

Sequential Max-Min Bilevel Linear Programming with Incomplete Information and Learning

Juan S. Borrero, Oleg A. Prokopyev

Department of Industrial Engineering, University of Pittsburgh, Pittsburgh, PA 15261, {jsb81,droