of $\mathcal{C}^1$ in Example 2

Figure 2 The labeling of arcs in Figure 2(a) is given by $[\ell_a, u_a], c_a$. Figure 2(b) illustrates non-convexity of $\mathcal{C}^1$ in Example 2: the feasible values for $\hat{c}_{(1,2)}$ and $\hat{c}_{(1,3)}$ in $\mathcal{C}^1$ form two line segments.

A2 : $\mathcal{C}^{t+1} \subseteq \mathcal{C}^t$ for all $t \in \mathcal{T} \setminus \{T\}$.

Assumption **A1** indicates that information updates do not rule out the actual cost vector, and **A2** that "uncertainty" surrounding c does not increase in time. Because update mechanisms do not necessarily incorporate all relevant information in $\mathcal{F}^t$, the interdictor cannot rule out the possibility of "getting stuck" on implementing a sub-optimal interdiction action. With these ideas in mind, next we propose a family of policies that ensures that the interdictor does not "get stuck" in such situations *independent of the update mechanism used.*

GRN Policies. In this paper we focus on greedy, robust and non-repetitive (GRN) interdiction policies. These policies are greedy in the sense that, at each period, the interdictor seeks to maximize the immediate cost for the evader; they are robust in that they assume the worst-case (for the evader) arc costs realizations in $\mathcal{C}^t$; and are non-repetitive in a sense that their goal is to avoid solutions implemented previously by the interdictor unless they are optimal.

In order to introduce the GRN policies, define $z_R(I^t)$ as the cost that the interdictor would expect to observe in the worst case scenario (for the evader) when interdicting the set I^t. That is,

$$z_R(I^t) := \min_{P^t} \left\{ \max_{\hat{c}} \left\{ \sum_{a \in P^t} \hat{c}_a : \ \hat{c} \in \mathcal{C}^t \right\} : \ P^t \text{ is a } 1-n \text{ path in graph } G(I^t) \right\}. \tag{4}$$

Remark 2. The right-hand side (r.h.s.) of (4) belongs to the class of robust shortest path problem with absolute robust objective (Yu and Yang 1998). Initially, when $\mathcal{C}^t = \mathcal{C}^0$, it belongs to the class of robust shortest path problems with interval cost, and as such, its inner maximization can be solved simply by setting $\hat{c} = u$. When $\mathcal{C}^t$ is either finite or a polyhedron, the problem is

NP-hard; see Buchheim and Kurtz (2018), Bertsimas and Sim (2003), Poss (2013). In our setting $\mathcal{C}^t$ is not necessarily convex (recall Example 2) and, to the best of our knowledge, such formulations has not been studied so far, with the notable exception of Borrero and Lozano (2020). ■

From the previous section, the information update mechanism determining $\mathcal{C}^t$ might not necessarily incorporate all available information on the cost vector. Thus, the interdictor's expectations should be corrected to account for the fact that, if I^t has been implemented in the past, then she should expect to see the cost $z(I^t)$ if she implements I^t again. Define $\hat{z}_R(I^t)$ as the cost the interdictor would expect to see, accounting for the aforementioned correction. That is,

$$\hat{z}_R(I^t) := \begin{cases} z_R(I^t) & \text{if } I^t \neq I^s \quad \forall s < t, \\ z(I^s) & \text{if } I^t = I^s \text{ for some } s < t. \end{cases} \tag{5}$$

The proposed GRN policies are such that, in each period $t \in \mathcal{T}$, the interdictor greedily implements a solution I^t that maximizes $\hat{z}_R(I^t)$. That is, the GRN policies implement a solution to the problem

$$z_R^{t,*} := \max\big\{\hat{z}_R(I^t) : \ |I^t| \leq k, \ I^t \subseteq A\big\}, \quad t \in \mathcal{T}. \tag{6}$$

For any policy π, let ξ^π denote the earliest time at which the interdictor's expectations, as given in (6), match the observed cost. That is,

$$\xi^\pi := \min\{t \in \mathcal{T} : \ z_R^{t,*} = z^{t,\pi}\}, \tag{7}$$

where one needs to recall that $z^{t,\pi} := z(I^{t,\pi})$, i.e., the cost incurred by the evader given interdiction decision $I^{t,\pi}$. We are ready to define Λ, the set of the GRN policies.

Definition 4. Policy λ belongs to the class of GRN policies Λ if and only if

$$I^{t,\lambda} \in \arg\max\big\{\hat{z}_R(I^t) : \ |I^t| \leq k, \ I^t \subseteq A\big\} \quad \forall t \leq \xi^\lambda,$$

and $I^{t,\lambda} = I^{\xi^\lambda,\lambda}$ for all $\xi^\lambda < t \leq T$. ■

Computing a policy $\lambda \in \Lambda$ requires solving (6) for each $t \in \mathcal{T}$. At first glance, such a formulation is a bilevel optimization problem that is difficult to solve in general. However, in Section 5 we show that for some class of update mechanisms, (6) can be either reduced to, or approximated by, a single-level mixed-integer program, and hence, solved using standard MIP solvers.

Finally, we note that the GRN policies are similar to those proposed in Borrero et al. (2016, 2019), where perfect feedback is assumed. In their setting all information from the (stronger) feedback can be included into $\mathcal{C}^t$ without compromising its convexity, while in our setting it is impossible to do so. Therefore, GRN policies instead "penalize" the interdictor if she repeats a solution that is not optimal (which amounts to a crude but easy way to implement "correction" on expectations). In Sections 3 and 4, we showcase the related properties of GRN policies under different types of feedback.

3. GRN Policies under Standard Feedback

In this section, we analyze the GRN policies under standard feedback. In Section 3.1, we show that time-stability of the GRN policies is guaranteed to converge. Then, in Section 3.2 we justify the greedy, robust and non-repetitive nature of policies in Λ by showing that these qualities are required in order to attain efficiency in a specific sense.

3.1. Convergence

Consider a setting with standard feedback so that, in each period, the interdictor only observes the cost incurred by the evader. The following "sandwich" result, whose proof can be found in Appendix A, provides a stopping criteria for the GRN policies by exploring the relationship between what the interdictor expects to see and what she actually observes.

Theorem 1. *For $t \in \mathcal{T} \setminus \{0\}$ given and $\lambda \in \Lambda$, one has that $z^{t,\lambda} \leq z^* \leq z_R^{t,*}$.*

Theorem 1 provides a certificate of optimality for policies in Λ. That is, whenever what the interdictor observes matches her expectations, her decision is guaranteed to be optimal. Thus, from the interdictor's perspective, the decision-making process at period $t \in \mathcal{T}$ under GRN policies Λ can be described as follows:

(i) The interdictor uses $((z^{s,\lambda}, I^{s,\lambda}): \ s < t)$ to formulate and solve (6), thus finding $z_R^{t,*}$ and $I^{t,\lambda}$.

(ii) The evader incurs on a cost $z^{t,\lambda}$, which is observed by the interdictor.

(iii) The process is repeated until $z_R^{t,*} = z^{t,\lambda}$, and $I^{t,\lambda}$ gets implemented from there on.

By construction, the GRN policies do not repeat solutions unless there is a guarantee about their optimality. This observation is formalized in the next corollary, whose proof follows directly from (5) and (6) and thus is omitted.

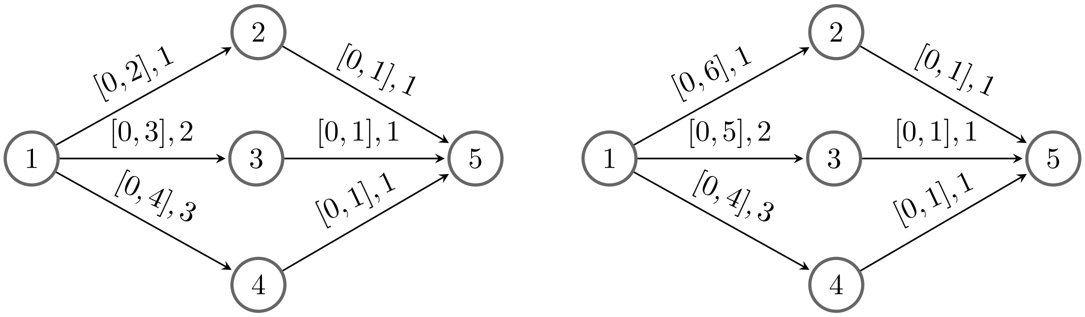
Corollary 1. *For any $\lambda \in \Lambda$, if at time period t $I^{t,\lambda} = I^{s,\lambda}$ for some $s < t$, then $z^{t,\lambda} = z_R^{t,*}$.*

The next result establishes that for any policy $\lambda \in \Lambda$, time-stability is reached in finite time. In particular, it establishes an upper bound on time-stability that is a function of the number of arcs in the network and the interdiction budget.

Proposition 1. *Consider $\lambda \in \Lambda$ and standard feedback. Then,*

$$\tau^\lambda \leq \xi^\lambda \leq \binom{m}{k} + 1.$$

Proposition 1 shows that under standard feedback, policies in Λ (which do not necessarily update the set $\mathcal{C}^t$) may need an exponentially large number of periods to find the full-information solution, in the worst case. In Section 4 we show how different updating mechanisms can be used to improve the performance of the proposed policies in Λ.



(a) Network G_1

(b) Network G_2

Figure 3 Networks used in Remark 3, the labeling of the arcs is given by $[\ell_a, u_a], c_a$

Remark 3. The next example suggests that the bound in Proposition 1 is tight. Consider graphs G_1 and G_2 depicted in Figure 3.

Consider $k = 2$, and note that in network G_1 a policy $\lambda \in \Lambda$ satisfies $\tau^{\lambda} = 5$. Indeed, when $t = 0$ we have that $I^{0,\lambda} = \emptyset$, $z^{0,\lambda} = 2$; then when $t = 1$, $I^{1,\lambda} = \{(1,2),(1,3)\}$, $z_R^{1,*} = 5$, $z^{1,\lambda} = 4$; $t = 2$, $I^{2,\lambda} = \{(1,2),(3,5)\}$, $z_R^{2,*} = 5$, $z^{2,\lambda} = 4$; $t = 3$, $I^{3,\lambda} = \{(1,3),(2,5)\}$, $z_R^{3,*} = 5$, $z^{3,\lambda} = 4$; $t = 4$, $I^{4,\lambda} = \{(2,5),(3,5)\}$, $z_R^{4,*} = 5$, $z^{4,\lambda} = 4$; $t = 5$, $I^{5,\lambda} = \{(1,2),(1,3)\}$, $z_R^{5,*} = 4$, $z^{5,\lambda} = 4$, and up to this point, the optimal solution is obtained.

In network G_2, on the other hand, time-stability is exactly the upper bound in Proposition 1, which is $\binom{m}{k} + 1 = 16$. (See Appendix B for the sequence of interdiction and evasion actions.) ■

The next result formalizes the tightness of the bound in Proposition 1.

Proposition 2. *For any $k \geq 0$ and $n \geq k+3$, there exist $\alpha \in (0,1]$, a graph G and information update mechanism such that, if $T > \binom{m}{k}$, then $\tau^{\lambda} \geq \alpha(\binom{m}{k} + 1)$.*

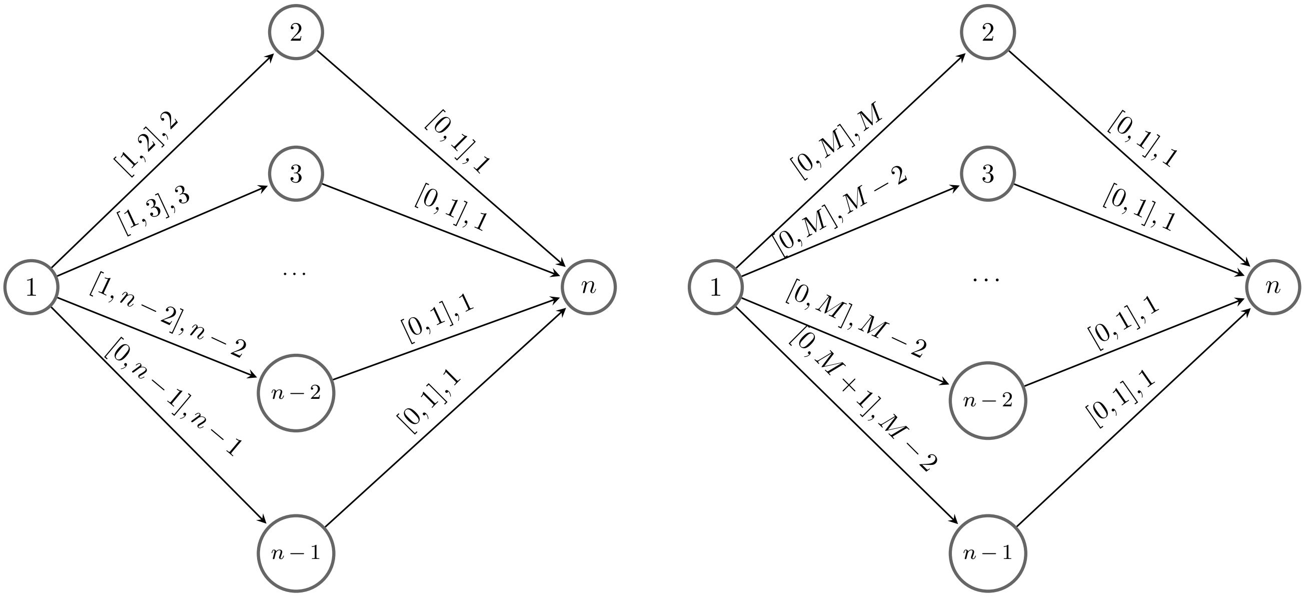
3.2. Necessity of Being Greedy, Robust and Non-repetitive

In this section we argue that it is necessary for the interdictor to act *consistently* in a greedy, robust, and non-repetitive manner. Our starting point here is a policy that is greedy, robust and non-repetitive, all at the same time; we show that removing one of such features might deteriorate policy performance. Formally, we say that a policy $\pi \notin \Lambda$ is *non-consistently* greedy if for every instance there exist time periods (which might depend on the instance) such that $I^{t,\pi}$ does not solve (6). Define in an analogous way non-consistently robust and non-consistently non-repetitive policies. Next, we give counter-examples showing that for any non-consistent policy π there exist instances such that $\tau^{\lambda} < \tau^{\pi}$ for all $\lambda \in \Lambda$.

Necessity of Being Greedy. Assume π is a non-consistently greedy policy and consider an instance where the upper bound and the real cost are the same for each arc. Thus, the optimal solution is to directly block the k-most vital arcs, which is the solution for the GRN policies because

$u_a = c_a$ for all $a \in A$. Consequently, the GRN policies can find an optimal solution in the first time period, that is, $\tau^\lambda = 1$. However, under π at time s the interdictor tries other non-greedy solution which implies that a non-k-most vital solution is implemented at time s. In other words, $z_R^{s,\pi} < z^*$ and hence, in this instance $\tau^\pi > \tau^\lambda = 1$ for any $\lambda \in \Lambda$.

Necessity of Being Robust. Consider the network G_1 depicted in Figure 4(a) and let $k = n-3$ $(n \geq 4)$. Under any GRN policy, at time $t = 1$ the interdictor blocks $I^{1,\lambda} = \{(1,2),(1,3),\ldots,(1,n-2)\}$, which coincides with the solution under full information I^* of value n. Moreover, it follows that $\tau^\lambda = 1$ for any λ. Suppose next that the interdictor is not robust, and instead assumes that at some time $s \geq 1$ the cost vector is a convex combination between the lower and upper bounds such that $I^{t,\pi} \neq I^{t,\lambda}$. For instance, assume that the interdictor uses the weight $\sigma \in (0,1)$ to combine lower and upper bounds, with σ such that $\sigma(n-1) < \sigma(n-2) + 1 - \sigma$, i.e., $0 < \sigma < \frac{1}{2}$. Then we can see that the path $1 \to (n-1) \to n$ is evaluated by a lower cost, and since π is greedy, we have that $I^{s,\pi} = \{(1,2),(1,3),\ldots,(1,n-3),(1,n-1)\}$, $z_R^{s,*} = \sigma(n-2) + 1 - \sigma + \sigma = \sigma(n-2) + 1$, $z^{s,\pi} = n-1$. It can be concluded that $\tau^\pi > s \geq 1 = \tau^\lambda$, as desired.



(a) Network G_1

(b) Network G_2

Figure 4 **Networks used in the discussion in Section 3.2, the labeling of the arcs is given by** $[\ell_a, u_a], c_a$

Necessity of Non-repetitiveness. Consider graph G_2 in Figure 4(b), and note that the cost for every path is $M - 1$ except for that of path $1 \to 2 \to n$, which is $M + 1$. The robust cost for every path is $M + 1$ except for that of path $1 \to (n-1) \to n$, which is $M + 2$. Let $k = n - 3$. Thus, the full-information optimal solution is $I^* = \{(1,3),(1,4),\ldots,(1,n-1)\}$ and $z^* = M + 1$.

At $t=0$, $I^{0,\lambda}=\emptyset$, $P^0=1\to 3\to n$, and the interdictor observes $z^{0,\pi}=M-1$. At $t=1$, a GRN policy interdicts paths $1\to 2\to n, 1\to 3\to n, \ldots, 1\to n-2\to n$ and expects to observe $z_R^{1,*}=M+2$. This decision results in $P^1=1\to(n-1)\to n$ and the interdictor observes $z^{1,\lambda}=M-1$. Note that in the next time period, with (5) and (6), if the interdictor repeats $I^{1,\lambda}$, then $z^{2,\lambda}=z^{1,\lambda}=M-1$. However, there are other better solutions given by not repeating $I^{1,\lambda}$; for example, if $I^{2,\lambda}=\{(1,2),(1,3),\ldots,(1,n-3),(1,n-1)\}$, $z_R^{2,*}=M+1>z^{1,\lambda}$. Therefore, the solution is improved by forcing the interdictor to explore different solutions.

Finally, we note that a non-consistently non-repetitive policy that solves (6), without the non-repetitiveness constraints, always have $I^{t,\pi}=I^{1,\lambda}=\{(1,2),(1,3),\ldots,(1,n-2)\}$ for all $t\geq 2$. This solution is suboptimal under full information.

4. GRN Policies under Imperfect Feedback

In this section, we consider the properties of the policies in Λ under imperfect feedback. Recall from Definition 2 that under response-imperfect feedback, the interdictor observes a set $P_r^t\subseteq P^t$, and that under value-imperfect feedback, the interdictor also learns the costs of the arcs in a further subset $P_v^t\subseteq P_r^t$. Because there is uncertainty surrounding the feedback, we measure the performance of the GRN policies using the expected time-stability criterion.

We begin analyzing how the feedback in each setting can be used to update the beliefs on the cost vector. Let $\mathcal{R}^t$ denote the cost vectors at time t that agree with the information of the response-imperfect feedback. Recalling that P_r^t is the subset of arcs in P^t observed by the evader, we have that

$$\mathcal{R}^t := \Big\{\hat{c}\in\mathbb{R}^m:\ \exists P\in S_{\hat{c}}(I^t) \text{ s.t. } P_r^t\subseteq P \text{ and } \sum_{a\in P}\hat{c}_a = z(I^t)\Big\},$$

where we recall that $S_{\hat{c}}(I^t)$ refers to shortest paths in $G(I^t)$ when costs are given by vector $\hat{c}$. In this case, the "best" (most informative) update mechanism is given by $\mathcal{C}^{t+1}=\mathcal{C}^t\cap\mathcal{R}^t$.

Similarly, let $\mathcal{V}^t$ denote the set of cost vectors in period t which satisfy the additional information given by value-imperfect feedback, i.e.,

$$\mathcal{V}^t := \big\{\hat{c}\in\mathbb{R}^m:\ \hat{c}_a=c_a,\ \forall\, a\in P_v^t\big\}, \tag{8}$$

thus, we have that, in this setting, the best update mechanism is given by $\mathcal{C}^{t+1}:=\mathcal{C}^t\cap\mathcal{R}^t\cap\mathcal{V}^t$.

The next result establishes an upper bound on the expected time-stability for the case when $\mathcal{C}^t$ is updated only with the information contained in $\mathcal{V}^t$. In particular, it shows that $\mathbb{E}(\tau^\lambda)=O(m)$ for fixed values of p_r and p_v.

Proposition 3. *Let $\lambda\in\Lambda$ and consider value-imperfect feedback, where $p_r>0$ and $p_v>0$. If the interdictor updates the uncertainty set by $\mathcal{C}^{t+1}=\mathcal{C}^t\cap\mathcal{V}^t$ for all $t\in\mathcal{T}$, then*

$$\mathbb{E}(\tau^\lambda)\leq\frac{m}{p_r p_v}.$$

The upper bound for $\mathbb{E}(\tau^\lambda)$ in Proposition 3 may be loose. However, the next result establishes a lower bound on the probability that the time-stability is of the same order as the upper bound.

COROLLARY 2. *Let $\lambda \in \Lambda$ and consider value-imperfect feedback, where $p_r > 0$ and $p_v > 0$. If the interdictor updates the uncertainty set by $\mathcal{C}^{t+1} = \mathcal{C}^t \cap \mathcal{V}^t$ for all $t \in \mathcal{T}$, then there exists $0 < \alpha < 1$ such that*

$$\Pr\Big(\tau^\lambda > \gamma\alpha \frac{m}{p_r p_v}\Big) \geq (1-\gamma)^2\alpha^2/2,$$

for any $0 \leq \gamma \leq 1$.

We end this section by noting that, just as in the case of standard feedback, updates involving $\mathcal{R}^t$ in settings under response-imperfect and value-imperfect feedback result in sets that are not necessarily convex. Such non-convexity implies that problem (6) cannot be reformulated or approximated in a straightforward way into a single-level MIP problem, which is a common approach to solving multi-level optimization problems, see, e.g., Audet et al. (1997), Zare et al. (2019) and our further discussions in Section 5.

With these considerations in mind, we explore the "weak" update $\mathcal{R}^t_w$ (that is weaker than $\mathcal{R}^t$). In particular, under response-imperfect feedback, because $P^t_r \subseteq P^t$, we consider the following alternative update mechanism

$$\mathcal{C}^{t+1} = \mathcal{C}^t \cap \mathcal{R}^t_w := \mathcal{C}^t \cap \Big\{\hat{c} \in \mathbb{R}^m : \sum_{a \in P^t_r} \hat{c}_a \leq z^{t,\lambda}\Big\}. \tag{9}$$

Similarly, we define a "weak" update mechanism under value-imperfect feedback as follows

$$\begin{aligned} \mathcal{C}^{t+1} &= \mathcal{C}^t \cap \mathcal{R}^t_w \cap \mathcal{V}^t \\ &= \mathcal{C}^t \cap \Big\{\hat{c} \in \mathbb{R}^m : \sum_{a \in P^t_r} \hat{c}_a \leq z^{t,\lambda}\Big\} \cap \Big\{\hat{c} \in \mathbb{R}^m : \ \hat{c}_a = c_a \text{ for } a \in P^t_v\Big\} \\ &= \mathcal{C}^t \cap \Big\{\hat{c} \in \mathbb{R}^m : \sum_{a \in P^t_r \setminus P^t_v} \hat{c}_a \leq z^{t,\lambda} - \sum_{a \in P^t_v} c_a, \ \hat{c}_a = c_a \text{ for } a \in P^t_v\Big\}. \end{aligned} \tag{10}$$

Under these weak update mechanisms, the uncertainty set $\mathcal{C}^t$ is a polyhedron, in which case the r.h.s. of (4) belongs to the class of robust shortest path problems with a polyhedral uncertainty set, which are NP-hard (see Remark 2). While, the three-level problem (6) remains computationally difficult in general, it is better suited for approximation methods. For this reason, in Section 5 we propose computing a certain approximation of (6) in each time period, which can be implemented using an off-the-shelf MIP solver.

5. Computing GRN Policies and Their Approximations

The ability to solve problem (6) depends on the update mechanism used: we know that under the strongest update, uncertainty sets are not necessarily convex, and said problem is in general intractable. Hence, in this section we focus on polyhedral uncertainty sets, which arise, for example, when considering the weak updates introduced in the previous section. As mentioned in Remark 2, robust shortest path problems with polyhedral uncertainty are also NP-hard in general and require specialized solution approaches. Thus, we focus on the development of approximate policies that are more tractable and can be implemented using off-the-shelf MIP solvers. Note that the approximate nature of the proposed policies arise from approximately solving (6), in the context of Definition 4, rather than from focusing on providing approximability guarantees with respect to $z_R^{t,*}$. We show that the resulting approximate GRN policies enjoy the same theoretical properties as GRN policies, in particular, with respect to their convergence.

5.1. Preliminaries

Consider decision variable x^t to denote the interdictor's decisions in period t. That is,

$$x_a^t = \begin{cases} 1 & \text{if arc } a \text{ is interdicted,} \\ 0 & \text{otherwise,} \end{cases}$$

thus $I^t = \{a \in A \colon x_a^t = 1\}$. (In the sequel we use $x^{t,\pi}$ and $I^{t,\pi}$ interchangeably when it is clear from the context.) We impose that $x^t \in X$, where $X := \left\{x^t \in \{0,1\}^m : \sum_{a\in A} x_a^t \leq k\right\}$. For each node $i \in N$, we define the sets of outgoing and incoming arcs as $\delta^+(i)$ and $\delta^-(i)$, respectively. For a given decision x^t, the evader traverses through a $1-n$ shortest path, which admits the following linear programming formulation:

$$z(I^t) = \min_y \ c^\top y \tag{11a}$$

$$\text{s.t. } \ y_a \leq 1 - x_a^t \ \forall a \in A, \tag{11b}$$

$$\sum_{a\in\delta^+(i)} y_a - \sum_{a\in\delta^-(i)} y_a = \begin{cases} 1 & i=1, \\ -1 & i=n, \\ 0 & i \in A\setminus\{1,n\}, \end{cases} \tag{11c}$$

$$y_a \geq 0 \ \forall a \in A. \tag{11d}$$

Constraints (11b) ensure that the evader cannot use interdicted arcs. Constraints in (11c) correspond to a network flow formulation of the shortest path problem, see Ahuja et al. (1993). For brevity, we write constraints (11c) as $\boldsymbol{B}y = b$, where $\boldsymbol{B}$ is the node-arc adjacency matrix induced by the graph and $b = [1,0,\ldots,0,-1]^\top \in \mathbb{R}^n$.

Consider now reformulating (6). For $s < t$, define decision variable v^s as:

$$v^s = \begin{cases} 0 & \text{if } x^t = x^s \\ 1 & \text{otherwise.} \end{cases}$$

Note that by time $t > s$, x^s is known and fixed, as well as is $z(I^s)$. Then, defining $z_R(x^t) \equiv z_R(I^t)$, we can write (6) as follows:

$$\begin{aligned} z_R^{t,*} = \max_{x^t,f,v^s} \; & f && && \text{(12a)}\\ \text{s.t. } & f - z(I^s) \le M\, v^s && \forall s < t, && \text{(12b)}\\ & f - z_R(x^t) \le M \sum_{s<t} (1 - v^s), && && \text{(12c)}\\ & \left\|x^t - x^s\right\|_1 \le n\, v^s \le n \left\|x^t - x^s\right\|_1 && \forall s < t, && \text{(12d)}\\ & x^t \in X, v^s \in \{0,1\} \;\forall s < t, && && \text{(12e)} \end{aligned}$$

where $\|\cdot\|_1$ denotes an ℓ_1 norm, and M is a sufficiently large constant; for example, we can set $M = (n-1)\max_{a\in A}\{u_a\}$. Constraints (12b)-(12d) encourage the interdictor to explore new solutions not implemented previously. Formally, if $x^t = x^s$ for some $s < t$, then $v^s = 0$ from (12d) and constraints (12b) ensure that f is equal to z^s. However, if $v^s = 1$ for all $s < t$, then constraints (12c) force f to take the value of $z_R(x^t)$.

Two observations are due. First, while constraints (12d) are nonlinear, they can be linearized using standard techniques. Second, the term $z_R(x^t)$ in constraints (12c) admit the following reformulation:

$$z_R(x^t) = \min_y \big\{ \max_{\hat{c}} \{ (\hat{c} + M\ x^t)^\top y :\ \hat{c} \in \mathcal{C}^t \} :\ \boldsymbol{B} y = b,\ y \in \{0,1\}^m \big\}. \tag{13}$$

Note that this formulation has an additional penalty term $(M\,x^t)$ in its objective function, so that y_a is forced to be 0 whenever $x_a^t = 1$ (see Israeli and Wood (2002) for more details on this reformulation approach). Thus, a constraint similar to (11b) is not needed in problem (13).

5.2. Approximate GRN Policy under Polyhedral Uncertainty Sets

Consider settings where $\mathcal{C}^t$ can be written as a polyhedron. In particular, assume that

$$\mathcal{C}^t = \{\hat{c} \in \mathbb{R}_+^m :\ \boldsymbol{G}^t\, \hat{c} \le g^t\},$$

where $\mathcal{C}^0 = \{\hat{c} \in \mathbb{R}_+^m : \boldsymbol{G}^0\, \hat{c} \le g^0\}$, $\boldsymbol{G}^0 = [\boldsymbol{I}; -\boldsymbol{I}]$, $g^0 = [u_1, \ldots, u_m, -\ell_1, \ldots, -\ell_m]^\top$ and $\boldsymbol{I}$ is the identity matrix in $\mathbb{R}^{m\times m}$. Polyhedral uncertainty sets arise, for example, when using the weak update mechanisms (9) and (10) from Section 4, under response- and value-imperfect feedback, respectively. In these settings, (13) is a robust shortest path problem with polyhedral uncertainty, which is known to be NP-hard (Buchheim and Kurtz 2018). Furthermore, because of (13), standard techniques in bilevel optimization that reformulate (12) as a single-level problem (see, e.g., Audet et al. (1997), Zare et al. (2019)), cannot be applied. While it is customary to solve such difficult problems using tailored decomposition methods (see, e.g., Zeng and Zhao (2013)), here we focus on policies that can be implemented using off-the-shelf MIP solvers. With that in mind, consider a relaxation of (13), where we drop the integrality restrictions for y, i.e.,

$$\underline{z}_R(x^t) = \min_y \big\{ \max_{\hat{c}} \{ (\hat{c} + M\ x^t)^\top y :\ \hat{c} \in \mathcal{C}^t \} :\ \boldsymbol{B} y = b,\ y \in \mathbb{R}^m_+ \big\}. \tag{14}$$

We propose the use of approximate GRN policies that replace $\hat{z}_R(x^t)$ in (5) with

$$\hat{\underline{z}}_R(x^t) := \begin{cases} \underline{z}_R(x^t) & \text{if } x^t \neq x^s \quad \forall s < t, \\ z(I^s) & \text{if } x^t = x^s \text{ for some } s < t. \end{cases} \tag{15}$$

Thus, the approximate policy implements an interdiction solution x^t that solves the problem

$$\underline{z}_R^{t,*} = \max_{x^t} \{ \hat{\underline{z}}_R(x^t) : x^t \in X \}$$

in each time period t. That is, the approximate GRN policies solve (12) after replacing $z_R(x^t)$ in constraint (12c) with its relaxed version $\underline{z}_R(x^t)$ as defined above, resulting in the following optimization problem:

$$\underline{z}_R^{t,*} = \max_{x^t, f, v^s} \ f \tag{16a}$$

$$\text{s.t. } f - z^s \le M\, v^s \qquad \forall s < t, \tag{16b}$$

$$f - \underline{z}_R(x^t) \le M \sum_{s<t} (1 - v^s), \tag{16c}$$

$$\left\| x^t - x^s \right\|_1 \le n\, v^s \le n\, \left\| x^t - x^s \right\|_1 \qquad \forall s < t, \tag{16d}$$

$$x^t \in X, v^s \in \{0,1\}\ \forall s < t. \tag{16e}$$

Formally, we define the set of approximate GRN policies $\underline{\Lambda}$ as follows.

Definition 5. Policy λ belongs to the set of approximate GRN policies $\underline{\Lambda}$ if and only if $x^{t,\lambda}$ solves (16) for $t \le \underline{\xi}^\lambda$, and $x^{t,\lambda} = x^{\underline{\xi}^\lambda, \lambda}$ for all $\underline{\xi}^\lambda < t \le T$, where for a policy λ we define

$$\underline{\xi}^\lambda := \min\{ t \in \mathcal{T} :\ \underline{z}_R^{t,*} = z^{t,\lambda} \}.$$

The robust nature of GRN policies can be interpreted as emanating from the assumption that the follower implements a solution to the Stackelberg game in (13) between the follower and the nature: once the follower selects an $1-n$ path, the nature responds by selecting the cost vector that is least favorable for the follower. Under this interpretation, approximate GRN policies can be viewed as policies that allow the follower to commit to a mixed strategy; see, e.g. von Stengel and Zamir (2010) for an analysis of Stackelberg games with mixed strategies.

We show next that the theoretical properties discussed in Sections 3 and 4 continue to hold for the case of the proposed approximate GRN policies. For that, we first establish that the inequalities in Theorem 1 continue to hold:

PROPOSITION 4. *For $t \in \mathcal{T} \setminus \{0\}$ given and $\lambda \in \underline{\Lambda}$, one has that $z^{t,\lambda} \leq z^* \leq \underline{z}_R^{t,*}$.*

This result implies that the convergence results (e.g., Propositions 1 and 3, and Corollary 2) established for GRN policies also hold for their approximations in Definition 5.

Under the game-theoretic interpretation of approximate GRN policies, Proposition 4 says that, while the follower's commitment to mixed strategies should improve its position, it does not lead to unattainable costs (in optimality). In this regard, the results holds more generally, independent of the properties of $\mathcal{C}^t$, as long as **A1** ($c \in \mathcal{C}^t$) holds (and nature's problem is well defined).

Next, we reformulate (16) as a single-level MIP for the case of polyhedral uncertainty sets, which enables implementation of approximate GRN policies using off-the-shelf solvers. Our starting point is (14), which for polyhedral uncertainty sets becomes:

$$\underline{z}_R(x^t) = \min_y \left\{ \max_{\hat{c}} \{ (\hat{c} + M\ x^t)^\top y : \ \boldsymbol{G}^t \hat{c} \leq g^t \} : \ \boldsymbol{B} y = b,\ y \in \mathbb{R}_+^m \right\}.$$

For any given y, the inner maximization of the objective function above is a linear program. Thus, we can use strong duality to obtain the following single-level reformulation:

$$\begin{aligned} \underline{z}_R(x^t) := \min_{y,p}\ & (g^t)^\top p + (Mx^t)^\top y \\ \text{s.t. } & (\boldsymbol{G}^t)^\top p = y, \\ & \boldsymbol{B} y = b, \\ & y \in \mathbb{R}_+^m,\ p \in \mathbb{R}_+^{|g^t|}. \end{aligned}$$

Defining Q as the feasible region in the formulation above, we have that constraints (16c) become

$$\min\{ (g^t)^\top p + (Mx^t)^\top y : \ (y,p) \in Q \} \geq f - M \sum_{s<t} (1 - v^s).$$

Noting that the formulation in the l.h.s. above is a linear program, we use strong duality once again, and conclude that f and v^s satisfy constraints (16c) if and only if there exist vectors $\hat{c}$ and w satisfying the following constraints:

$$b^\top w \geq f - M \sum_{s<t} (1 - v^s), \quad \boldsymbol{G}^t \hat{c} \leq g^t, \quad \boldsymbol{B}^T w - \hat{c} \leq M x^t.$$

Thus, summarizing the above, we have that for the case of polyhedral uncertainty sets, formulation (16) admits the following MIP reformulation:

$$\begin{aligned} \mathrm{MIP}(\boldsymbol{G}^t, g^t) := \max_{x^t, f, v^s, p}\ & f && && (17a) \\ \text{s.t. } & f - z^s \leq M v^s && \forall s < t, && (17b) \\ & f - b^\top w \leq M \sum_{s<t} (1 - v^s), && && (17c) \end{aligned}$$

$$\boldsymbol{G}^t \hat{c} \leq g^t \tag{17d}$$

$$\boldsymbol{B}^\top w - \hat{c} \leq M x^t, \tag{17e}$$

$$\left\|x^t - x^s\right\|_1 \leq n\, v^s \leq n\, \left\|x^t - x^s\right\|_1 \qquad \forall s < t, \tag{17f}$$

$$x^t \in X, v^s \in \{0,1\}\ \forall s < t. \tag{17g}$$

Next, we show that under some conditions, approximate GRN polices coincide with GRN polices, i.e., $\underline{\Lambda} \equiv \Lambda$, and the latter set can also be computed using single-level MIPs.

5.3. Special case: GRN Policies without Uncertainty Set Updates

Consider a setting without any uncertainty set update, i.e., $\mathcal{C}^t = \mathcal{C}^0$ for all $t \in \mathcal{T}$. Recall that $\mathcal{C}^0 = \{\hat{c} \in \mathbb{R}^m :\ \ell_a \leq \hat{c}_a \leq u_a\ \forall a \in A\}$. Then the inner maximization problem of (13) admits an optimal solution $u := (u_1, \ldots, u_m)^T$, where u is the vector of arc costs' upper bounds. Thus, problem (13) reduces to the following single-level MIP:

$$z_R(x^t) = \min_y \{(u + Mx^t)^\top y :\ \boldsymbol{B} y = b,\ y \in \{0,1\}^m\}.$$

Moreover, recall that $\boldsymbol{B}$ is the node-arc adjacency matrix induced by graph $G(N, A)$. Hence, $\boldsymbol{B}$ is totally unimodular and $z_R(x^t)$ can be computed as the following linear program (LP):

$$z_R(x^t) = \min_y \{(u + Mx^t)^\top y :\ \boldsymbol{B} y = b,\ y \in \mathbb{R}^m_+\},$$

where the integrality restrictions for y are relaxed. This observation also implies that approximate GRN policies coincide with GRN policies whenever $\mathcal{C}^t = \mathcal{C}^0$ for all $t \in \mathcal{T}$.

Using strong duality of the above LP formulation, we can further rewrite constraint (12c) as

$$\max_p \{b^T p :\ \boldsymbol{B}^\top p \leq u + Mx^t\} \geq f - M \sum_{s<t} (1 - v^s).$$

Moreover, f and v^s satisfy the above constraint if and only if there exists a vector $p \in \mathbb{R}^m$ such that $b^\top p \geq f - M \sum_{s<t} (1 - v^s)$ and $\boldsymbol{B}^\top p \leq u + Mx^t$. Therefore, formulation (12) reduces to:

$$\text{MIP}(\mathcal{C}^0) := \max_{x^t, f, v^s, p}\ f \tag{18a}$$

$$\text{s.t. } f - z^s \leq M v^s \qquad \forall s < t, \tag{18b}$$

$$f - b^\top p \leq M \sum_{s<t} (1 - v^s), \tag{18c}$$

$$\boldsymbol{B}^\top p \leq u + Mx^t, \tag{18d}$$

$$(12\text{d}) - (12\text{e}),$$

which is a single-level MIP model.

Finally, from the above discussion it is also clear that approximate GRN policies coincide with GRN policies whenever only value-imperfect feedback with $\mathcal{V}^t$, see (8), is used. For the latter, the same MIP can be applied after replacing u_a with the appropriate value of $\hat{c}_a$ whenever it is observed.

6. The Case of a General Evader

In this section we analyze the implications of relaxing our assumption on the evader's response, namely that his response must be a shortest path in the interdicted network. Instead, we consider settings in which such a response is constrained to be any valid $1-n$ path. The latter implies that the evader may commit to implement non-optimal responses to the interdictor's actions indefinitely. That is, the cost incurred by the evader and then observed by the interdictor after an interdiction $I^{t,\pi}$ is not necessarily equal to the shortest path cost in $G(I^{t,\pi})$, i.e., $z^{t,\pi} \geq z(I^{t,\pi})$ for any policy π. Thus, it is necessary to revisit the performance criterion used.

When the evader responds with a shortest path, time-stability provides a bound on a policy's regret. That is,

$$R^{t,\pi} = \sum_{s \leq t} (z^* - z^{s,\pi}) \leq \mu \tau^{\pi},$$

where μ is an upper bound on $(z^* - z^{s,\pi})$ for each s; recall Remark 1.

Note that, when the evader's response is instead a valid $1-n$ path, there is no guarantee that $z^* \geq z^{s,\pi}$. Nonetheless, one can bound the regret by discarding periods on which the evader's actions are sub-optimal. Specifically, we have that

$$R^{t,\pi} = \sum_{s \leq t} (z^* - z^{s,\pi}) \leq \tilde{R}^{t,\pi} := \sum_{s \leq t} (z^* - z^{s,\pi})^{+} \leq \mu \tilde{\tau}^{\pi}, \tag{19}$$

where $(\cdot)^{+}$ denotes the positive part of the argument, and we refer to $\tilde{R}^{t,\pi}$ as the *generalized regret.* Similarly, we define the *generalized time-stability* as:

$$\tilde{\tau}^{\pi} := 1 + \left| \{ t \in \mathcal{T} \, : \, z^* > z^{t,\pi} \} \right|. \tag{20}$$

Note that the generalized time-stability: (i) coincides with the traditional time-stability when the evader responds in a greedy fashion, for the proposed policies, as they both indicate the time it takes the interdictor to achieve the full information solution; and (ii) accounts for the time periods on which the evader effectively takes advantages of the interdictor's initial uncertainty to increase the regret.

Information Update. In this setting, information updates are weaker in the sense that only the fact of the existence of a response can be incorporated, not its optimality (from the evader's perspective). For example, under standard feedback, the update in (3) becomes

$$\mathcal{C}^{t+1} = \mathcal{C}^{t} \cap \left\{ \hat{c} \in \mathbb{R}^{m}_{+} : \; \exists P \in S_{\hat{c}}(I^{t}) \text{ s.t. } \sum_{a \in P} \hat{c}_a \leq z^{t} \right\}, \quad t \in \mathcal{T} \setminus T. \tag{21}$$

Note however, that properties **A1** and **A2** continue to hold under this update. Also, as expected, the update continues to be non-convex in general, as illustrated next.

Example 3. Consider the instance in Figure 2(a), and suppose that $k=1$. Note that $\mathcal{C}^0 = [0,4]\times[0,5]\times[5,5]\times[5,5]$. Suppose that $I^0=\emptyset$, so that $P^0 = 1\to 2\to 4$ and the interdictor observes $z^0=6$. Using update (21) results in $\mathcal{C}^1 = \mathcal{C}^0 \cap \{\hat{c}_{1,2}=1 \text{ or } \hat{c}_{1,3}=1\}$, which is non-convex. ■

GRN Policies. The proposed policies operate as in the case of a greedy evader, i.e., the interdictor assumes that the evader's response is an $1-n$ shortest path; hence, the expected cost $z_R(I^t)$ continues to be given by (4). However, corrections to such an expectation on solutions implemented in the past must account for the fact that said solutions might have not been $1-n$ shortest paths. With this, (5) is replaced by:

$$\tilde{z}_R(I^t) := \begin{cases} z_R(I^t) & \text{if } I^t \neq I^s \quad \forall s<t, \\ \min\{z^s :\ s<t,\, I^s = I^t\} & \text{if } I^t = I^s \quad \text{for some } s<t. \end{cases} \tag{22}$$

Moreover, we define

$$\tilde{z}_R^{t,*} := \max\{\tilde{z}_R(I^t) :\ I^t \subseteq A, |I^t| \leq k\},\ \forall t \in \mathcal{T}$$

as the problem that the interdictor solves in each time period. Thus, we define the tailored GRN policies (Λ^G) as follows.

Definition 6. Policy $\lambda \in \Lambda^G$ if and only if $I^{t,\lambda} = I^{t-1,\lambda}$ when $z^{t-1,\lambda} \geq \tilde{z}_R^{t-1,*}$, and

$$I^{t,\lambda} \in \arg\max\big\{\tilde{z}_R(I^t) :\ |I^t| \leq k,\ I^t \subseteq A\big\},$$

otherwise. ■

Convergence under Standard Feedback. As emphasized earlier, unlike in Section 3, the observed costs do not necessarily provide a lower bound to the expected cost, as the evader might act suboptimaly at any time. This observation is formalized in Lemma 1, which is the equivalent of Theorem 1 in this more general setting.

Lemma 1. *For $t \in \mathcal{T}\setminus\{0\}$ given and $\lambda \in \Lambda^G$, one has that $z^* \leq \tilde{z}_R^{t,*}$.*

Note, however, that because the expected cost $\tilde{z}_R^{t,*}$ is a valid upper bound to z^*, if the observed cost turns out to be not lower than $\tilde{z}_R^{t,*}$, then this implies that the evader is acting suboptimaly. Hence, such a period does not contribute to increasing the modified regret $\tilde{R}^{t,\pi}$. Thus, the optimality certificate alluded in Theorem 1 still applies in the sense that, whenever the observed cost is greater than the expected one, then the interdictor is sure that the regret is not growing. Similar to Section 5, when the update mechanism used results in a polyhedral uncertainty set, we can compute approximate GRN policies via the single-level MIP formulation (17) (see below).

As in Section 3, the non-repetitive nature of the proposed policies ensure the finiteness of the generalized time-stability $\tilde{\tau}^\lambda$, which is formalized next.

Proposition 5. *Consider $\lambda \in \Lambda^G$ and standard feedback. Then,*

$$\tilde{\tau}^{\lambda} \leq \binom{m}{k} + 1.$$

The tightness of the bound above follows from Proposition 1, as the greedy evader is a particular case of the general one.

One can see that information updates under different forms of imperfect feedback admit rather straightforward extensions to the case of general evader. Also, the bounds derived in Section 4 hold for the generalized time-stability and the weaker information updates from Section 5.2 are also applicable for the case of a general evader.

MIP Formulations for approximate GRN Policies. Relative to the case of greedy evader, computation of approximate policies differ in that the cost expectations and information updates ought to be adjusted differently. With respect to the the first of these issues, from Section 5 we see that (6) can be written as

$$\tilde{z}_R^{t,*} = \max_{x^t, f, v^s} \; f \tag{23a}$$

$$\text{s.t. } f - \min\{z(I^u) : \; u \leq t, I^u = I^s\} \leq M\, v^s \qquad \forall s < t, \tag{23b}$$

$$(12c) - (12e) \text{ hold.} \tag{23c}$$

Thus, similar MIP formulations for approximate policy computation can be derived, provided that information updates maintain the polyhedral representation of the uncertainty set. This is certainly the case when there is no update (see Section 5.3). As for the weak update mechanisms explored for the case of imperfect feedback, as noted above, they are both compatible with the weak feedback available for the general evader. Thus, the formulations in Section 5.2 still apply to this more general setting with minor modifications to take into account that (17b) is replaced by (23b).

7. Computational Study

In Section 7.1, we describe our test instances and three benchmark policies, which are compared against the GRN policies in Section 7.2. In Section 7.3, we explore to what degree the performance of the approximate GRN policies depends on the information revealed to the interdictor. Sections 7.4 and 7.5 perform sensitivity analysis of the approximate GRN policies with respect to the quality of feedback and the initial information available to the interdictor. In Section 7.6, we study the performance of our policies for a general evader.

7.1. Test Instances, Benchmark Policies, and Implementation Details

Graph and cost structure. We test our policies on three different graph instances: uniform random graphs (Erdös and Rényi 1959), layered graphs (Bastert and Matuszewski 2001), and Watts-Strogatz graphs (Watts and Strogatz 1998). For brevity, we focus our analysis on the results for uniform random graphs; the results for the latter two types of graphs are fairly similar and thus, they are provided in Appendices D and E.

The uniform random graphs used in this paper are generated following the model of Erdös and Rényi (1959).[6] Based on the cost structure we divide our instances into four categories: *random*, *right-skewed*, *symmetric* and *left-skewed*. For the random cost structure, ℓ_a, c_a and u_a are three integers randomly generated from $[0,50]$ for each arc $a \in A$, and then sorted so that $\ell_a \leq c_a \leq u_a$. For the other types of the cost structure, for each arc $a \in A$, we first generate ℓ_a and u_a randomly from uniform integer distributions $U(0,50)$ and $U(\ell_a,50)$, respectively. Consequently, cost c_a is computed as $\ell_a + (u_a - \ell_a)\beta_a$, where β_a is drawn from a Beta(δ,θ) distribution. We set (δ,θ) to be $(2,10)$, $(10,2)$ and $(10,10)$ for the left-skewed, right-skewed and symmetric cost structure, respectively.

Benchmark policies. Similar to Borrero et al. (2016), for settings with no uncertainty set updates, we consider three benchmark policies. For each of them, the interdictor is greedy and non-repetitive, but does not consider a worst-case realization for the cost, but instead inputs a (single) value for c based on the known lower and upper bounds. Specifically:

- *Lower bound policies* Π_L: the interdictor assumes that the cost of each arc is given by

$$\hat{c}_a = \ell_a \quad \forall a \in A.$$

- *Mean bound policies* Π_M: the interdictor assumes that the costs of arcs are given as

$$\hat{c}_a = (\ell_a + u_a)/2 \quad \forall a \in A.$$

- *Random bound policies* Π_R: for each arc $a \in A$, the interdictor randomly chooses either lower or upper bound of the arc as its real cost. That is,

$$\hat{c}_a = \begin{cases} \ell_a & \text{with probability } \frac{1}{2}, \\ u_a & \text{with probability } \frac{1}{2}. \end{cases}$$

Using the techniques in the previous sections, we reformulate the problem faced by the interdictor each period as the single-level MIP (17) where $\hat{c}$ is evaluated based on the above realizations and this is no longer a decision variable. Note that when no uncertainty set update is implemented ($\mathcal{C}^t = \mathcal{C}^0$ for all $t \in \mathcal{T}$), the problem faced by an interdictor implementing benchmark policies is

[6] That is, a uniform random graph with n nodes is generated in a way that for each pair of nodes, there is an arc between them with probability p. We use $p = 0.5$ in all our experiments.

(18) where u is replaced with above realizations of $\hat{c}$. Note that as discussed in Section 5.3, when we assume that $\mathcal{C}^t = \mathcal{C}^0$ for all $t \in \mathcal{T}$, GRN policies reduce to upper bound policies where the cost vector is composed with upper bounds; i.e. $\hat{c}_a = u_a$ for all $a \in A$. However, when we have more information in the update mechanism, e.g., $\mathcal{R}^t_w$ and $\mathcal{V}^t$, approximate GRN policies are more effective than the upper-bound policies.

Implementation details. The algorithms are coded in C++ using CPLEX 12.6 as the MIP solver. The experiments are performed on a Windows PC with 3.7 GHz CPU and 32 GB RAM.

Table 2 Performance of policies in Λ and benchmark policies without information updates ($n = 15$, uniform random graph, greedy evader).

Graph		k=2 (T=500)				k=4 (T=1000)			
		λ	π_L	π_M	π_R	λ	π_L	π_M	π_R
Right-skewed	Time-stability	**42.6**	450.2^{18}	201.2^{8}	450.2^{18}	$\mathbf{181.9}^{1}$	950.1^{19}	501.0^{10}	1000.0^{20}
	MAD	55.9	89.6	239.0	89.6	189.4	94.8	499.0	0.0
	Relative difference	**0.0%**	21.2%	3.8%	21.6%	**0.7%**	23.6%	4.2%	18.6%
	MAD	0.0%	13.5%	4.9%	10.4%	1.3%	14.2%	4.8%	7.4%
	Total regret	**104.9**	4850.0	925.0	5081.5	**727.4**	13900.0	2350.0	10759.3
	MAD	147.6	2835.0	1195.0	2323.4	956.7	8700.0	2655.0	4874.5
Symmetric	Time-stability	244.5^{8}	350.6^{14}	$\mathbf{74.7}^{2}$	400.4^{16}	871.4^{17}	800.4^{17}	$\mathbf{232.3}^{4}$	800.6^{16}
	MAD	178.9	209.2	104.5	159.4	218.6	319.4	329.3	319.1
	Relative difference	7.3%	16.7%	**0.3%**	12.4%	17.5%	16.7%	**1.0%**	10.3%
	MAD	9.3%	13.9%	0.6%	9.7%	10.0%	11.3%	1.6%	7.3%
	Total regret	1498.2	3075.0	**101.7**	2698.4	6857.7	7300.0	**442.3**	4561.0
	MAD	1254.9	2647.5	147.1	2206.9	3016.1	5060.0	656.9	2996.1
Left-skewed	Time-stability	408.4^{15}	$\mathbf{176.3}^{7}$	243.4^{7}	237.3^{8}	1000.0^{20}	$\mathbf{501.0}^{12}$	912.2^{17}	729.7^{12}
	MAD	137.4	226.6	182.1	213.8	0.0	499.0	149.3	327.2
	Relative difference	26.9%	**4.3%**	7.2%	10.2%	36.6%	**4.5%**	14.6%	18.7%
	MAD	18.3%	5.6%	9.4%	12.5%	17.1%	4.9%	24.6%	22.3%
	Total regret	3191.8	**450.0**	1255.9	1317.3	9715.2	**1300.0**	6268.1	7228.4
	MAD	1983.8	585.0	882.2	1175.9	4479.9	1450.0	3309.9	5729.6
Random	Time-stability	338.4^{12}	425.3^{17}	$\mathbf{211.4}^{6}$	401.1^{16}	1000.0^{20}	850.5^{19}	$\mathbf{702.0}^{13}$	919.1^{18}
	MAD	194.0	127.0	217.2	158.2	0.0	254.2	362.4	145.6
	Relative difference	16.8%	27.4%	**6.8%**	17.5%	36.3%	29.5%	**15.9%**	21.0%
	MAD	16.8%	17.5%	9.7%	15.0%	14.0%	17.2%	15.6%	14.9%
	Total regret	3066.4	4825.0	**1255.9**	3210.0	14264.3	13400.0	**6849.7**	9478.0
	MAD	2329.5	3190.0	1474.8	2822.2	5500.0	8440.0	4803.6	6237.4

Notes: Entries in bold denote the best policy in each setting; the numbers in superscript of time-stability denote the number of instances out of 20 for which the corresponding policy failed to converge within T time periods.

7.2. Comparison of Policies without Information Updates

In our first set of the computational experiments we compare the performance of the GRN policies $\lambda \in \Lambda$ against the benchmark policies when no information updates are used, i.e., $\mathcal{C}^t = \mathcal{C}^0$ for all periods t. We test the policy performance using the right-skewed, symmetric, left-skewed and random cost structures as outlined in Section 7.1. For each structure we randomly generate 20 instances with $n = 15$. Finally, we have $k \in \{2, 4\}$, and set either $T = 500$ or $T = 1000$, respectively.

We compare the policies' average time-stability and average total regret, as well as their mean absolute deviation (MAD). Also, we compute the *relative difference* between the cost returned by the full information optimal solution, and the cost observed from the evader under a policy either at time T or at the time period when the interdictor starts repeating the same solution, i.e.,

$$\frac{\left| z^{\hat{t},\pi} - z^* \right|}{z^*} \cdot 100\%,$$

where $\hat{t}$ is the time period when $z^{\hat{t},\pi} = z^{\hat{t}+1,\pi} = \ldots = z^{T,\pi}$ holds. We use the average relative difference and its MAD to measure policy performance. Table 2 summarizes the performance of policies Λ and the benchmark polices across all cost structures. There, λ, π_L, π_M and π_R denote the policies in Λ, Π_L, Π_M and Π_R, respectively.

It appears that policies λ and π_M outperform the other policies in general as they perform reasonably well in all of the scenarios and perform best in most of them. In particular, policy λ performs best in settings with right-skewed costs, and policy π_M performs best in settings with symmetric and random costs. In contrast, policy π_L performs reasonably well only with left-skewed costs which are arguably favorable to π_L.

Comparing among different settings, we see that for right-skewed costs, GRN policies perform significantly better than all other benchmark policies, which is rather intuitive given its robust nature. However, λ loses its relative advantage for $k = 4$, especially for networks with left-skewed and random costs. Furthermore, observe that the benchmark policies fail to find an optimal solution in at least one of the instances for all cost structures within T steps, see the numbers in superscript of Table 2. Recall that unless the interdictor performs in a robust manner (as in Λ policies), there is no guarantee that a policy achieves an optimal solution, see our discussion in Section 3.2.

A similar setting as the one shown above is presented in Borrero et al. (2016). There, the authors demonstrate that the greedy and pessimistic policies are better than the benchmark policies and, moreover, that all the instances are solved to optimality. In contrast to our experiments, in Borrero et al. (2016) the authors assume perfect feedback, which yields the worst-case time-stability linear upper bound of $|A|$. Here, given the limited feedback, time-stability for the GRN policies is exponentially bounded by the number of arcs. As we set T to be small for the sake of computational tractability, there are some instances where the GRN policies take more than T periods to find an optimal solution. However, for sufficiently large values T the GRN policies should outperform the benchmark ones, under standard feedback for at least some test instances.

To summarize the discussion above we conclude that the GRN polices demonstrate overall good performance across all cost structures for sufficiently small values of k as shown by the results in Table 2 for $k = 2$. On the other hand, the benchmark policies fail to converge for at least one instance in each of the considered cost structures. The performance of the GRN policies deteriorates significantly as k increases as shown by the results in Table 2 for $k = 4$. These observations are not surprising given the worst-case time-stability result derived in Proposition 2. In addition, the observed results emphasize that with very limited feedback (i.e., only the total cost of the evader's shortest path is revealed to the interdictor) it is rather difficult to converge to a full-information optimal interdiction solution, in particular, as the value of k increases. Similar results hold for layered and Watts-Strogatz graphs, see Appendices D.2 and E.2, respectively.

7.3. Improvements When Information Updates Are Applied

In this section, we study the performance of approximate GRN policies under response-imperfect (λ_r) and value-imperfect feedback (λ_v) and compare their performance to that of policies that assume standard feedback (λ). Recall from Section 5.3 that under standard feedback, when uncertainty sets are either not updated or updated only through value-imperfect updates with $\mathcal{V}^t$, then approximate GRN and GRN policies coincide. We implement the weaker versions of the response- and value-imperfect updates, as defined in (9) and (10), respectively, due to the convexity requirement for the uncertainty set updates in our MIP models, see Sections 2.3 and 5.2.

We consider graphs with a number of nodes $n \in \{15, 20, \ldots, 50\}$ for all four cost structures. The time horizon is set as $T = 500$ and for each cost structure we randomly generate 20 instances. For response-imperfect feedback we set $p_r = 0.5$, and in value-imperfect feedback we set $p_r = p_v = 0.5$. We measure policy performance using the average time-stability and MAD, see Table 3. Furthermore, because the mean-bound policy π_M performs better than the other benchmark policies in Section 7.2, we also choose to explore the performance of π_M with the value-imperfect updates, see the results in the column denoted by $\pi_M(+\mathcal{V})$ in Table 3.

As expected, the performance of the policies improve as more information is revealed to the interdictor, see the results for λ_v, λ_r and $\pi_M(+\mathcal{V})$ in Table 3. We also observe that policies in λ_v and λ_r significantly outperform the other considered policies, which emphasizes the importance of having sufficiently good information feedback.

Finally, we note that policies π_M are outperformed by policies λ_v and λ_r even when additional information is used from value-imperfect feedback as in $\pi_M(+\mathcal{V})$. More importantly, policies π_M and $\pi_M(+\mathcal{V})$ fail to converge for at least one test instances in all cost structures. These observations are consistent with our theoretical derivations and negative examples discussed in Section 3. Similar results hold for layered and Watts-Strogatz graphs, see Appendices D.3 and E.3, respectively.

Table 3 Average time-stability and MAD (in parenthesis) for $\lambda \in \underline{\Lambda}$ and $\pi_M \in \Pi_M$ policies when information updates are applied ($k = 6$, $T = 500$, uniform random graphs, greedy evader).

n	Right-skewed					Symmetric					Left-skewed					Random				
	λ	λ_r	λ_v	π_M	$\pi_M(+\mathcal{V})$	λ	λ_r	λ_v	π_M	$\pi_M(+\mathcal{V})$	λ	λ_r	λ_v	π_M	$\pi_M(+\mathcal{V})$	λ	λ_r	λ_v	π_M	$\pi_M(+\mathcal{V})$
15	178.5^{4}	33.0^{1}	**6.4**	350.6^{14}	325.7^{13}	475.3^{19}	47.8^{1}	**13.7**	134.3^{5}	103.1^{4}	500.0^{20}	29.1	**20.2**	476.1^{19}	20.8	500.0^{20}	33.6	**20.7**	453.0^{18}	231.4^{9}
	(198.7)	(46.7)	(3.3)	(209.2)	(226.6)	(52.8)	(45.3)	(5.7)	(182.9)	(158.8)	(0.0)	(7.5)	(5.5)	(45.5)	(8.0)	(0.0)	(11.0)	(6.5)	(84.7)	(241.7)
20	307.9^{10}	11.9	**7.1**	350.6^{14}	350.6^{14}	500.0^{20}	50.8^{1}	**16.5**	196.8^{7}	129.8^{5}	500.0^{20}	57.6^{1}	**20.8**	500.0^{20}	27.6	500.0^{20}	41.4	**21.5**	475.2^{19}	113.3^{4}
	(205.9)	(6.1)	(3.5)	(209.2)	(209.2)	(0.0)	(45.1)	(4.8)	(212.3)	(185.1)	(0.0)	(44.2)	(7.6)	(0.0)	(8.3)	(0.0)	(19.7)	(4.5)	(47.1)	(154.7)
25	317.3^{12}	11.7	**6.0**	350.6^{14}	350.6^{14}	500.0^{20}	31.0	**17.6**	179.4^{7}	54.5^{2}	500.0^{20}	40.6	**32.1**	500.0^{20}	30.6	500.0^{20}	40.2	**24.7**	500.0^{20}	260.0^{10}
	(219.3)	(4.7)	(2.9)	(209.2)	(209.2)	(0.0)	(15.5)	(3.7)	(224.4)	(89.1)	(0.0)	(12.1)	(17.1)	(0.0)	(9.4)	(0.0)	(10.6)	(7.1)	(0.0)	(240.1)
30	362.1^{14}	15.6	**8.8**	375.5^{15}	375.5^{15}	500.0^{20}	30.1	**16.2**	205.8^{7}	153.9^{6}	500.0^{20}	44.3	**27.5**	500.0^{20}	27.9	500.0^{20}	35.9	**23.8**	500.0^{20}	96.7^{3}
	(193.1)	(8.2)	(3.3)	(186.8)	(186.8)	(0.0)	(12.3)	(6.0)	(235.4)	(207.7)	(0.0)	(10.7)	(6.9)	(0.0)	(6.2)	(0.0)	(12.2)	(5.7)	(0.0)	(121.0)
35	$314.5^{1}2$	18.3	**8.4**	400.4^{16}	375.5^{15}	500.0^{20}	36.0	**15.3**	111.1^{7}	55.6^{2}	500.0^{20}	40.3	**27.3**	500.0^{20}	36.5	500.0^{20}	39.2	**23.3**	500.0^{20}	239.2^{9}
	(22.7)	(9.9)	(7.2)	(159.4)	(186.8)	(0.0)	(14.5)	(5.6)	(155.6)	(88.9)	(0.0)	(8.6)	(7.0)	(0.0)	(7.8)	(0.0)	(16.2)	(5.4)	(0.0)	(234.7)
40	332.8^{13}	12.2	**7.23**	325.7^{13}	400.4^{16}	500.0^{20}	29.8	**16.8**	242.9^{7}	105.3^{4}	500.0^{20}	66.1^{1}	**25.4**	500.0^{20}	34.9	500.0^{20}	65.5^{1}	**29.2**	500.0^{20}	167.7^{6}
	(217.4)	(7.8)	(4.7)	(226.6)	(159.4)	(0.0)	(9.4)	(3.8)	(231.4)	(157.9)	(0.0)	(47.9)	(7.0)	(0.0)	(8.6)	(0.0)	(4.5)	(6.5)	(0.0)	(199.4)
45	380.3^{14}	22.9	**7.3**	425.3^{17}	400.4^{16}	500.0^{20}	27.6	**17.5**	468.6^{18}	179.5^{7}	500.0^{20}	69.5	**27.0**	500.0^{20}	40.6	500.0^{20}	42.3	**28.4**	500.0^{20}	166.1^{6}
	(167.6)	(13.9)	(2.5)	(127.0)	(159.4)	(0.0)	(6.3)	(4.1)	(56.6)	(224.4)	(0.0)	(46.3)	(7.5)	(0.0)	(9.7)	(0.0)	(19.9)	(8.9)	(0.0)	(200.3)
50	293.3^{11}	21.4	**10.8**	425.3^{17}	425.3^{16}	500.0^{20}	34.5	**17.7**	366.2^{12}	130.6^{5}	500.0^{20}	50.2	**25.4**	500.0^{20}	58.7^{1}	500.0^{20}	50.2	**27.3**	500.0^{20}	195.1^{7}
	(227.4)	(13.9)	(4.7)	(127.0)	(127.0)	(0.0)	(9.8)	(5.3)	(166.5)	(184.7)	(0.0)	(22.5)	(7.2)	(0.0)	(44.1)	(0.0)	(18.8)	(9.3)	(0.0)	(213.4)

Note. The numbers in superscript of time-stability denote the number of instances out of 20 for which the corresponding policy failed to converge within T time periods.

7.4. Policy Performance: Sensitivity with Respect to p_r and p_v

We now study the approximate GRN policies' performance under response-imperfect and value-imperfect feedback as a function of the probability of learning information. To this end, we set $k = 6$ and $n = 50$ for all the experiments and generate 20 instances with the different cost structures.

Figure 5 depicts the behaviour of time-stability for response- and value-imperfect updates with the right-skewed costs (see Appendix C for the results with the symmetric and left-skewed costs). As expected, the time-stability of both policies decreases as p_r and p_v increase. Note that λ_v policies have better time-stability than λ_r policies; moreover, their time-stability also decreases faster than that of λ_r. These observations show that the policies are highly sensitive to the availability of information, and emphasize the importance of having access to a high quality information feedback. For the results on layered and Watts-Strogatz graphs, see Appendices D.4 and E.4, respectively.

7.5. Policy Performance: Sensitivity with Respect to the Quality of the Bounds

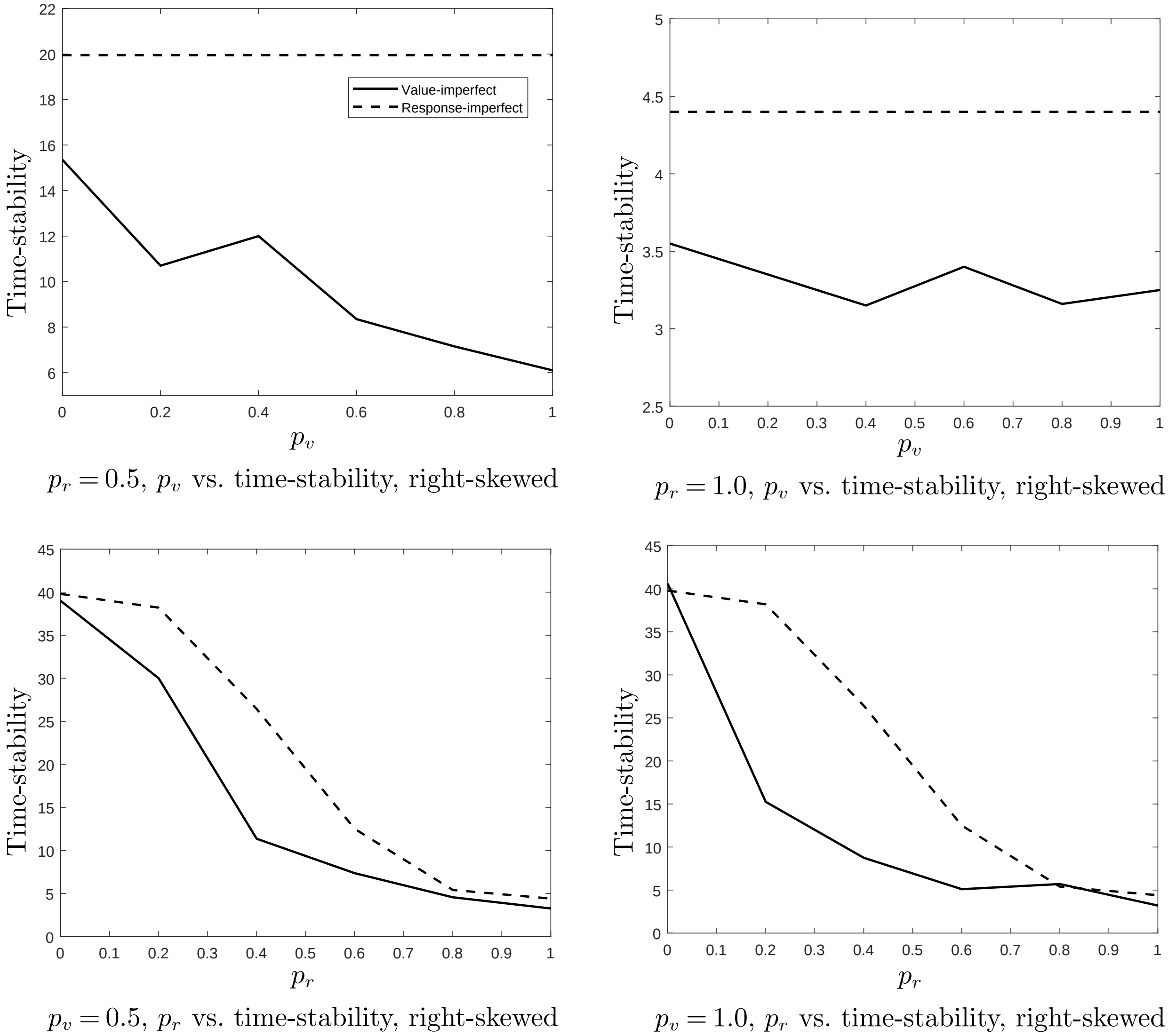
Next, we study the performance of our approximate GRN policies with respect to the quality of initial information, that is, the magnitude of $u_a - l_a$ for all $a \in A$. The value of c_a is generated uniformly from $U(500, 1000)$ and, as in Borrero et al. (2016), we divide the test instances into three categories: $(c_a - \chi_a^-, c_a + \chi_a^+)$, $(c_a - 5\chi_a^-, c_a + 5\chi_a^+)$ and $(c_a - 25\chi_a^-, c_a + 25\chi_a^+)$, where χ_a^- and χ_a^+ are drawn uniformly from $[1, 20]$ for all $a \in A$. We refer to these three sets of instances as "I.1", "I.2" and "I.3", respectively. Clearly, I.1 has the best quality bounds, and I.3 has the worst quality bounds. We generate 20 instances for I.1, I.2 and I.3, and set $k = 6$ and $T = 200$. We consider uniform random graphs with $n = 50$, and study policy performance for various values for probabilities p_r and p_v. Table 4 summarizes the results. The results on layered and Watts-Strogatz graphs can be found in Appendices D.5 and E.5, respectively.

The results show that policy performance is rather sensitive with respect to the quality of initial information; particularly, as the width of the intervals increase, the time-stability increases. Note that the effect is amplified under response-imperfect feedback. In addition, the effect that the quality of bounds have on policy performance is smaller when p_r and p_v take larger values, i.e., when the interdictor can learn more information from the evader's actions. This behavior is indicative of an important trade-off, namely, in order to improve the performance of the GRN policies, the interdictor can either seek to improve the quality of the initial deterministic information, or seek to improve the probabilities of observing real information from the evader.

7.6. Policy Performance for the Case of a General Evader

In this section we consider the generalization introduced in Section 6 for the case of a general evader, who may implement a non-optimal response (i.e., a path that is not necessarily shortest)

Figure 5 Average time-stability for policies in $\underline{\Lambda}$ for different types of feedback as p_r and p_v increase for $k=6$ and the right-skewed costs ($T=50$, uniform random graphs, greedy evader).



$p_r=0.5$, p_v vs. time-stability, right-skewed

$p_r=1.0$, p_v vs. time-stability, right-skewed

$p_v=0.5$, p_r vs. time-stability, right-skewed

$p_v=1.0$, p_r vs. time-stability, right-skewed

to the interdictor's actions in each time period. In our experiments we assume that the evader's feasible solutions are contained in the path set:

$$\widetilde{S}^t = S(I^t) \cup \{P:\ P \in S(I^t \cup a)\ \forall a \text{ such that } \exists P' \in S(I^t),\, a \in P'\}, \tag{24}$$

that is, $\widetilde{S}^t$ contains shortest paths in the interdicted network along with an additional set of evasion paths generated in the following manner. For every shortest path in the interdicted graph, we assume that at least one arc in the path cannot be used by the evader and then generate another evasion path, namely, the shortest possible one, that does not contain the said arc in the interdicted graph. We repeat this procedure for each arc in all shortest paths. By construction, the set $\widetilde{S}^t$

Table 4 Behaviour of policies in $\underline{\Lambda}$ with respect to the cost bound quality ($k=6$, $T=200$, uniform random graphs, greedy evader).

	$p_r = p_v$		0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1
Policy λ_r	I.1	Time-stability	122.5^{11}	76.0^{2}	27.4	20.5	9.9	6.9	5.3	4.9	3.8	3.3	2.8
		MAD	85.3	58.0	13.6	10.0	5.2	3.4	2.2	2.8	1.3	1.1	0.6
		Relative difference	0.6%	0.1%	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
		MAD	0.6%	0.2%	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
	I.2	Time-stability	200.0^{20}	179.5^{17}	83.9^{1}	44.4	26.4	16.8	11.7	8.5	7.0	6.4	5.2
		MAD	0.0	34.9	37.1	16.7	10.3	4.4	4.5	2.0	1.3	1.4	0.6
		Relative difference	7.1%	4.7%	0.2%	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
		MAD	1.4%	2.6%	0.5%	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
	I.3	Time-stability	200.0^{20}	194.0^{18}	132.0^{3}	67.3	36.6	30.9	16.2	19.9	10.2	9.3	7.4
		MAD	0.0	10.9	42.3	14.9	11.5	17.4	3.4	10.5	1.6	1.5	0.8
		Relative difference	34.6%	28.2%	3.15%	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
		MAD	5.5%	9.0%	5.4%	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Policy λ_v	I.1	Time-stability	130.4^{12}	58.5^{1}	23.8	11.7	8.0	5.7	4.3	3.0	3.2	2.1	1.8
		MAD	83.5	42.1	11.9	6.8	3.9	2.9	2.0	1.1	1.8	0.8	0.6
		Relative difference	0.6%	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
		MAD	0.6%	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
	I.2	Time-stability	200.0^{20}	161.4^{11}	63.2	25.9	17.4	13.8	9.0	6.0	5.9	4.6	4.1
		MAD	0.0	44.1	19.0	9.8	5.2	4.5	2.8	1.7	1.5	0.8	0.6
		Difference	7.1%	1.4%	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
		MAD	1.4%	1.6%	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
	I.3	Time-stability	200.0^{20}	197.3^{18}	90.9	43.5	27.1	19.9	13.7	10.2	7.7	6.6	5.7
		MAD	0.0	5.0	33.1	14.1	7.1	4.0	2.6	2.1	1.2	0.8	0.5
		Difference	34.7%	16.4%	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
		MAD	5.3%	10.4%	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0

Note. The numbers in superscript of time-stability denote the number of instances out of 20 for which the corresponding policy failed to converge within T time periods

contains shortest paths and the second shortest paths in $G(I^t)$, along with some additional evasion paths, which have at least one arc distinct from a shortest path.

Given $\widetilde{S}^t$ we consider three types of randomized evasion policies:

- "Random–0.9" policy: in each time period t the evader chooses the shortest path from $\widetilde{S}^t$ with probability 0.9 and the second shortest path from $\widetilde{S}^t$ with probability 0.1.
- "Random–0.5" policy: in each time period t the evader chooses either the shortest path from $\widetilde{S}^t$ or the second shortest path from $\widetilde{S}^t$ with equal probability.
- "Random–All" policy: in each time period t randomly chooses one of the paths from $\widetilde{S}^t$ with equal probability.

As mentioned in Section 7.6, we compute the approximate version of GRN policies based on its single-level MIP formulation. Because the mean-bound policies Π_M are the only ones that are comparable to our GRN policies, we compare the performance of tailored approximate GRN policies $\lambda_v^G \in \underline{\Lambda}_v^G$ (refer to Definition 6) and mean-bound policies $\pi_M^G \in \Pi_M^G$, under value–imperfect feedback. Note that policies Π_M^G follow the same pattern of as those in $\underline{\Lambda}^G$ except for the calculation of $\tilde{z}_R^{t,*}$, which evaluates $\hat{c}$ with $\frac{\ell+u}{2}$.

We summarize the policy performance under Random–All evader in Table 5. The results under Random–0.9 and Random–0.5 evaders can be found in Table 6 and Table 7, respectively (see Appendix C.2). Note that we set $T = 100$ and calculate the generalized time-stability $\tilde{\tau}^\pi$ as defined in (20) and generalized total regret $\tilde{R}^{T,\pi}$ as in (19). The results show that the approximate GRN policies outperform mean-bound policies in many cases, for example, for the graphs with the right-skewed, left-skewed and random cost structure, only lagging behind π_M for some symmetric instances. These results show that the GRN policies can retain their performance over strategic evaders, outperforming π_M in this class of challenging instances.

8. Conclusions

This paper studies the sequential shortest path network interdiction problem in a directed graph, where the interdictor has incomplete information about the arc costs and limited feedback from the evader's actions. By observing feedback from the evader's actions, the interdictor adjusts her decisions so as to maximize the total cumulative cost incurred by the evader.

We study settings with various forms of feedback and propose the GRN policies, a class of policies that follow some rather simple rules, which can also be approximated by solving mixed integer optimization programs, for a certain class of tractable uncertainty updates. With the performance of a policy measured by time-stability, we show that such policies find an optimal solution within $O(\binom{m}{k})$ periods under standard feedback. If more information is available in the feedback, then we show that the interdictor finds an optimal solution with the expected time-stability that is linear in terms of the number of arcs of the network.

We also extend our analysis to settings where the evader does not necessarily respond optimally. By generalizing the concept of time-stability, we show that GRN policies can be adapted so that their theoretical guarantees are preserved. These results imply, for example, that the proposed policies (and the principles behind them) are robust with respect to possible strategic behaviour on the evader's side, and that there is a limit on the advantage that the evader might have because of the interdictor's initial limited information.

Our theoretical results are supported by the numerical experiments. Relative to benchmark policies, GRN policies are guaranteed to find optimal solutions across different types of graphs.

Table 5 **Average time-stability and MAD (in parenthesis) for $\lambda_v^G \in \underline{\Lambda}_v^G$ and $\pi_M^G \in \Pi_M^G$ policies (with value-imperfect feedback) for the general evader ("Random–all" policy, see Section 7.6) on uniform random graphs with $k = 6$ and $T = 100$. Note that the "Random-all" evader randomly chooses a path from the set of evasion paths given by $\widetilde{S}^t$, see (24).**

n		Right-skewed		Symmetric		Left-skewed		Random	
		λ_v^G	$\pi_M^G(+\mathcal{V})$	λ_v^G	$\pi_M^G(+\mathcal{V})$	λ_v^G	$\pi_M^G(+\mathcal{V})$	λ_v^G	$\pi_M^G(+\mathcal{V})$
15	Time-stability ($\tilde{\tau}$)	**21.4**	50.6	14.5	**6.1**	21.7	**19.3**	**18.8**	38.2
	MAD	23.9	39.6	7.5	5.7	6.1	8.6	6.1	26.1
	Total regret ($\tilde{R}^T$)	595.3	717.5	118.3	48.1	161.3	128.7	172.3	245.0
	MAD	851.9	874.5	60.8	15.4	49.4	54.0	51.3	122.0
20	Time-stability ($\tilde{\tau}$)	**29.6**	48.8	**17.0**	19.0	**22.0**	25.0	**18.9**	33.8
	MAD	35.2	33.5	5.1	18.9	6.6	7.2	7.4	20.6
	Total regret ($\tilde{R}^T$)	871.5	948.3	106.1	74.9	143.7	127.3	161.1	176.4
	MAD	1200.9	1161.7	30.4	58.8	53.4	29.9	60.5	97.4
25	Time-stability ($\tilde{\tau}$)	**17.2**	40.2	14.8	**6.9**	**22.9**	23.0	**22.9**	35.8
	MAD	16.6	30.4	4.5	7.6	6.4	8.2	7.2	22.7
	Total regret ($\tilde{R}^T$)	290.7	364.3	91.8	40.0	129.1	108.1	159.9	229.4
	MAD	413.9	426.1	29.5	23.8	35.5	41.3	54.6	158.6
30	Time-stability ($\tilde{\tau}$)	**35.8**	60.9	**25.1**	25.4	**20.8**	30.6	**21.3**	27.5
	MAD	38.6	39.2	15.7	27.5	5.3	8.1	4.8	12.6
	Total regret ($\tilde{R}^T$)	844.1	901.2	236.5	213.1	90.7	101.1	134.8	137.6
	MAD	1092.7	1054.5	268.6	272.6	27.1	25.6	35.5	58.4
35	Time-stability ($\tilde{\tau}$)	**25.8**	63.1	16.1	**10.7**	**23.9**	28.6	**21.2**	42.9
	MAD	29.7	37.2	5.8	12.7	6.5	9.5	5.9	18.7
	Total regret ($\tilde{R}^T$)	597.9	674.6	88.3	33.4	110.2	119.3	145.0	166.5
	MAD	859.6	794.7	32.0	21.0	39.5	49.8	61.6	75.7
40	Time-stability ($\tilde{\tau}$)	**26.4**	70.6	**14.4**	16.6	**24.0**	27.9	**25.3**	40.0
	MAD	29.4	33.7	4.7	19.0	6.2	9.7	6.3	21.4
	Total regret ($\tilde{R}^T$)	445.1	610.2	71.5	43.2	93.6	98.4	138.9	144.7
	MAD	634.4	588.9	28.1	35.0	39.8	36.9	47.4	67.6
45	Time-stability ($\tilde{\tau}$)	**37.1**	61.9	20.9	**20.7**	**23.4**	26.0	**23.1**	38.5
	MAD	37.8	38.1	10.7	17.4	6.1	8.4	4.9	17.1
	Total regret ($\tilde{R}^T$)	647.3	705.6	126.1	50.4	89.7	87.9	134.4	177.3
	MAD	831.9	769.7	109.4	28.1	35.2	32.2	43.8	96.6
50	Time-stability ($\tilde{\tau}$)	**25.9**	65.1	**23.7**	25.0	**21.7**	29.2	**22.3**	39.1
	MAD	29.6	40.0	15.5	25.5	6.9	6.7	6.8	18.2
	Total regret ($\tilde{R}^T$)	311.4	417.7	209.2	181.7	67.3	98.5	122.0	175.7
	MAD	412.2	382.1	242.3	249.9	30.0	36.9	44.3	100.6

Also, consistent with intuition and the theoretical results, GRN policies perform significantly better when the probability of learning more information increases and the quality of bounds improves.

One the main conclusions of our analysis is that policies that ignore the repeated interaction with the evader (and therefore act greedily in each period) and are optimistic to their own benefit, regarding the evader's costs, are efficient, provided that: (*i*) previous feedback is incorporated into the decision-making process in each period; (*ii*) results from optimization models are not followed blindly and are contrasted against the actual feedback obtained from the evader in the previous time periods. Surprisingly, this insight still holds in settings where the evader might act strategically. This latter feature, and (*ii*) above distinguish our work from extant literature.

While the GRN policies focus on minimizing the number of opportunities an (strategic) evader has to increase the regret, it is not clear that such policies are efficient in terms of minimizing said regret. In this regard, designing policies that are robust with respect to regret minimization is an interesting and promising direction for future research.

Our results show that implementing approximate policies is possible by solving a series of MIPs, whenever policy updates maintain the polyhedral structure of the uncertainty set. Furthermore, our results show that the strongest update does not necessarily maintain such structure. Thus, a promising direction for future research amounts to propose tight approximations to this strongest update that maintain such a polyhedral structure, and to study their practical performance. Alternatively, it might be possible to propose non-linear approximations to the strongest update that allows for implementation of the GNR-like policies by solving a series of structured (possibly non-linear) MIP problems.

Acknowledgements. This material is based on work partially supported by the National Science Foundation [Grant CMMI-1634835] and the Central Research Development Fund (CRDF) at the University of Pittsburgh. The research of the second author is partially supported by the Office of Naval Research (ONR) [Grant N00014-19-1-2329]. The fourth author's research was partially funded by the Complex Engineering Systems Institute (CONICYT PIA/BASAL AFB180003) and project FONDECYT 1181513. The authors thank three anonymous referees and the Associate Editor for constructive and helpful comments.

References

Ahuja R, Magnanti T, Orlin J (1993) *Network flows: Theory, algorithms, and applications* (Prentice-Hall).

Audet C, Hansen P, Jaumard B, Savard G (1997) Links between linear bilevel and mixed 0-1 programming problems. *Journal of Optimization Theory and Applications* 93(2):273–300.

Ball M, Golden B, Vohra R (1989) Finding the most vital arcs in a network. *Operations Research Letters* 8(2):73–76.

Bastert O, Matuszewski C (2001) Layered drawings of digraphs. *Drawing graphs*, 87–120 (Springer).

Bayrak H, Bailey M (2008) Shortest path network interdiction with asymmetric information. *Networks* 52(3):133–140.

Bertsimas D, Sim M (2003) Robust discrete optimization and network flows. *Mathematical programming* 98(1-3):49–71.

Borrero JS, Lozano L (2020) Modeling defender-attacker problems as robust linear programs with mixed-integer uncertainty sets. *INFORMS Journal on Computing.* Forthcoming.

Borrero JS, Prokopyev OA, Sauré D (2016) Sequential shortest path interdiction with incomplete information. *Decision Analysis* 13(1):68–98.

Borrero JS, Prokopyev OA, Sauré D (2019) Sequential interdiction with incomplete information and learning. *Operations Research* 67(1):72–89.

Buchheim C, Kurtz J (2018) Robust combinatorial optimization under convex and discrete cost uncertainty. *EURO Journal on Computational Optimization* 6(3):211–238.

Buehn A, Eichler S (2009) Smuggling illegal versus legal goods across the U.S.-mexico border: A structural equations model approach. *Southern Economic Journal* 76(2):328–350.

Cesa-Bianchi N, Lugosi G (2006) *Prediction, Learning, and Games* (Cambridge University Press).

Corley H, Sha D (1982) Most vital links and nodes in weighted networks. *Operations Research Letters* 1(4):157–160.

Erdös P, Rényi A (1959) On random graphs, I. *Publicationes Mathematicae (Debrecen)* 6:290–297.

Fulkerson D, Harding G (1977) Maximizing the minimum source-sink path subject to a budget constraint. *Mathematical Programming* 13(1):116–118.

Gathmann C (2008) Effects of enforcement on illegal markets: Evidence from migrant smuggling along the southwestern border. *Journal of Public Economics* 92(10-11):1926–1941.

Gift PD (2010) *Planning for an adaptive evader with application to drug interdiction operations.* Ph.D. thesis, Monterey, California. Naval Postgraduate School.

Hausken K, Zhuang J (2011) Governments' and terrorists' defense and attack in a t-period game. *Decision Analysis* 8(1):46–70.

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

Johnson MP, Gutfraind A, Ahmadizadeh K (2014) Evader interdiction: algorithms, complexity and collateral damage. *Annals of Operations Research* 222(1):341–359.

Magliocca NR, McSweeney K, Sesnie SE, Tellman E, Devine JA, Nielsen EA, Pearson Z, Wrathall DJ (2019) Modeling cocaine traffickers and counterdrug interdiction forces as a complex adaptive system. *Proceedings of the National Academy of Sciences* 116(16):7784–7792.

Malik K, Mittal A, Gupta S (1989) The k-most vital arcs in the shortest path problem. *Operations Research Letters* 8(4):223–227.

Morton D, Pan F, Saeger K (2007) Models for nuclear smuggling interdiction. *IIE Transactions* 39(1):3–14.

Petrov VV (2007) On lower bounds for tail probabilities. *Journal of Statistical Planning and Inference* 137:2703–2705.

Poss M (2013) Robust combinatorial optimization with variable budgeted uncertainty. *4OR* 11(1):75–92.

Salmerón J (2012) Deception tactics for network interdiction: A multiobjective approach. *Networks* 60(1):45–58.

Sefair JA, Smith JC (2016) Dynamic shortest-path interdiction. *Networks* 68(4):315–330.

Shaked M, Shanthikumar JG (2007) *Stochastic orders* (Springer Science & Business Media).

Sinha A, Malo P, Deb K (2018) A review on bilevel optimization: From classical to evolutionary approaches and applications. *IEEE Transactions on Evolutionary Computation* 22(2):276–295.

Smith JC, Prince M, Geunes J (2013) Modern network interdiction problems and algorithms. Pardalos P, Du DZ, Graham R, eds., *Handbook of Combinatorial Optimization*, 1949–1987 (Springer).

Smith JC, Song Y (2019) A survey of network interdiction models and algorithms. *European Journal of Operational Research* ISSN 0377-2217.

Song Y, Shen S (2016) Risk-averse shortest path interdiction. *INFORMS Journal on Computing* 28(3):527–539.

von Stengel B, Zamir S (2010) Leadership games with convex strategy sets. *Games and Economic Behavior* 69(2):446 – 457.

Watts DJ, Strogatz SH (1998) Collective dynamics of “small-world” networks. *Nature* 393(6684):440.

Xu J, Zhuang J (2016) Modeling costly learning and counter-learning in a defender-attacker game with private defender information. *Annals of Operations Research* 236(1):271–289.

Yu G, Yang J (1998) On the robust shortest path problem. *Computers & operations research* 25(6):457–468.

Yürekli A, Sayginsoy Ö (2010) Worldwide organized cigarette smuggling: an empirical analysis. *Applied Economics* 42(5):545–561.

Zare MH, Borrero JS, Zeng B, Prokopyev OA (2019) A note on linearized reformulations for a class of bilevel linear integer problems. *Annals of Operations Research* 272(1):99–117.

Zeng B, Zhao L (2013) Solving two-stage robust optimization problems using a column-and-constraint generation method. *Operations Research Letters* 41(5):457–461.

Zheng J, Castañón DA (2012) Dynamic network interdiction games with imperfect information and deception. *2012 IEEE 51st IEEE Conference on Decision and Control (CDC)*, 7758–7763 (IEEE).

Zhuang J, Bier VM, Alagoz O (2010) Modeling secrecy and deception in a multiple-period attacker–defender signaling game. *European Journal of Operational Research* 203(2):409–418.

Appendix A: Proofs for Theoretical Results

Proof of Theorem 1. The left-most inequality follows directly from the definitions, see equations (1) and (2). For the right-most inequality, let $\hat{I}^t \in \arg\max\{\hat{z}_R(I^t) : |I^t| \leq k, I^t \subseteq A\}$ and I^* be an optimal solution under full information, i.e., I^* is such that $z(I^*) = z^*$. Then, according to (6), $z_R^{t,*} = \hat{z}_R(\hat{I}^t) \geq \hat{z}_R(I^*)$. Using (5) we have that

$$\hat{z}_R(I^*) = \begin{cases} z_R(I^*) & \text{if } I^* \neq I^s, \quad \forall\, s < t \\ z^s & \text{for some } s < t\text{, otherwise.} \end{cases}$$

Next, we consider two cases: $I^* = I^s$ for some $s < t$; and $I^* \neq I^s$ for all $s < t$. In the first case, we have that $z_R^{t,*} \geq \hat{z}_R(I^*) = z(I^s) = z(I^*) = z^*$. For the second case, observe that $z_R^{t,*} \geq \hat{z}_R(I^*) = z_R(I^*)$. Because $c \in \mathcal{C}^t$, then $z(I^*) \leq z_R(I^*)$ from (4). Therefore, $z^* = z(I^*) \leq z_R(I^*)$. Accordingly, we have that $z_R^{t,*} \geq z^*$. ■

Proof of Proposition 1. First, we show that $\tau^\lambda \leq \xi^\lambda$. Suppose that $t = \xi^\lambda$, then from equation (7) we have that $z^{t,\lambda} = z_R^{t,*}$. Thus, by Theorem 1 we know that $z^{t,\lambda} = z^*$. We claim that $z^{s,\lambda} = z^*$ for $s > t$ (note that this would imply that $\tau^\lambda \leq t$). We prove this claim by contradiction.

Suppose that $z^{s,\lambda} < z^*$. By construction one has that $I^{s,\lambda} = I^{t,\lambda}$ for all $s \geq t$. Thus, the shortest path $P^{s,\lambda}$ must satisfy that $P^{s,\lambda} \in S(I^{t,\lambda})$. This in turn implies that $z(I^{t,\lambda}) \leq z(I^{s,\lambda})$ and thus, $z^{t,\lambda} \leq z^{s,\lambda} < z^*$, which contradicts the assumption that $z^{t,\lambda} = z^*$. Therefore, we conclude that $\tau^\lambda \leq t = \xi^\lambda$.

The second inequality follows directly from noting that policy λ does not repeat solutions unless an optimal interdiction solution is found, and that there are $\binom{m}{k}$ different interdiction decisions. Thus, a solution is repeated with certainty by period $\binom{m}{k}$+1. ■

Proof of Proposition 2. Consider graph G, depicted in Figure 6, that generalizes graph G_2 in Figure 3. Note that G is such that $m = 2(k+1)$. We consider the worst possible update mechanism consistent with Assumptions **A1-A2**, i.e. $\mathcal{C}^t = \mathcal{C}^0$ for all $t \in \mathcal{T}$. Also, without loss of generality we assume that k is odd.

In the first period, $I^{0,\lambda} = \emptyset$ and the evader uses path $P^0 = 1 \to 2 \to (k+3)$. Then the interdictor blocks $I^{1,\lambda} = \{(1,3), (1,4), \ldots, (1, k+2)\}$, because this solution is optimal for problem (6) with $z_R^{1,*} = 2k+3$. Consequently, the interdictor blocks different combinations of the arcs from paths $\{1 \to 3 \to (k+3), 1 \to 4 \to (k+3), \ldots, 1 \to (k+2) \to (k+3)\}$ and each of them returns the same objective function value as $z_R^{1,*}$. Note that there are 2^k possible solutions corresponding to blocking paths $\{1 \to 3 \to (k+3), 1 \to 4 \to (k+3), \ldots, 1 \to (k+2) \to (k+3)\}$. Every time after these solutions are implemented, the evader traverses through the same path $P = 1 \to 2 \to (k+3)$ whose total cost is 2.

For $t = 2^k + 1$, due to (5) and (6), repeating the previous solutions returns an objective function value of 2, which is no longer optimal. Therefore, the interdictor explores a new solution that makes path $1 \to 3 \to (k+3)$ available to the evader, which gives $z_R^{t,*} = 2k+2$. Observe that there are two ways to make path $1 \to 3 \to (k+3)$ available: either blocking paths $\{1 \to 2 \to (k+3), 1 \to 4 \to (k+3), \ldots, 1 \to (k+2) \to (k+3)\}$ or blocking paths $\{1 \to 4 \to (k+3), 1 \to 5 \to (k+3), \ldots, 1 \to (k+2) \to (k+3)\}$. There are 2^k and $\binom{k-1}{1} 2^{k-2}$ possible solutions corresponding to the above two path sets, respectively.

Proceeding in this fashion, in the next period, the interdictor makes path $1 \to 4 \to (k+3)$ available for the evader. Then paths $1 \to 5 \to (k+3), \ldots, 1 \to (k+2) \to (k+3)$ are available in the subsequent periods. Next,

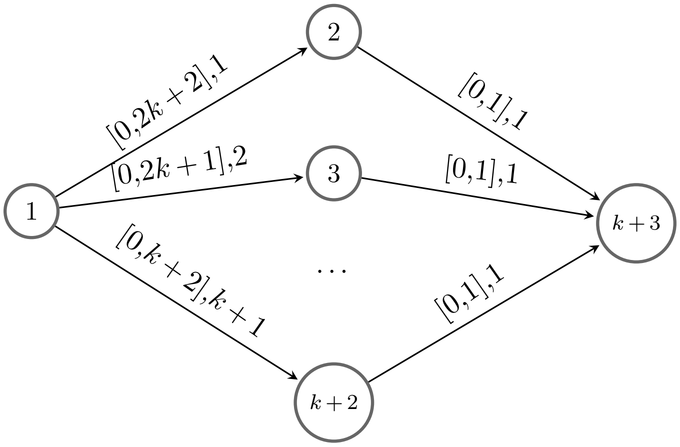


Figure 6 Network G used in the proof of Proposition 2. The labeling of the arcs is given by $[\ell_a, u_a], c_a$.

we assume that the interdictor blocks exactly k arcs in each period. (Otherwise, other feasible interdiction solutions may need to be explored which can only increase time-stability.) Thus, the process described above ends when the interdictor makes $\frac{k+1}{2}$ paths available to the evader because, by blocking exactly k arcs, she can make at most $\frac{k+1}{2}$ paths open for the evader in order to maximize $z_R^{t,*}$, see the graph structure in Figure 6. The reason why the interdictor attempts to proceed through all of these aforementioned steps is that all paths have their expected (by the interdictor) costs (based on the known upper bounds) strictly greater than their actual costs.

Implementing all these interdiction solutions results on a time-stability of at least

$$\sum_{i=1}^{\frac{k+1}{2}} \binom{k+1}{i} \binom{k+1-i}{i-1} 2^{k-2i+2} = \binom{2k+2}{k}.$$

At this point, the interdictor repeats a solution corresponding to the maximum $z^{t,\lambda}$ until time horizon T. Therefore, $\tau^\lambda \geq \binom{2k+2}{k} + 1$ and the result follows. ■

Proof of Proposition 3. First, we show that the shortest path used by the evader in each period before an optimal solution is found contains at least one arc whose cost is not known by the interdictor, i.e., $|P^t \setminus \cup_{s<t} P_v^s| \geq 1$ for $t < \tau^\lambda$. We prove this statement by contradiction.

Suppose that $t < \tau^\lambda$ and that the interdictor knows with certainty the cost of all the arcs in P^t. It follows that $\hat{c}_a = c_a$ for all $a \in P^t$. Therefore, we have that

$$\sum_{a \in P^t} \hat{c}_a = z^{t,\lambda}. \tag{25}$$

Note that because $P^t \subseteq A \setminus I^{t,\lambda}$, by (4) we have $\sum_{a \in P^t} \hat{c}_a \geq z_R(I^{t,\lambda})$. Moreover, from Corollary 1, we have that $z_R(I^{t,\lambda}) = \hat{z}_R(I^{t,\lambda})$ as $t < \tau^\lambda$ and there is no repetition until time t. Thus,

$$\sum_{a \in P^t} \hat{c}_a \geq z_R(I^{t,\lambda}) = \hat{z}_R(I^{t,\lambda}) = z_R^{t,*}.$$

According to Theorem 1, $z_R^{t,*} > z^{t,\lambda} = \sum_{a \in P^t} c_a$ for all $t < \tau^\lambda$. Thus, we have that $\sum_{a \in P^t} \hat{c}_a \geq z_R^{t,*} > z^{t,\lambda}$, which contradicts (25).

Note now that before τ^λ, because in each period there is at least one arc whose cost has not been observed, the probability of observing *at least* one new arc is lower bounded by $(p_r\,p_v)$. Because the number of arcs observed before τ^λ is at most m, we conclude that

$$Pr(\tau^\lambda \geq r) \leq Pr(m + X \geq r), \quad r \geq 0,$$

where X denotes the number of trials until m arcs' costs are learned. Note that X is a random variable with a negative binomial distribution of parameters $p_r\,p_v$ (probability of success, i.e., probability of learning an arc's cost) and m (number of successes). Therefore, $m+X$ is greater than τ^λ, in the first-order stochastic dominance sense. Given that the expected number of trials until learning m arcs' costs is $m(1-p_rp_v)/(p_rp_v)$, we conclude (see Shaked and Shanthikumar (2007)) that

$$\mathbb{E}[\tau^\lambda] \leq m + \mathbb{E}[X] = m + m\left(\frac{1-p_r\,p_v}{p_r\,p_v}\right) = \frac{m}{p_r\,p_v}.$$

This concludes the proof. ■

Proof of Corollary 2. Given that $E[\tau^\lambda] \leq m/(p_rp_v)$ from Proposition 3, there exists a constant $\alpha \in (0,1]$ such that $\mathbb{E}[\tau^\lambda] = \alpha\frac{m}{p_rp_v}$. By the Paley-Zygmund inequality (Petrov 2007), we have that

$$\Pr(\tau^\lambda > \gamma\mathbb{E}[\tau^\lambda]) \geq \frac{(1-\gamma)^2\mathbb{E}[\tau^\lambda]^2}{\mathbb{E}[(\tau^\lambda)^2]},$$

for any $0 \leq \gamma \leq 1$. It follows that,

$$\Pr\Big(\tau^\lambda > \gamma\alpha\frac{m}{p_rp_v}\Big) \geq \frac{(1-\gamma)^2\mathbb{E}[\tau^\lambda]^2}{\mathbb{E}[(\tau^\lambda)^2]}.$$

Now, from the proof of Proposition 3, we know that τ^λ is lower (in the first-order stochastic dominance sense) than a m plus a random variable X with negative binomial distribution with parameters $p_r\,p_v$ and m. This implies (see Shaked and Shanthikumar (2007)) that

$$\begin{aligned}
\mathbb{E}[(\tau^\lambda)^2] &\leq \mathbb{E}\Big[\Big(m+X\Big)^2\Big] \\
&= \mathbb{E}[m^2] + 2m\mathbb{E}[X] + \mathbb{E}[X^2] \\
&= m^2 + 2\,m^2\left(\frac{1-p_r\,p_v}{p_r\,p_v}\right) + \mathbb{E}[X^2] \\
&= m^2 + 2\,m^2\left(\frac{1-p_r\,p_v}{p_r\,p_v}\right) + \mathrm{Var}(X) + \mathbb{E}[X]^2 \\
&= m^2 + 2\,m^2\left(\frac{1-p_r\,p_v}{p_r\,p_v}\right) + m\left(\frac{1-p_r\,p_v}{(p_r\,p_v)^2}\right) + m^2\left(\frac{1-p_r\,p_v}{p_r\,p_v}\right)^2 \\
&= \frac{m^2 + m(1-p_r\,p_v)}{(p_r\,p_v)^2} \leq \frac{2m^2}{(p_r\,p_v)^2}.
\end{aligned}$$

The first inequality above satisfies from the results in Shaked and Shanthikumar (2007)) given that $m+X$ is greater than τ^λ in the first-order stochastic dominance sense, and that $\phi(x) = x^2$ is an increasing function with $x \geq 0$.

Because $\mathbb{E}[\tau^\lambda] = \alpha\frac{m}{p_rp_v}$, we have that

$$\begin{aligned}
\Pr\Big(\tau^\lambda > \gamma\alpha\frac{m}{p_rp_v}\Big) &\geq (1-\gamma)^2\Big(\frac{\alpha\;m}{p_rp_v}\Big)^2\frac{1}{\mathbb{E}[(\tau^\lambda)^2]} \\
&\geq (1-\gamma)^2\Big(\frac{\alpha\;m}{p_rp_v}\Big)^2\frac{(p_rp_v)^2}{2m^2} \quad = (1-\gamma)^2\alpha^2/2
\end{aligned}$$

as desired. ■

Proof of Proposition 4. The left-most inequality follows directly Theorem 1. For the right-most inequality, let $x^t \in \arg\max\{\hat{\underline{z}}_R(x) : \sum_{a\in A} x_a \le k,\ x \in \{0,1\}^m\}$ and x^* be an optimal solution under full information. That is, $z(x^*) = z^*$. Then, $\underline{z}_R^{t,*} = \hat{\underline{z}}_R(x^t) \ge \hat{\underline{z}}_R(x^*)$. Using (15), we have that

$$\hat{\underline{z}}_R(x^*) = \begin{cases} \underline{z}_R(x^*) & \text{if } x^* \ne x^s, \quad \forall\, s < t \\ z^s & \text{for some } s < t, \text{ otherwise.} \end{cases}$$

Next, we consider two cases: $x^* = x^s$ for some $s < t$; and $x^* \ne x^s$ for all $s < t$. In the first case, following from the proof of Theorem 1, we have $\underline{z}_R^{t,*} \ge \hat{\underline{z}}_R(x^*) = z(x^s) = z(x^*) = z^*$. For the second case, since $x^t \ne x^s$ for all $s < t$, we have $\underline{z}_R^{t,*} \ge \hat{\underline{z}}_R(x^*) = \underline{z}_R(x^*)$.

Note that real cost vector c is never cut from our uncertainty set update $\mathcal{C}^{t+1} = \mathcal{C}^t \cap \mathcal{R}_w^t \cap \mathcal{V}^t$, that is,

$$c \in \mathcal{C}^t,\ \forall t \in \mathcal{T}.$$

Therefore, for any x^t and y, we have

$$\max_{\hat{c}}\{(\hat{c} + Mx^t)^T y :\ \hat{c} \in \mathcal{C}^t\} \ge (c + Mx^t)^T y.$$

Now recall that

$$\underline{z}_R(x^t) = \min_{y}\left\{\max_{\hat{c}}\{(\hat{c} + Mx^t)^T y :\ \hat{c} \in \mathcal{C}^t\} :\ By = b,\ y \ge 0\right\},$$

and after the interdiction solution x^t in each time period $t \in \mathcal{T}$, the evader solves for $z(x^t)$:

$$z(x^t) = \min\{(c + Mx^t)^T y :\ By = b,\ y \ge 0\}.$$

Suppose $y_R \in \arg\min\{\max\{(\hat{c} + Mx^t)^T y :\ \hat{c} \in \mathcal{C}^t\} :\ By = b,\ y \ge 0\}$ and $y^* \in \arg\min\{(c + Mx^t)^T y :\ By = b,\ y \ge 0\}$. Then we have

$$\underline{z}_R(x^t) = \max\{(\hat{c} + Mx^t)^T y_R :\ \hat{c} \in \mathcal{C}^t\} \ge (c + Mx^t)^T y_R \ge (c + Mx^t)^T y^* = z(x^t).$$

Thus, $\underline{z}_R(x^*) \ge z(x^*)$. Hence, $\underline{z}_R^{t,*} \ge \hat{\underline{z}}_R(x^*) = \underline{z}_R(x^*) \ge z(x^*) = z^*$. ■

Proof of Lemma 1. Let $\tilde{I}^t \in \arg\max\{\tilde{z}_R(I^t) : |I^t| \le k, I^t \subseteq A\}$, where $\tilde{z}_R(I^t)$ is tailored for the general evader as defined in equation (22). Also, let I^* be an optimal solution under full information. That is, $z(I^*) = z^*$. Then, $\tilde{z}_R^{t,*} = \tilde{z}_R(\tilde{I}^t) \ge \tilde{z}_R(I^*)$. Using (22) we have that

$$\tilde{z}_R(I^*) = \begin{cases} z_R(I^*) & \text{if } I^* \ne I^s, \quad \forall\, s < t \\ \min\{z^s :\ I^* = I^s,\ s < t\} & \text{otherwise.} \end{cases}$$

Next, we consider two cases: $I^* = I^s$ for some $s < t$; and $I^* \ne I^s$ for all $s < t$. In the first case, we have that $\tilde{z}_R^{t,*} \ge \tilde{z}_R(I^*) = \min\{z^s :\ I^* = I^s,\ s < t\}$. Note that the general evader can be suboptimal, thus the cost observed by the interdictor is at least z^* for all $s \in \{s' :\ I^* = I^{s'}\}$, that is $z^s \ge z^*$. Therefore $z_R^{t,*} \ge \tilde{z}_R(I^*) = \min\{z^s :\ I^* = I^s,\ s < t\} \ge z^*$. The proof for the second case directly follows the proof of Theorem 1. ■

Proof of Proposition 5. We define S_{sub} as the set of time periods where the cost observed by the interdictor is less than the optimal cost, that is, $S_{\text{sub}} = \{t \ge 0 :\ z^{t,\lambda} < z^*\}$. Then we claim that for any two distinct time periods $s_1, s_2 \in S_{\text{sub}}$, we have that $I^{s_1,\lambda} \neq I^{s_2,\lambda}$. In other words, under policy λ, the solutions $I^{s,\lambda}$ are all distinct for $t \in S_{\text{sub}}$. We prove the claim by contradiction. Suppose that the interdictor repeats $I^{s_1,\lambda}$ at time s_2, that is, $I^{s_2,\lambda} = I^{s_1,\lambda}$. By Lemma 1, we have that $z^* \leq \tilde{z}_R^{t,*}$. Therefore, we have that $\tilde{z}_R^{s_2,*} = \min\{z^{s',\lambda}:\ s' \leq s_1,\ I^{s',\lambda} = I^{s_1,\lambda}\}$ according to (22). Putting this together, get the following contradiction:

$$z^{s_2,\lambda} < z^* \leq \tilde{z}_R^{s_2,*} = \min\{z^{s',\lambda}:\ s' \leq s_1,\ I^{s',\lambda} = I^{s_1,\lambda}\} \leq z^{s_1,\lambda} < z^* \leq \tilde{z}_R^{s_1,*}.$$

Because for all $t \in S_{\text{sub}}$, $z^{t,\lambda} < z^*$, solutions $I^{t,\lambda}$ are all suboptimal. Notice that the total number of solutions is $\binom{m}{k}$. Then given that $I^{t,\lambda}$ are distinct for all $t \in S_{\text{sub}}$, we have the following results:

$$|S_{\text{sub}}| = |\{t \geq 0:\ z^{t,\lambda} < z^*\}| \leq \binom{m}{k} + 1,$$

which implies the required result. ∎

Appendix B: Decision-making process for network G_2 in Figure 3

- **Step 0**: $I^{0,\lambda} = \emptyset$, $z^0 = 2$;
- **Step 1**: $I^{1,\lambda} = \{(1,3),(1,4)\}$, and the interdictor would expect the evader to traverse path $1-2-5$. This implies that $z_R^{1,*} = 7$ and that the evader would go through path $1-2-5$, implying that $z^{1,\lambda} = 2$.
- **Step 2**: $I^{2,\lambda} = \{(1,3),(4,5)\}$, $z_R^{2,*} = 7$, $z^{2,\lambda} = 2$;
- **Step 3**: $I^{3,\lambda} = \{(1,4),(3,5)\}$, $z_R^{3,*} = 7$, $z^{3,\lambda} = 2$;
- **Step 4**: $I^{4,\lambda} = \{(3,5),(4,5)\}$, $z_R^{4,*} = 7$, $z^{4,\lambda} = 2$;
- **Step 5**: $I^{5,\lambda} = \{(1,2),(1,4)\}$, $z_R^{5,*} = 6$, $z^{5,\lambda} = 3$;
- **Step 6**: $I^{6,\lambda} = \{(1,2),(4,5)\}$, $z_R^{6,*} = 6$, $z^{6,\lambda} = 3$;
- **Step 7**: $I^{7,\lambda} = \{(1,4),(2,5)\}$, $z_R^{7,*} = 6$, $z^{7,\lambda} = 3$;
- **Step 8**: $I^{8,\lambda} = \{(2,5),(4,5)\}$, $z_R^{8,*} = 6$, $z^{8,\lambda} = 3$;
- **Step 9**: $I^{9,\lambda} = \{(1,4),(4,5)\}$, $z_R^{9,*} = 6$, $z^{9,\lambda} = 2$;
- **Step 10**: $I^{10,\lambda} = \{(1,2),(1,3)\}$, $z_R^{10,*} = 5$, $z^{10,\lambda} = 4$;
- **Step 11**: $I^{11,\lambda} = \{(1,2),(3,5)\}$, $z_R^{11,*} = 5$, $z^{11,\lambda} = 4$;
- **Step 12**: $I^{12,\lambda} = \{(1,3),(2,5)\}$, $z_R^{12,*} = 5$, $z^{12,\lambda} = 4$;
- **Step 13**: $I^{13,\lambda} = \{(2,5),(3,5)\}$, $z_R^{13,*} = 5$, $z^{13,\lambda} = 4$;
- **Step 14**: $I^{14,\lambda} = \{(1,2),(2,5)\}$, $z_R^{14,*} = 5$, $z^{14,\lambda} = 3$;
- **Step 15**: $I^{15,\lambda} = \{(1,3),(3,5)\}$, $z_R^{15,*} = 5$, $z^{15,\lambda} = 2$;
- **Step 16**: $I^{16,\lambda} = \{(1,2),(2,3)\}$, $z_R^{16,*} = 4$, $z^{16,\lambda} = 4$.

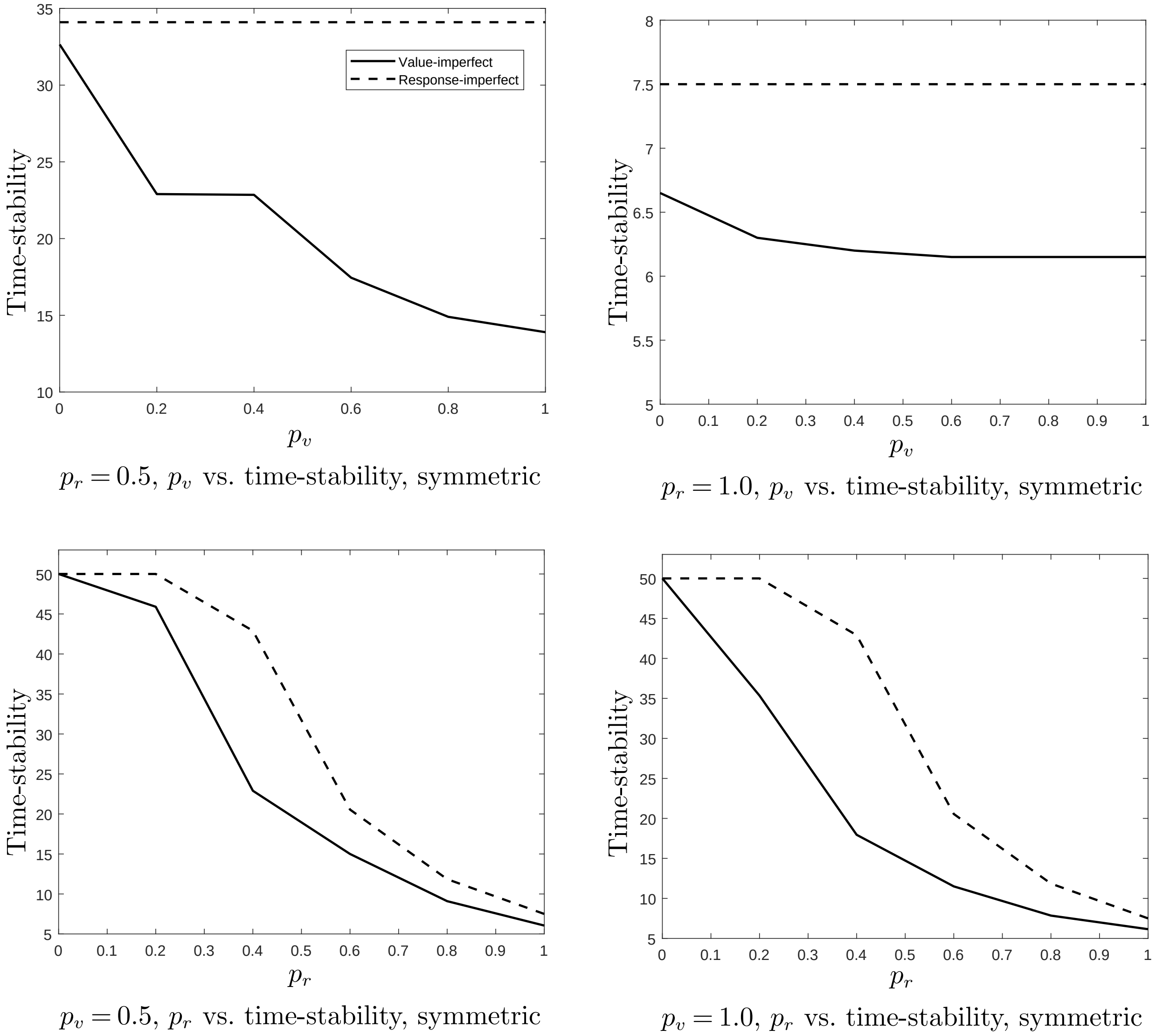
We can see that after 16 steps, the interdictor finally identifies an optimal solution for the full information problem.

Appendix C: Supplementary Computational Results for Uniform Random Graphs

C.1. Policy Performance: Sensitivity with Respect to p_r and p_v

In Section 7.4, we test the performance of the approximate GRN policies on uniform random graphs with right-skewed costs. For left-skewed and symmetric cost structures, we use the same graph size as in Section 7.4, i.e. $n = 50$ and probability of having an arc between any two nodes is 0.5. We set $T = 50$.

Figure 7 **Average time-stability for policies in $\underline{\Lambda}$ for different types of feedback as p_r and p_v increase for $k=6$ and the symmetric costs ($T=50$, uniform random graphs, greedy evader).**



$p_r = 0.5$, p_v vs. time-stability, symmetric

$p_r = 1.0$, p_v vs. time-stability, symmetric

$p_v = 0.5$, p_r vs. time-stability, symmetric

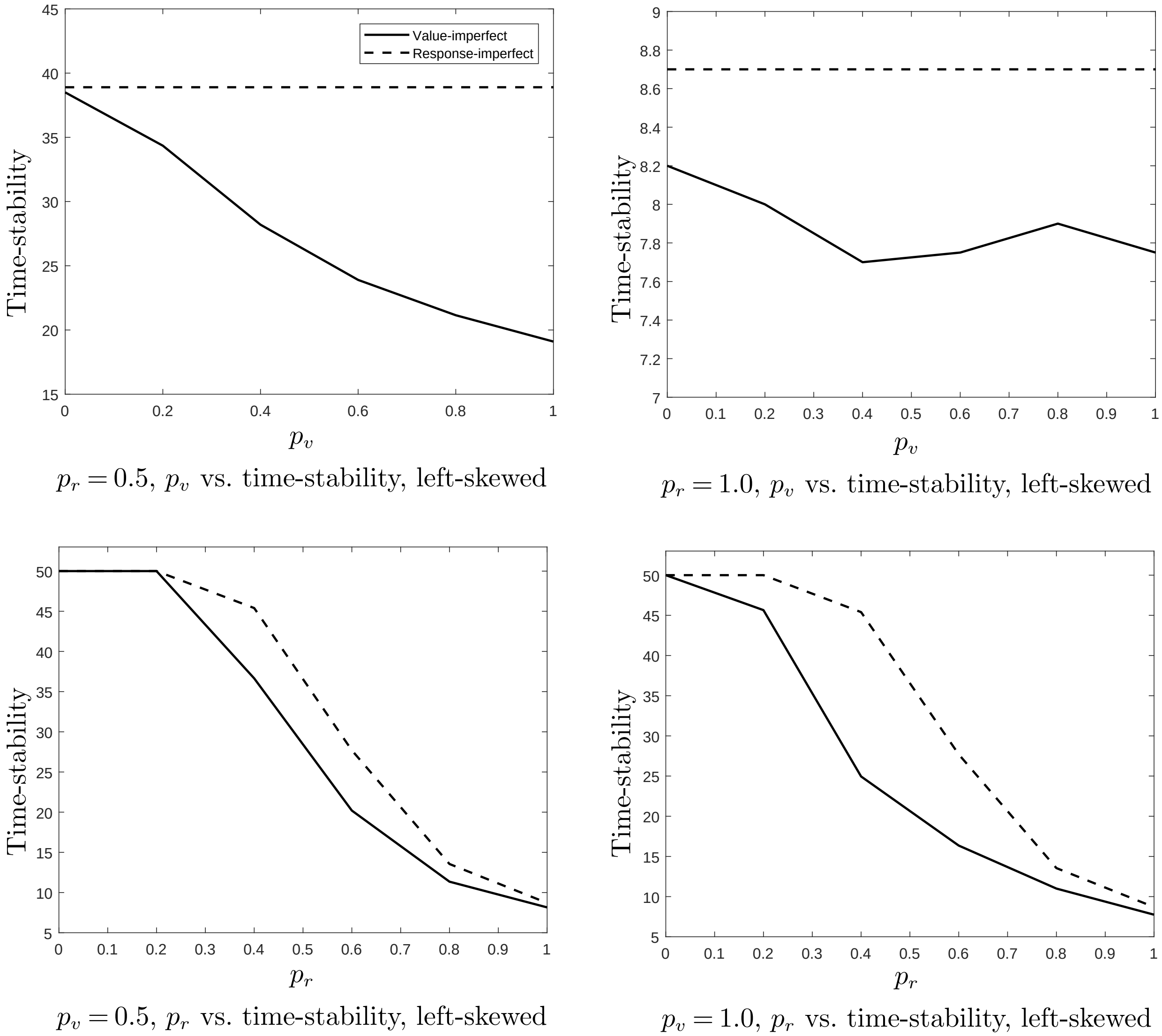
$p_v = 1.0$, p_r vs. time-stability, symmetric

The results for symmetric and left-skewed costs are shown in Figure 7 and 8, respectively. We observe that for both response-imperfect and value-imperfect feedback, GRN policies obtain optimal solutions in less time steps as p_r or p_v increases. Moreover, under value-imperfect feedback, the policies converge faster than under response-imperfect feedback.

C.2. Performance under a General Evader

In Section 7.6, we test policy performance when we relax the assumption on the greedy nature of the evader and consider the setting introduced in Section 6, where the evader's response is constrained to a $1-n$ (not necessarily shortest) path on the interdicted graph. Policy performance under Random–0.9 and Random–0.5 in depicted in Tables 6 and 7, respectively. The definitions of Random–0.9 and Random–0.5

Figure 8 **Average time-stability for policies in $\underline{\Lambda}$ for different types of feedback as p_r and p_v increase for $k = 6$ and the left-skewed costs ($T = 50$, uniform random graphs, greedy evader).**



$p_r = 0.5$, p_v vs. time-stability, left-skewed

$p_r = 1.0$, p_v vs. time-stability, left-skewed

$p_v = 0.5$, p_r vs. time-stability, left-skewed

$p_v = 1.0$, p_r vs. time-stability, left-skewed

evader can be found in Section 7.6. Observe that with *less randomness*, for example under Random–0.9 evader, there are more cases where the approximate GRN policies outperform the mean-bound policies.

Appendix D: Computational Results for Layered Graphs

D.1. Graph generation

We generate layered graphs using parameters (θ, ϕ), where ϕ and θ denote the number of layers and nodes in each layer, respectively. We add a source node before the first layer and a destination node after the last layer. Thus, the total number of nodes is $\theta \times \phi + 2$. There is an arc from source node to all the nodes in the first layer and from all the nodes in the last layer to the destination node. Moreover, there is an arc with probability 0.5 from every node in layer i to nodes in layers $i+1, i+2, \ldots, \phi$.

Table 6 Average time-stability and MAD (in parenthesis) for $\lambda_v^G \in \underline{\Lambda}_v^G$ and $\pi_M^G \in \Pi_M^G$ policies (with value-imperfect feedback) when evader type is Random–0.9 on uniform random graphs with $k=6$ and $T=100$. Note that a Random-0.9 evader chooses the shortest path with probability 0.9 and the second shortest path with probability 0.1.

n		Right-skewed		Symmetric		Left-skewed		Random	
		λ_v^G	$\pi_M^G(+\mathcal{V})$	λ_v^G	$\pi_M^G(+\mathcal{V})$	λ_v^G	$\pi_M^G(+\mathcal{V})$	λ_v^G	$\pi_M^G(+\mathcal{V})$
15	Time-stability ($\tilde{\tau}$)	**20.2**	70.9	**13.7**	23.5	20.0	**19.0**	**20.6**	59.7
	MAD	23.9	35.0	5.5	31.7	6.9	8.0	5.2	41.6
	Total regret ($\tilde{R}^T$)	580.8	806.8	106.6	62.5	166.5	140.3	221.0	310.6
	MAD	895.8	850.1	35.1	29.2	47.3	45.7	61.5	154.7
20	Time-stability ($\tilde{\tau}$)	**32.9**	71.3	**19.8**	21.6	**18.9**	24.4	**19.5**	34.7
	MAD	40.1	35.2	7.7	28.7	7.1	8.0	6.5	30.5
	Total regret ($\tilde{R}^T$)	1065.5	1189.9	132.4	54.8	118.8	130.5	164.8	181.1
	MAD	1438.0	1334.1	50.5	36.6	39.0	43.4	67.6	117.1
25	Time-stability ($\tilde{\tau}$)	**15.3**	70.6	15.2	**12.9**	25.3	**22.3**	**21.8**	60.8
	MAD	17.0	34.8	4.9	16.4	7.1	7.7	4.4	38.8
	Total regret ($\tilde{R}^T$)	315.4	549.0	96.3	42.0	149.3	115.2	168.7	314.5
	MAD	491.4	529.7	28.4	26.1	39.0	35.9	53.9	210.8
30	Time-stability ($\tilde{\tau}$)	**34.8**	77.0	**19.8**	35.9	**24.9**	28.0	**20.3**	38.8
	MAD	39.2	30.4	9.7	41.3	5.8	12.4	4.4	25.7
	Total regret ($\tilde{R}^T$)	965.0	1091.0	122.3	253.4	118.9	118.6	144.8	178.2
	MAD	1290.7	1190.3	68.3	332.5	36.7	49.0	39.1	76.4
35	Time-stability ($\tilde{\tau}$)	**24.6**	78.2	**13.9**	21.2	**20.1**	28.0	**28.7**	57.4
	MAD	30.2	30.8	5.8	29.1	5.4	7.7	10.5	37.3
	Total regret ($\tilde{R}^T$)	676.4	824.5	87.4	44.9	103.7	140.8	225.7	216.4
	MAD	1023.0	965.9	31.6	34.1	47.2	76.2	128.0	112.6
40	Time-stability ($\tilde{\tau}$)	**24.6**	86.5	**15.3**	22.3	28.6	**28.4**	**22.8**	42.9
	MAD	30.2	17.6	4.0	29.0	8.8	6.3	7.8	31.9
	Total regret ($\tilde{R}^T$)	473.8	754.8	80.7	53.8	118.9	101.8	137.8	165.0
	MAD	699.4	631.7	20.5	50.3	46.3	27.9	45.6	77.6
45	Time-stability ($\tilde{\tau}$)	**36.1**	86.6	**15.5**	35.8	**23.0**	30.5	**25.3**	46.8
	MAD	38.4	17.0	5.2	40.8	7.4	11.0	6.9	30.6
	Total regret ($\tilde{R}^T$)	739.2	942.7	74.5	69.4	91.1	104.9	163.1	206.0
	MAD	979.6	873.2	24.0	60.9	31.9	43.3	62.1	125.1
50	Time-stability ($\tilde{\tau}$)	**29.4**	77.7	**24.3**	34.8	**24.6**	28.5	**22.4**	53.1
	MAD	34.3	30.7	15.5	40.9	7.4	9.4	5.6	32.9
	Total regret ($\tilde{R}^T$)	506.4	636.6	222.4	197.1	74.9	96.5	122.0	179.4
	MAD	708.0	595.1	263.5	265.9	25.5	39.2	41.0	68.9

D.2. Comparison of Policies without Information Updates

We compare policy performance when no information updates are used, i.e., $\mathcal{C}^t = \mathcal{C}^0$ for all periods t. We test policy performance using the right-skewed, symmetric, left-skewed and random cost structures on layered graphs. For each structure we randomly generate 20 instances with $(\theta, \phi) = (7, 3)$. Finally, we

Table 7 Average time-stability and MAD (in parenthesis) for $\lambda_v^G \in \underline{\Lambda}_v^G$ and $\pi_M^G \in \Pi_M^G$ policies (with value-imperfect feedback) when evader type is Random–0.5 on uniform random graphs with $k=6$ and $T=100$. Note that a Random–0.5 evader randomly chooses from the shortest path and the second shortest path.

n		Right-skewed		Symmetric		Left-skewed		Random	
		λ_v^G	$\pi_M^G(+\mathcal{V})$	λ_v^G	$\pi_M^G(+\mathcal{V})$	λ_v^G	$\pi_M^G(+\mathcal{V})$	λ_v^G	$\pi_M^G(+\mathcal{V})$
15	Time-stability ($\tilde{\tau}$)	**20.9**	57.0	**13.5**	14.5	20.1	19.1	**17.7**	45.3
	MAD	23.7	30.9	6.3	19.1	5.6	5.8	5.2	33.2
	Total regret ($\tilde{R}^T$)	570.5	738.6	99.1	49.0	163.4	134.4	177.1	241.2
	MAD	855.2	820.1	36.2	18.1	56.9	42.4	66.7	120.2
20	Time-stability ($\tilde{\tau}$)	**30.7**	58.5	17.1	**16.6**	**20.9**	26.6	**19.0**	28.2
	MAD	34.7	29.1	6.8	17.5	5.6	7.9	5.4	22.2
	Total regret ($\tilde{R}^T$)	912.0	1022.5	113.0	55.1	124.0	135.0	153.3	163.1
	MAD	1285.0	1223.6	42.9	31.3	36.7	44.1	61.8	97.3
25	Time-stability ($\tilde{\tau}$)	**16.3**	47.6	17.0	**8.0**	**23.8**	26.9	**23.4**	42.9
	MAD	16.8	24.8	5.5	8.6	6.1	8.2	5.6	27.9
	Total regret ($\tilde{R}^T$)	327.6	421.3	109.9	36.8	132.7	146.6	165.1	268.4
	MAD	474.2	468.4	42.5	19.8	34.2	47.5	44.3	184.2
30	Time-stability ($\tilde{\tau}$)	**35.2**	68.8	**24.9**	28.8	**25.4**	30.7	**20.6**	36.2
	MAD	38.9	34.4	16.6	32.1	6.4	7.6	4.8	21.4
	Total regret ($\tilde{R}^T$)	844.7	932.0	232.5	214.1	109.9	118.3	136.8	165.3
	MAD	1096.3	1056.8	261.8	274.0	31.6	35.2	49.7	53.3
35	Time-stability ($\tilde{\tau}$)	**25.7**	67.9	16.0	**15.3**	**24.9**	31.0	**22.2**	45.5
	MAD	29.7	35.3	5.0	20.0	7.0	9.4	5.6	23.2
	Total regret ($\tilde{R}^T$)	594.5	703.0	86.2	44.9	122.2	138.4	158.3	195.6
	MAD	871.4	829.1	28.7	36.8	56.4	52.7	65.6	80.0
40	Time-stability ($\tilde{\tau}$)	**25.1**	68.2	**15.3**	20.0	**24.1**	27.2	**21.7**	45.9
	MAD	30.0	31.8	4.3	23.3	5.4	10.1	7.0	25.7
	Total regret ($\tilde{R}^T$)	435.2	602.3	68.8	51.4	98.2	108.1	127.9	172.0
	MAD	632.7	574.8	17.7	42.5	34.2	46.4	48.2	80.0
45	Time-stability ($\tilde{\tau}$)	**37.3**	71.3	**17.5**	25.6	**25.3**	33.0	**24.7**	39.5
	MAD	37.7	31.6	5.2	25.6	8.5	9.9	5.7	21.4
	Total regret ($\tilde{R}^T$)	643.5	768.0	73.5	57.2	104.2	118.5	135.9	180.1
	MAD	832.5	771.3	18.1	38.2	47.4	45.3	54.1	98.2
50	Time-stability ($\tilde{\tau}$)	**26.2**	68.1	29.9	**28.9**	**23.3**	28.2	**24.6**	37.5
	MAD	29.5	35.1	14.3	31.4	7.7	6.1	7.6	20.1
	Total regret ($\tilde{R}^T$)	372.2	482.7	239.5	192.3	77.6	89.2	131.8	159.3
	MAD	518.8	459.0	240.2	258.5	31.5	27.3	49.0	76.3

use $k \in \{2,4\}$, and set either $T=500$ or $T=1000$, respectively. Table 8 summarizes the results, which are similar to those for uniform random graphs, see Table 2.

D.3. Improvements When Information Updates Are Applied

We consider layered graphs with 7 nodes ($\theta = 7$) in each layer and number of layers $\phi \in \{3, 4, \ldots, 10\}$ for all four cost structures. We set $T = 500$ and for each cost structure we randomly generate 20 instances. For response-imperfect feedback we set $p_r = 0.5$, and for value-imperfect feedback we set $p_r = p_v = 0.5$. Note that in this section and the following two sections, since we have polyhedral cost update mechanism, the approximate GRN policies are implemented. We measure policy performance using the average time-stability and MAD, see Table 9. As in Section 7.3, we observe that policy performance improves as more information revealed in the feedback. However, note that there is no significant improvement for policies in Π, even if the interdictor learns more information. On the other hand, GRN policies outperform mean bound policies under response- and value-imperfect feedback.

D.4. Policy Performance: Sensitivity with Respect to p_r and p_v

We set $(\theta, \phi) = (7, 10)$ for all the experiments, and generate 20 instances with different cost structures (right-skewed, symmetric and left-skewed). Interdictor's resource limit and time horizon are set as $k = 6$ and $\mathcal{T} = \{0, 1, \ldots, 50\}$. Figure 9, 10 and 11 depict the behaviour of time-stability for response- and value-imperfect updates with the right-skewed, symmetric and left-skewed costs, respectively. As in the case of uniform random graphs in Section 7.4, performance of policies $\underline{\Lambda}_r$ and $\underline{\Lambda}_v$ improves as p_r and p_v increase. We observe that time-stability of policies Λ_v decreases faster than that of Λ_r.

D.5. Policy Performance: Sensitivity with Respect to Quality of Arc Cost Bounds

We generate 20 instances for I.1, I.2 and I.3 as in Section 7.5, and set $k = 6$. Test instances are layered graphs with $(\theta, \phi) = (7, 10)$ and we set $T = 200$. We test policy performance for various values of probabilities p_r and p_v. Table 10 summarizes the results obtained. There, we observe similar results to those presented in Section 7.5, which indicate that the quality of the initial bounds has a significant affect on policy performance.

Appendix E: Computational Results for Watts-Strogatz Graphs

E.1. Graph generation

We generate Watts-Strogatz graphs following the model in Watts and Strogatz (1998), using parameters (n, d, β). Mean degree of nodes and rewiring probability are denoted as d and β, respectively. Note that β denotes the graph instances' degree of randomness. Through all the experiments, we set $\beta = 1$.

E.2. Comparison of Policies without Information Updates

For each structure we randomly generate 20 instances with $(n, d) = (15, 8)$. Finally, we have $k \in \{2, 4\}$, and use $T \in \{500, 1000\}$. Table 11 summarizes the results, which are similar to those for uniform random and layered graphs, see Table 2 and Table 8, respectively.

E.3. Improvements When Information Updates Are Applied

We consider Watts-Strogatz graphs with

$$(n, d) \in \{(15, 8), (20, 15), (25, 14), (30, 16), (35, 22), (40, 22), (45, 28), (50, 30)\},$$

for all four cost structures. We set $T = 500$ and for each cost structure we randomly generate 20 instances. For response-imperfect feedback we set $p_r = 0.5$, and in value-imperfect feedback we set $p_r = p_v = 0.5$. Note that similar in Appendix D, in this section and the following two sections, since we have polyhedral cost update mechanism, the approximate GRN policies are implemented. Table 12 depicts our results, which are similar to those obtained for uniform random and layered graphs.

E.4. Policy Performance: Sensitivity with Respect to p_r and p_v

We set $(n,d)=(50,30)$ for all the experiments, and generate 20 instances with different cost structures (right-skewed, symmetric and left-skewed). Interdictor's resource limit and time horizon are set as $k=6$ and $\mathcal{T}=\{0,1,\ldots,50\}$. Figure 12, 13 and 14 depict the behaviour of time-stability for response- and value-imperfect updates with the right-skewed, symmetric and left-skewed costs, respectively.

E.5. Policy Performance: Sensitivity with Respect to Quality of Arc Cost Bounds

We generate 20 instances for I.1, I.2 and I.3 as in Section 7.5, and set $k=6$. We set $(n,d)=(50,30)$ and $T=200$. We test policy performance for various values of probabilities p_r and p_v. Table 13 summarizes the results obtained, which are similar results to those in Section 7.5 and Appendix D.5.

Table 8 **Performance of policies in Λ and benchmark policies without information updates (7×3, layered graphs, greedy evader).**

Graph		$k=2$ $(T=500)$				$k=4$ $(T=1000)$			
		λ	π_L	π_M	π_R	λ	π_L	π_M	π_R
Right-skewed	Time-stability	**82.0**^{1}	450.2^{18}	201.2^{8}	450.2^{18}	**445.3**^{7}	900.2^{18}	600.8^{12}	950.2^{18}
	MAD	116.1	89.6	239.0	89.6	420.3	179.6	479.0	94.7
	Relative difference	**0.2%**	18.5%	3.7%	18.2%	**2.9%**	19.9%	5.6%	17.0%
	MAD	0.3%	11.9%	4.4%	12.6%	2.6%	9.1%	5.6%	9.5%
	Total regret	**242.3**	4925.0	1025.0	5082.5	**2002.4**	12700.0	3650.0	11103.4
	MAD	360.4	3167.5	1237.5	3575.8	1973.0	5140.0	3680.0	6306.9
Symmetric	Time-stability	381.6^{12}	350.6^{14}	**133.4**^{4}	378.5^{15}	946.4^{17}	900.2^{18}	**153.2**^{3}	691.2^{14}
	MAD	142.1	209.2	167.2	182.3	91.1	179.6	246.5	401.5
	Relative difference	10.6%	11.9%	**0.5%**	10.6%	17.4%	15.3%	**0.3%**	10.5%
	MAD	9.5%	12.2%	0.8%	10.8%	8.4%	11.3%	0.6%	9.5%
	Total regret	2343.3	2450.0	**146.8**	2259.4	10055.2	7750.0	**382.7**	5945.2
	MAD	1516.7	2580.0	185.5	2249.0	3969.8	6025.0	641.3	4763.7
Left-skewed	Time-stability	411.9^{15}	**151.4**^{6}	337.5^{9}	324.4^{9}	978.6^{19}	**351.3**^{7}	867.6^{15}	876.0^{16}
	MAD	132.2	209.2	146.3	165.7	40.8	454.1	206.9	198.5
	Relative difference	27.4%	**3.5%**	14.1%	9.1%	34.4%	**3.2%**	19.4%	22.2%
	MAD	19.2%	4.9%	15.8%	10.0%	13.5%	4.3%	15.7%	18.8%
	Total regret	4085.8	**425.0**	2247.8	2334.9	10635.1	**1050.0**	7860.9	8042.5
	MAD	2592.7	595.0	1377.0	1811.9	3858.9	1370.0	3827.3	5773.4
Random	Time-stability	425.2^{15}	**226.1**^{9}	290.6^{9}	350.9^{14}	974.1^{19}	850.3^{17}	**781.2**^{14}	950.5^{18}
	MAD	115.3	246.5	189.9	208.7	49.3	254.5	306.3	94.1
	Relative difference	29.5%	9.0%	**8.1%**	14.8%	32.0%	22.3%	**18.6%**	25.9%
	MAD	18.9%	10.4%	9.2%	12.6%	15.4%	13.6%	14.3%	17.5%
	Total regret	4912.9	1850.0	**1719.5**	2895.8	16014.8	11600.0	**9979.1**	14468.1
	MAD	2889.4	2170.0	1587.4	2365.4	7935.9	7500.0	6205.8	9943.2

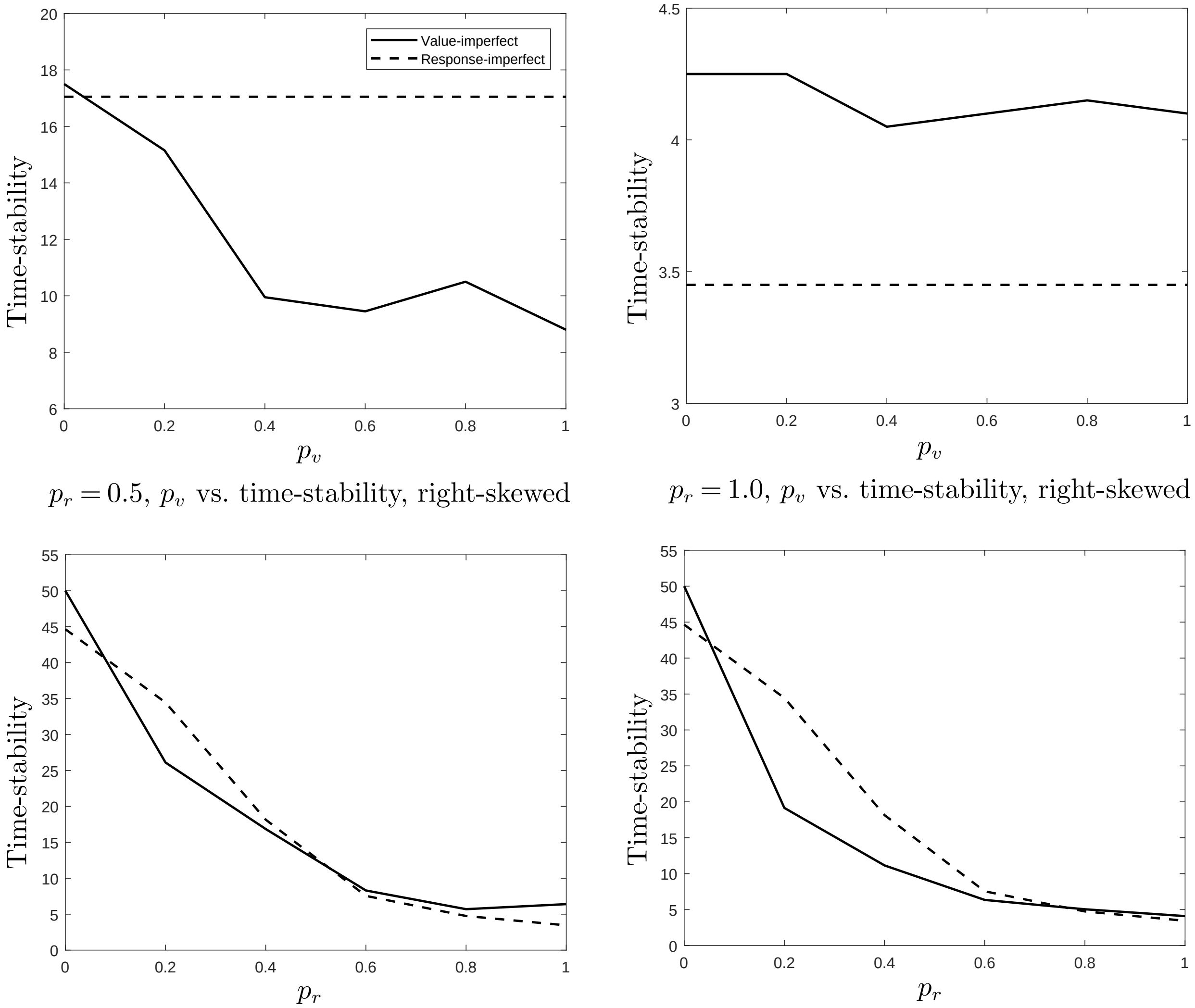
Notes: Entries in bold denote the best policy in each setting; the numbers in superscript of time-stability denote the number of instances out of 20 for which the corresponding policy failed to converge within T time periods.

Table 9 Average time-stability and MAD (in parenthesis) for $\lambda \in \underline{\Lambda}$ and $\pi_M \in \Pi_M$ policies when information updates are used ($k = 6$, $T = 500$, layered graphs, greedy evader).

Size	Right-skewed					Symmetric					Left-skewed					Random				
	λ	λ_r	λ_v	π_M	$\pi_M(+)$	λ	λ_r	λ_v	π_M	$\pi_M(+)$	λ	λ_r	λ_v	π_M	$\pi_M(+)$	λ	λ_r	λ_v	π_M	$\pi_M(+)$
7×3	341.6^{11} (195.8)	12.7 (5.6)	**7.3** (2.4)	325.7^{13} (226.6)	325.7^{13} (226.6)	500^{20} (0.0)	31.3 (12.0)	**19.2** (6.2)	226.9^{7} (232.7)	128.5^{5} (185.8)	500^{20} (0.0)	43.8 (19.9)	**24.7** (6.7)	489.4^{19} (20.2)	27.0 (6.0)	500^{20} (0.0)	37.5 (14.9)	**23.0** (5.3)	330.1^{17} (208.6)	329.16^{13} (222.2)
7×4	383.2^{15} (175.2)	11.4 (5.6)	**8.5** (3.3)	375.5^{15} (186.8)	375.5^{15} (186.8)	500^{20} (0.0)	33.1 (12.0)	**16.7** (6.0)	155.0^{6} (207.0)	176.6^{7} (226.4)	500^{20} (0.0)	39.1 (17.1)	33.6 (15.7)	500^{20} (0.0)	31.6 (7.7)	500^{20} (0.0)	40.9 (13.8)	**22.6** (4.9)	500^{20} (0.0)	282.3^{11} (239.5)
7×5	383.6^{15} (174.7)	15.4 (8.5)	**8.4** (3.4)	300.8^{12} (239.04)	300.8^{12} (239.0)	500^{20} (0.0)	29.8 (15.4)	**43.4**^{1} (45.7)	281.9^{11} (239.91)	153.3^{6} (208.0)	500^{20} (0.0)	43.2 (16.7)	50.0^{1} (45.0)	500^{20} (0.0)	29.7 (8.9)	500^{20} (0.0)	35.2 (9.7)	**24.4** (6.4)	500^{20} (0.0)	212.8^{8} (229.8)
7×6	345.7^{13} (200.7)	13.5 (7.6)	**11.6** (7.2)	350.6^{13} (209.2)	350.6^{14} (209.2)	500^{20} (0.0)	60.1^{1} (45.1)	**67.9**^{3} (86.4)	308.5^{11} (229.9)	228.1^{9} (244.8)	500^{20} (0.0)	68.9^{1} (44.5)	36.1 (10.4)	500^{20} (0.0)	34.1 (9.5)	500^{20} (0.0)	74.6^{1} (44.9)	**52.1**^{1} (47.5)	500^{20} (0.0)	283.3^{11} (238.4)
7×7	430.2^{16} (117.3)	17.9 (6.9)	**14.7** (9.1)	400.4^{16} (159.4)	400.4^{16} (159.4)	500^{20} (0.0)	55.9^{1} (47.0)	**45.7**^{1} (45.8)	259.8^{9} (240.2)	106.4^{4} (157.4)	500^{20} (0.0)	77.2^{1} (53.5)	76.8^{2} (84.7)	500^{20} (0.0)	37.5 (10.0)	500^{20} (0.0)	73.2^{1} (46.7)	**29.1** (5.5)	500^{20} (0.0)	307.8^{12} (230.7)
7×8	382.5^{14} (176.3)	38.9 (46.1)	**9.2** (3.4)	350.6^{15} (209.2)	400.4^{16} (159.4)	500^{20} (0.0)	122.8^{4} (150.9)	**93.5**^{3} (122.0)	335.7^{12} (213.6)	231.4^{9} (241.7)	500^{20} (0.0)	53.6 (15.8)	**28.1** (9.1)	500^{20} (0.0)	36.0 (11.7)	500^{20} (0.0)	128.9^{3} (117.2)	**30.6** (6.3)	500^{20} (0.0)	190.6^{7} (216.6)
7×9	453.2^{17} (84.2)	22.8 (11.0)	**11.5** (3.3)	325.7^{13} (226.6)	300.8^{12} (239.0)	500^{20} (0.0)	101.4^{3} (119.6)	**46.2** (45.4)	295.1^{11} (225.4)	82.4^{3} (125.3)	500^{20} (0.0)	83.8^{1} (50.3)	**31.6** (7.3)	500^{20} (0.0)	38.1 (9.0)	500^{20} (0.0)	41.8 (10.5)	**53.0** (44.7)	500^{20} (0.0)	214.5^{8} (228.4)
7×10	379.0^{15} (181.6)	11.1 (4.7)	**8.4** (4.5)	400.4^{16} (159.4)	400.4^{16} (159.4)	500^{20} (0.0)	114.9^{3} (115.5)	**69.2**^{2} (86.2)	299.0^{10} (221.2)	155.6^{6} (206.7)	500^{20} (0.0)	138.4^{4} (144.6)	**77.6**^{2} (84.5)	500^{20} (0.0)	107.5^{3} (117.8)	500^{20} (0.0)	46.5 (9.7)	**52.0**^{1} (44.8)	500^{20} (0.0)	166.9^{8} (199.9)

Note. The numbers in superscript of time-stability denote the number of instances out of 20 for which the corresponding policy failed to converge within T time periods.

Figure 9 **Average time-stability for policies in $\underline{\Lambda}$ for different types of feedback as p_r and p_v increase for $k = 6$ and the right-skewed costs ($T = 50$, layered graphs, greedy evader).**



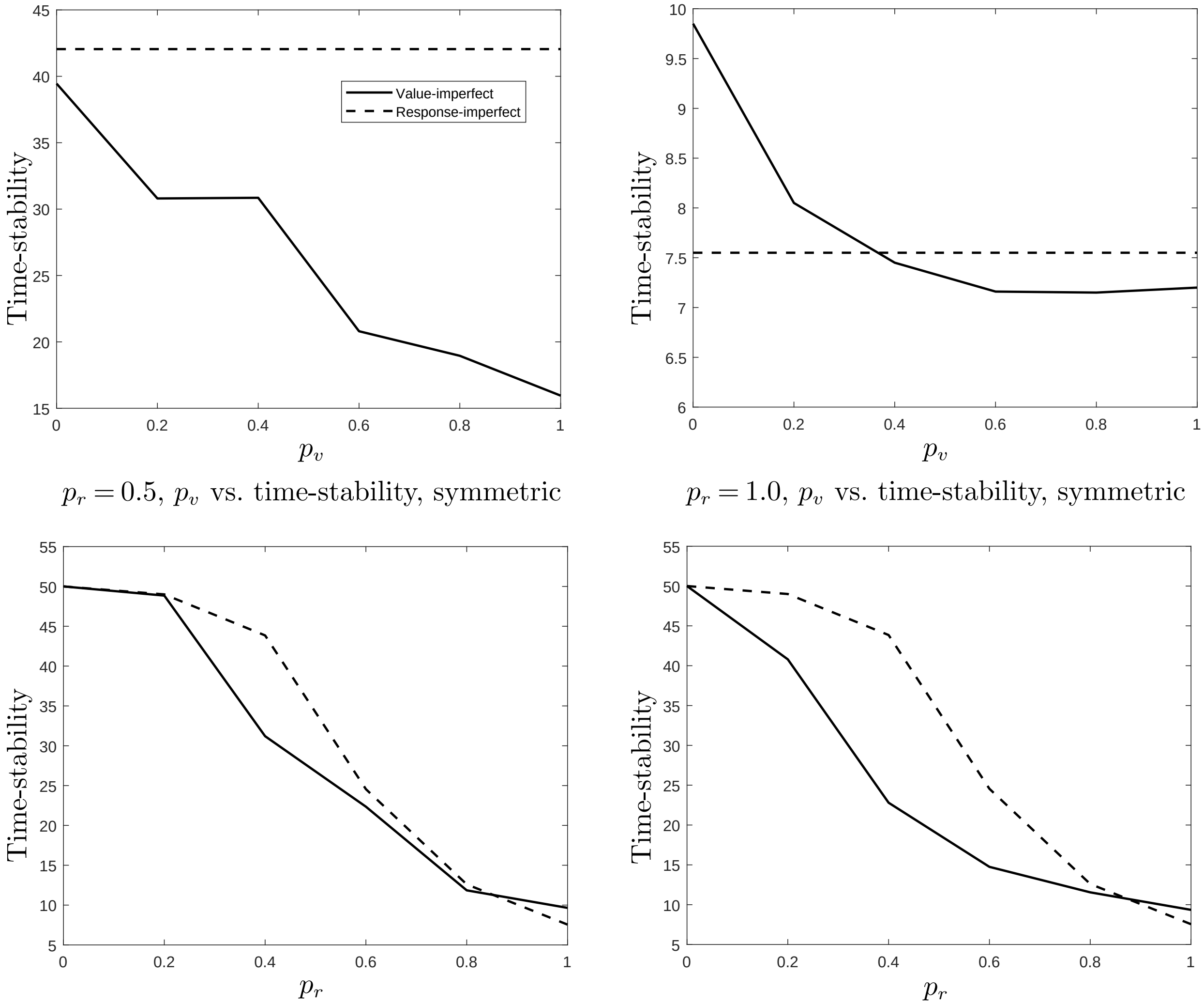
$p_r = 0.5$, p_v vs. time-stability, right-skewed

$p_r = 1.0$, p_v vs. time-stability, right-skewed

$p_v = 0.5$, p_r vs. time-stability, right-skewed

$p_v = 1.0$, p_r vs. time-stability, right-skewed

Figure 10 Average time-stability for policies in $\underline{\Lambda}$ for different types of feedback as p_r and p_v increase for $k = 6$ and the symmetric costs ($T = 50$, layered graphs, greedy evader).



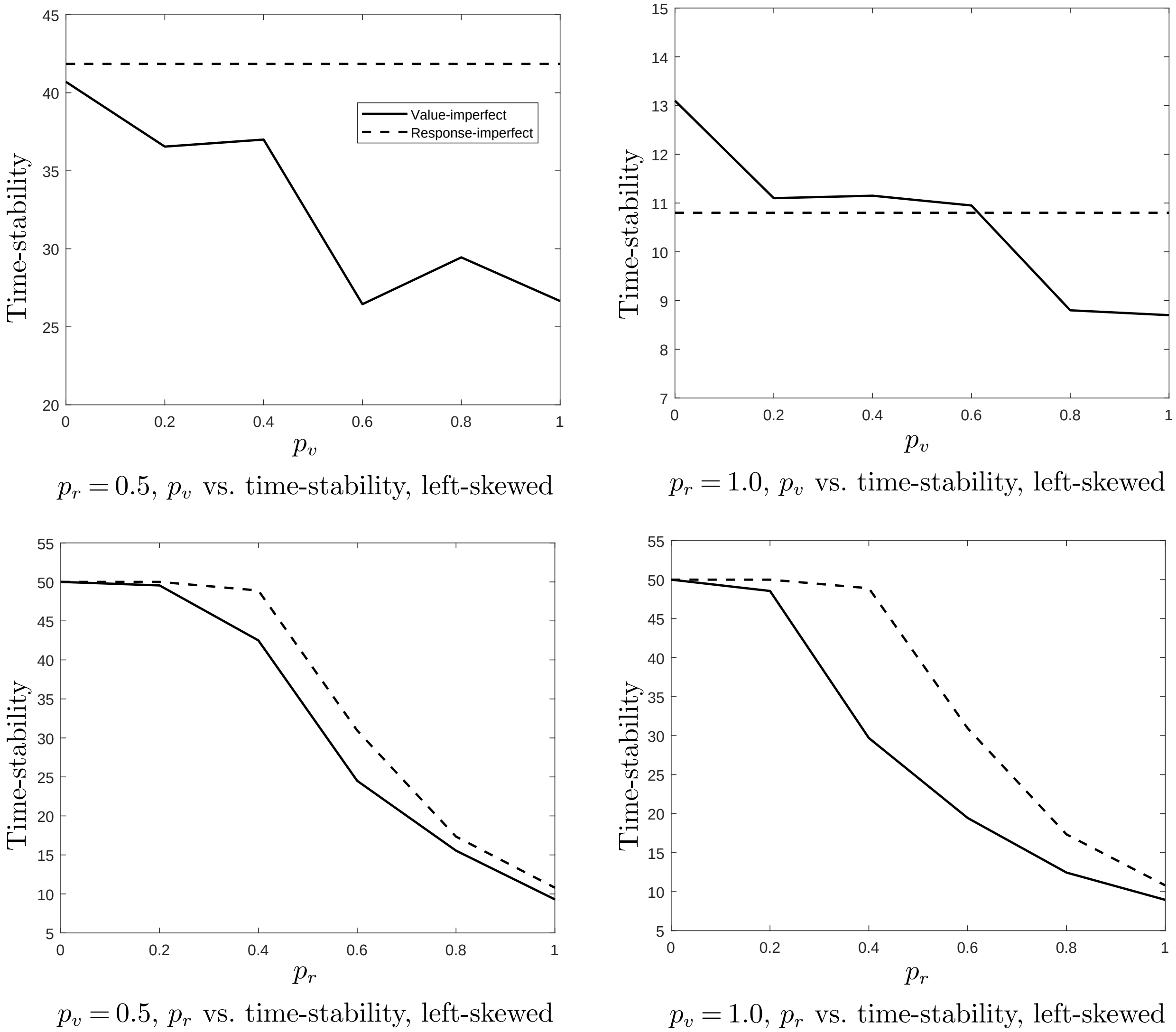
$p_r = 0.5$, p_v vs. time-stability, symmetric

$p_r = 1.0$, p_v vs. time-stability, symmetric

$p_v = 0.5$, p_r vs. time-stability, symmetric

$p_v = 1.0$, p_r vs. time-stability, symmetric

Figure 11 **Average time-stability for policies in $\underline{\Lambda}$ for different types of feedback as p_r and p_v increase for $k = 6$ and the left-skewed costs ($T = 50$, layered graphs, greedy evader).**



$p_r = 0.5$, p_v vs. time-stability, left-skewed

$p_r = 1.0$, p_v vs. time-stability, left-skewed

$p_v = 0.5$, p_r vs. time-stability, left-skewed

$p_v = 1.0$, p_r vs. time-stability, left-skewed

Table 10 **Behaviour of policies in $\underline{\Lambda}$ with respect to the cost bound quality ($k = 6$, $T = 200$, layered graphs, greedy evader).**

	$p_r = p_v$		0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1
Policy λ_r	I.1	Time-stability	**150.3**	**107.1**	**44.4**	**19.0**	**10.5**	**8.2**	**6.1**	**5.0**	**4.1**	**7.8**	**2.7**
		MAD	54.7	60.8	36.8	8.3	3.9	3.8	2.9	2.0	1.7	9.3	0.6
		Total regret	**6102.0**	**5447.0**	**3063.7**	**126.4**	**56.3**	**32.7**	**26.6**	**17.3**	**19.7**	**43.2**	**7.5**
		MAD	5112.9	5995.8	4908.5	142.8	52.5	34.3	25.9	17.0	23.1	65.1	7.2
	I.2	Time-stability	193.3	181.0	112.7	53.6	31.3	15.3	12.4	14.0	11.8	15.6	4.9
		MAD	12.8	30.5	58.0	26.9	14.4	6.6	4.0	10.3	9.7	18.0	1.1
		Total regret	13827.2	12969.3	8212.0	2822.1	1632.5	658.1	575.2	582.5	610.3	567.9	183.2
		MAD	9674.3	7633.7	6920.5	2336.6	1251.9	460.8	446.1	422.9	601.8	621.1	113.1
	I.3	Time-stability	200.0	198.1	155.1	70.2	54.4	30.4	27.6	13.7	14.4	9.0	7.1
		MAD	0.0	3.6	44.3	29.3	30.2	8.8	17.5	3.7	9.5	1.3	0.1
		Total regret	33018.8	21630.7	11753.0	5899.6	4605.6	2515.5	2334.1	1109.2	981.0	772.8	546.4
		MAD	12001.0	7696.5	5683.3	2858.8	2406.5	1113.8	1575.9	412.2	648.2	360.3	208.9
Policy λ_v	I.1	Time-stability	**161.5**	**75.4**	**28.8**	**13.1**	**6.6**	**6.2**	**4.9**	**4.3**	**4.4**	**3.5**	**3.2**
		MAD	57.8	55.2	16.5	7.0	2.7	2.7	2.0	1.7	1.3	1.3	1.1
		Total regret	**12643.3**	**3461.0**	**361.0**	**53.2**	**38.5**	**28.6**	**17.0**	**18.7**	**19.7**	**17.7**	**16.6**
		MAD	12171.1	4513.4	417.7	58.4	39.6	27.1	16.6	19.8	20.0	19.4	17.8
	I.2	Time-stability	193.3	163.9	81.6	30.8	17.6	20.3	9.5	8.1	15.7	5.4	4.9
		MAD	12.8	49.0	51.2	15.5	6.5	18.0	3.4	2.9	18.4	1.3	1.2
		Total regret	13894.5	9782.2	5299.3	1174.7	691.5	968.0	317.4	327.8	573.2	189.0	171.3
		MAD	8683.5	5049.2	5479.6	934.9	380.5	1181.5	196.9	238.3	667.4	119.4	102.0
	I.3	Time-stability	200.0	197.3	123.3	49.2	29.1	30.6	15.7	21.9	9.9	8.6	7.1
		MAD	0.0	5.0	29.8	12.0	6.0	18.0	3.5	17.8	1.4	0.9	0.1
		Total regret	33589.3	22066.9	11660.3	4200.6	2533.7	2043.9	1333.5	1285.0	795.4	637.6	546.4
		MAD	13519.1	9613.6	4940.6	1440.0	1143.8	956.6	611.9	809.8	345.0	262.4	208.9

Note. The numbers in superscript of time-stability denote the number of instances out of 20 for which the corresponding policy failed to converge within T time periods

Table 11 Performance of policies in Λ and benchmark policies without information updates ($n=15$, Watts-Strogatz graphs, greedy evader).

Graph		$k=2$ ($T=500$)				$k=4$ ($T=1000$)			
		λ	π_L	π_M	π_R	λ	π_L	π_M	π_R
Right-skewed	Time-stability	**28.4**	325.7^{13}	101.6^{5}	400.4^{16}	**280.0**^{3}	850.3^{17}	311.8^{8}	850.6^{17}
	MAD	37.8	226.6	159.4	159.4	303.2	254.5	412.9	254.1
	Relative difference	**0.0%**	13.8%	3.4%	17.9%	**1.4%**	23.2%	1.9%	21.0%
	MAD	0.0%	12.6%	5.3%	11.8%	2.4%	11.3%	2.1%	12.4%
	Total regret	**90.9**	3775.0	875.0	4978.4	**1629.1**	17050.0	1772.4	15650.2
	MAD	145.2	3425.0	1350.0	3328.4	1997.1	8255.0	1949.6	9414.9
Symmetric	Time-stability	182.0^{4}	375.5^{15}	**87.3**^{3}	195.2^{7}	843.9^{15}	800.4^{16}	**178.5**^{3}	690.0^{12}
	MAD	143.2	186.8	135.3	220.7	234.2	319.4	267.5	412.4
	Relative difference	2.6%	14.1%	**0.6%**	4.7%	14.9%	13.9%	**1.0%**	10.2%
	MAD	4.2%	13.2%	1.0%	6.1%	11.2%	9.0%	1.7%	8.9%
	Total regret	749.2	3025.0	**155.4**	1048.9	8347.9	7350.0	**655.9**	5715.5
	MAD	622.0	2935.0	247.1	1263.6	4912.7	4985.0	973.4	4588.0
Left-skewed	Time-stability	315.4^{10}	176.3^{7}	246.2^{7}	**88.6**^{1}	1000.0^{20}	**401.2**^{8}	702.2^{13}	723.7^{13}
	MAD	184.6	226.6	191.8	99.4	0.0	479.0	387.2	359.2
	Relative difference	12.4%	4.9%	7.6%	**0.5%**	33.9%	**5.2%**	18.6%	13.1%
	MAD	13.2%	6.3%	9.9%	1.0%	13.2%	6.6%	14.3%	13.3%
	Total regret	2362.7	850.0	1543.6	**646.0**	14373.8	**2150.0**	7452.3	6734.9
	MAD	1687.4	1105.0	1373.7	798.4	6177.3	2710.0	4383.1	4523.3
Random	Time-stability	281.4^{9}	350.6^{17}	**216.6**^{7}	301.6^{12}	940.9^{18}	900.2^{18}	**342.2**^{6}	751.1^{15}
	MAD	196.8	209.2	206.0	238.1	106.4	179.6	404.1	373.4
	Relative difference	10.7%	15.8%	**4.5%**	11.4%	31.6%	20.3%	**5.0%**	15.1%
	MAD	12.8%	8.8%	6.1%	12.1%	10.9%	10.9%	7.1%	13.9%
	Total regret	2715.5	3250.0	**1293.3**	2540.2	16614.3	11900.0	**2985.4**	9675.5
	MAD	2411.5	1925.0	1442.2	2825.9	4588.5	7190.0	3706.1	8355.2

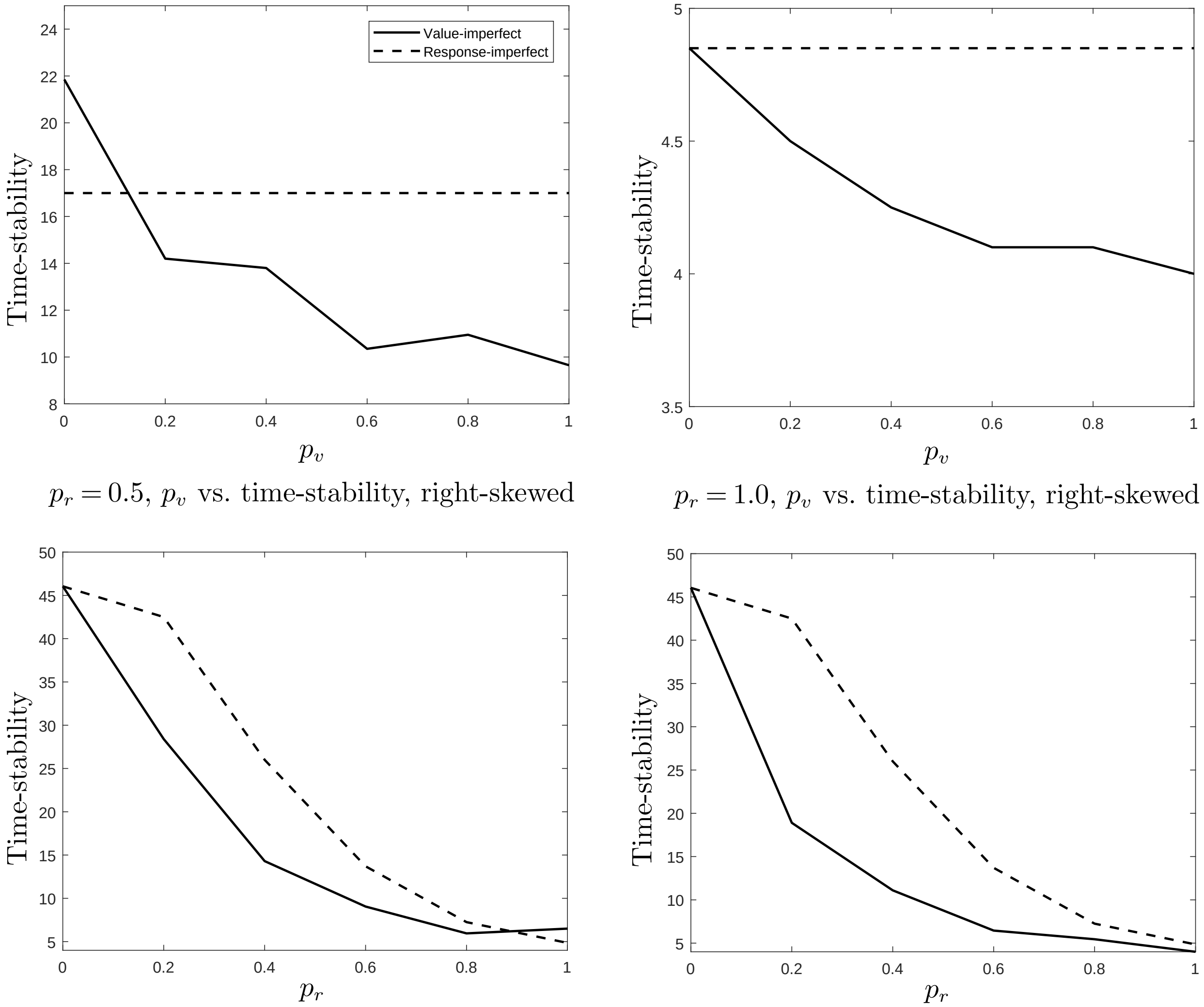
Notes: Entries in bold denote the best policy in each setting; the numbers in superscript of time-stability denote the number of instances out of 20 for which the corresponding policy failed to converge within T time periods.

Table 12 **Average time-stability and MAD (in parenthesis) for $\lambda \in \underline{\Lambda}$ and $\pi_M \in \Pi_M$ policies when information updates are used ($k = 6$, $T = 500$, Watts-Strogatz graphs, greedy evader).**

n	Right-skewed					Symmetric					Left-skewed					Random				
	λ	λ_r	λ_v	π_M	$\pi_M(+\mathcal{V})$	λ	λ_r	λ_v	π_M	$\pi_M(+\mathcal{V})$	λ	λ_r	λ_v	π_M	$\pi_M(+\mathcal{V})$	λ	λ_r	λ_v	π_M	$\pi_M(+\mathcal{V})$
15	218.6	36.6	**20.3**	215.4	251.0	500.0	66.3	**15.1**	74.2	123.0	500.0	82.1	**68.6**	500.0	157.0	500.0	89.8	**20.8**	364.7	337.6
	(211.2)	(11.0)	(26.5)	(243.9)	(249.0)	(0.0)	(8.5)	(3.6)	(121.7)	(170.0)	(0.0)	(4.8)	(79.5)	(0.0)	(196.0)	(0.0)	(9.1)	(8.4)	(193.3)	(208.8)
20	363.4	41.5	**32.8**	251.0	251.0	500.0	123.5	**18.8**	152.0	103.5	500.0	72.5	**48.9**	500.0	202.4	500.0	94.6	**53.0**	410.1	354.5
	(191.3)	(54.3)	(46.7)	(249.0)	(249.0)	(0.0)	(149.2)	(5.1)	(208.8)	(158.6)	(0.0)	(47.6)	(45.1)	(0.0)	(208.3)	(0.0)	(84.0)	(54.8)	(143.9)	(203.7)
25	347.2	43.1	**11.1**	251.0	251.0	500.0	59.5	**45.0**	220.9	61.1	500.0	61.6	**26.8**	500.0	143.3	500.0	58.4	**52.6**	476.8	333.3
	(200.3)	(45.8)	(4.1)	(249.0)	(249.0)	(0.0)	(45.9)	(45.5)	(231.2)	(87.8)	(0.0)	(21.4)	(4.9)	(0.0)	(156.3)	(0.0)	(22.3)	(44.7)	(44.2)	(216.7)
30	475.8	41.3	**9.9**	275.9	275.9	500.0	43.4	**19.7**	129.7	153.1	500.0	63.4	**26.7**	500.0	128.5	500.0	61.7	**28.2**	452.1	354.0
	(46.1)	(47.9)	(2.9)	(246.5)	(246.5)	(0.0)	(23.1)	(5.6)	(185.2)	(208.1)	(0.0)	(27.2)	(6.0)	(0.0)	(148.6)	(0.0)	(21.7)	(7.7)	(81.5)	(204.4)
35	428.9	21.0	**10.1**	350.6	375.5	500.0	61.4	**21.2**	257.1	179.4	500.0	72.8	**37.4**	500.0	180.8	500.0	50.5	**29.6**	479.6	253.3
	(121.0)	(10.7)	(4.0)	(209.2)	(186.8)	(0.0)	(44.1)	(5.1)	(242.9)	(224.5)	(0.0)	(34.3)	(9.1)	(0.0)	(191.5)	(0.0)	(18.9)	(7.2)	(38.9)	(233.8)
40	427.4	73.7	**60.1**	375.5	350.6	500.0	38.2	**25.6**	159.5	80.9	500.0	90.1	**56.4**	500.0	223.8	500.0	98.0	**75.9**	500.0	236.2
	(123.5)	(85.3)	(88.0)	(186.8)	(209.2)	(0.0)	(9.8)	(6.4)	(204.3)	(125.7)	(0.0)	(53.2)	(44.4)	(0.0)	(221.0)	(0.0)	(82.5)	(84.8)	(0.0)	(237.4)
45	377.9	46.7	**10.3**	350.6	350.6	500.0	65.6	**48.3**	304.0	83.2	500.0	93.6	**62.7**	500.0	271.1	500.0	94.9	**35.5**	500.0	288.1
	(183.2)	(47.0)	(3.4)	(209.2)	(209.2)	(0.0)	(48.8)	(45.4)	(235.3)	(125.0)	(0.0)	(53.7)	(45.8)	(0.0)	(228.9)	(0.0)	(49.4)	(6.7)	(0.0)	(233.1)
50	403.0	17.0	**11.5**	400.4	400.4	500.0	70.3	**21.4**	205.6	276.9	500.0	93.1	**34.0**	500.0	270.9	500.0	111.7	**31.2**	500.0	358.0
	(155.2)	(5.7)	(3.1)	(159.4)	(159.4)	(0.0)	(50.4)	(4.8)	(235.5)	(245.5)	(0.0)	(51.7)	(7.6)	(0.0)	(229.2)	(0.0)	(78.5)	(7.9)	(0.0)	(198.9)

Note. The numbers in superscript of time-stability denote the number of instances out of 20 for which the corresponding policy failed to converge within T time periods.

Figure 12 **Average time-stability for policies in $\underline{\Lambda}$ for different types of feedback as p_r and p_v increase for $k = 6$ and the right-skewed costs ($T = 50$, Watts-Strogatz graphs, greedy evader).**



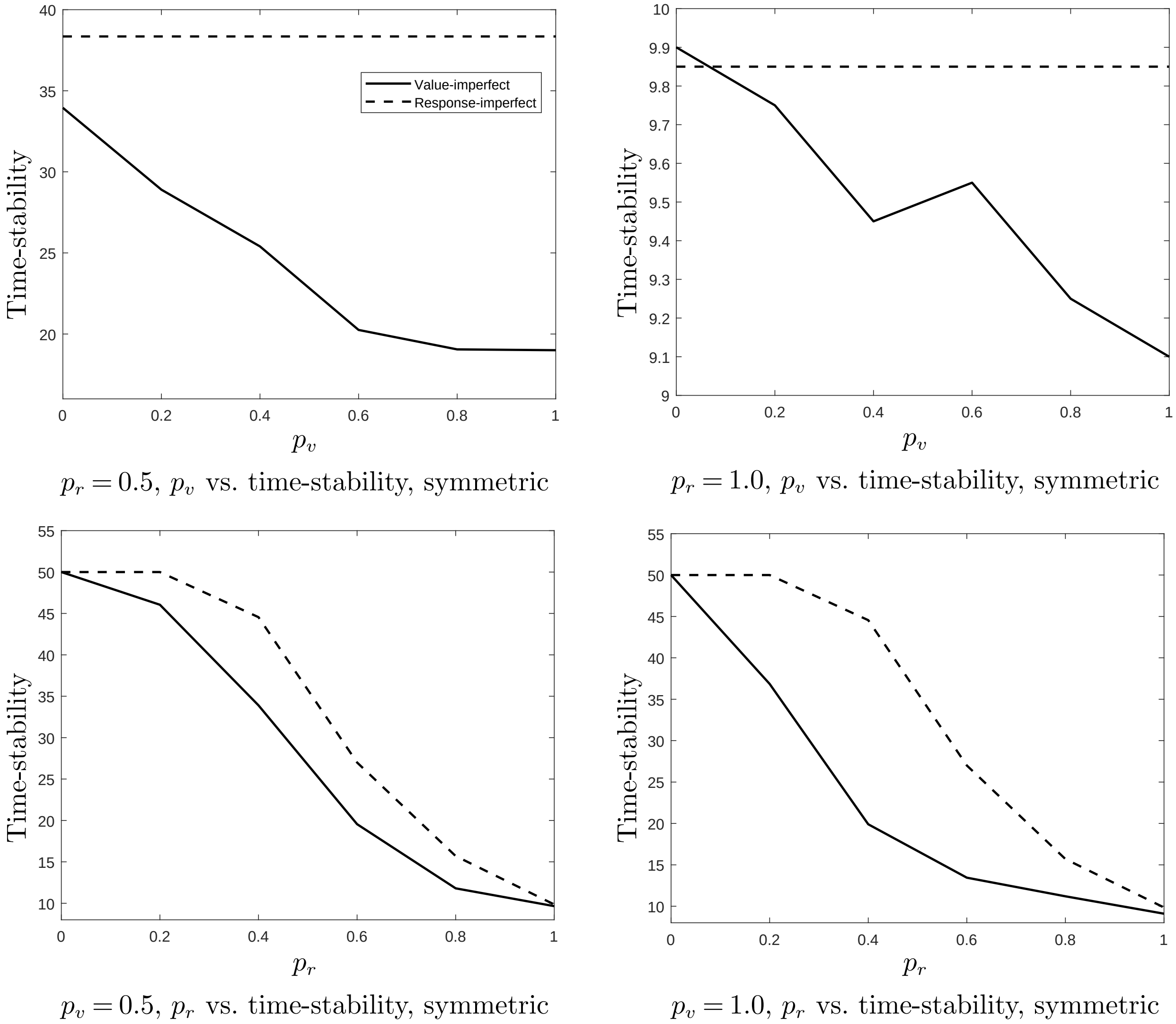
$p_r = 0.5$, p_v vs. time-stability, right-skewed

$p_r = 1.0$, p_v vs. time-stability, right-skewed

$p_v = 0.5$, p_r vs. time-stability, right-skewed

$p_v = 1.0$, p_r vs. time-stability, right-skewed

Figure 13 **Average time-stability for policies in $\underline{\Lambda}$ for different types of feedback as p_r and p_v increase for $k=6$ and the symmetric costs ($T=50$, Watts-Strogatz graphs, greedy evader).**



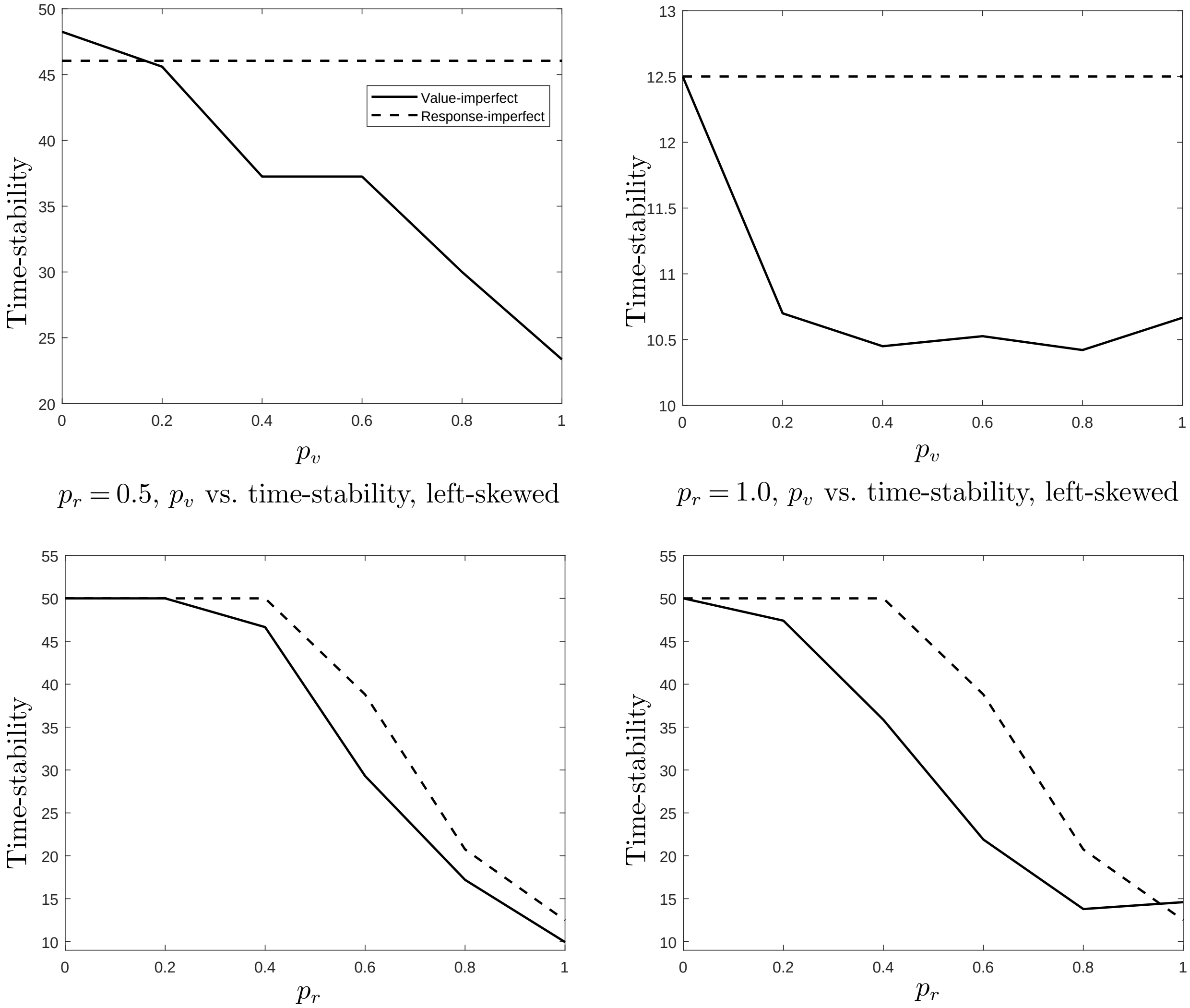
$p_r = 0.5$, p_v vs. time-stability, symmetric

$p_r = 1.0$, p_v vs. time-stability, symmetric

$p_v = 0.5$, p_r vs. time-stability, symmetric

$p_v = 1.0$, p_r vs. time-stability, symmetric

Figure 14 **Average time-stability for policies in $\underline{\Lambda}$ for different types of feedback as p_r and p_v increase for $k=6$ and the left-skewed costs ($T=50$, Watts-Strogatz graphs, greedy evader).**



$p_r = 0.5$, p_v vs. time-stability, left-skewed

$p_r = 1.0$, p_v vs. time-stability, left-skewed

$p_v = 0.5$, p_r vs. time-stability, left-skewed

$p_v = 1.0$, p_r vs. time-stability, left-skewed

Table 13 Behaviour of policies in $\underline{\Lambda}$ with respect to the cost bound quality ($k=6$, $T=200$, Watts-Strogatz graphs, greedy evader).

	$p_r = p_v$		0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1
Policy λ_r	I.1	Time-stability	**105.5**	**77.2**	**31.6**	**16.3**	**12.4**	**6.4**	**5.8**	**3.6**	**3.6**	**2.7**	**2.5**
		MAD	70.1	65.3	17.3	7.7	7.8	3.0	2.5	1.4	1.4	0.8	0.6
		Total regret	**1572.6**	**909.9**	**390.3**	**236.4**	**151.6**	**19.6**	**18.0**	**10.7**	**6.7**	**7.1**	**5.2**
		MAD	2193.9	1306.2	600.4	338.4	217.5	27.0	25.3	15.4	9.1	10.1	7.1
											7.7		
	I.2	Time-stability	200.0	151.6	84.6	41.5	21.3	17.3	9.0	7.8	6.3	5.1	4.3
		MAD	0.0	48.2	44.2	21.7	6.5	12.3	2.7	2.4	1.6	1.2	0.8
		Total regret	10297.9	9986.2	5490.7	1881.5	832.8	852.7	350.9	268.7	202.4	158.5	126.2
		MAD	6342.5	7238.2	4813.5	1731.2	490.5	946.4	265.6	190.0	139.6	98.0	76.3
											7.7		
	I.3	Time-stability	200.0	200.0	138.9	81.6	46.2	29.7	22.6	20.1	11.6	8.1	7.0
		MAD	0.0	0.0	39.7	31.4	18.7	13.4	11.9	10.4	2.1	0.7	0.0
		Total regret	27668.4	20968.8	14129.3	7069.2	3950.6	2552.3	2177.5	1744.6	797.9	591.7	502.2
		MAD	10027.8	10213.1	5741.8	3264.7	1774.7	1528.0	1648.3	1211.4	253.4	173.5	129.0
Policy λ_v	I.1	Time-stability	**109.1**	**60.5**	**22.3**	**15.4**	**6.2**	**6.1**	**3.4**	**3.8**	**3.1**	**2.6**	**2.5**
		MAD	74.1	36.5	8.5	9.7	2.6	2.7	1.0	1.5	0.6	0.6	0.5
		Total regret	**2244.2**	**1163.8**	**46.1**	**23.3**	**15.4**	**13.3**	**9.0**	**10.0**	**5.7**	**5.2**	**5.2**
		MAD	3268.4	1874.0	64.4	32.9	22.0	19.0	12.8	13.9	7.7	7.1	7.1
											7.7		
	I.2	Time-stability	200.0	144.5	58.4	22.4	15.1	9.7	8.3	6.0	5.6	4.7	4.3
		MAD	0.0	38.7	28.6	7.7	6.2	3.8	3.1	1.6	1.4	1.1	0.7
		Total regret	10785.2	6594.4	3361.4	742.1	517.2	341.1	228.6	180.5	156.7	142.6	132.9
		MAD	6942.5	4936.1	3504.3	549.0	409.9	263.9	138.1	106.9	95.1	89.4	84.3
											7.7		
	I.3	Time-stability	200.0	198.0	108.6	68.7	29.1	20.3	15.0	11.7	10.4	8.4	7.3
		MAD	0.0	3.5	31.4	33.2	8.0	3.6	2.4	1.9	1.3	0.9	0.5
		Total regret	27200.2	20596.5	9600.6	5408.1	2264.0	1610.5	1214.1	844.4	753.8	578.3	513.7
		MAD	11068.9	6487.9	4343.4	2991.2	1072.0	372.5	414.4	300.8	235.1	133.8	129.9

Note. The numbers in superscript of time-stability denote the number of instances out of 20 for which the corresponding policy failed to converge within T time periods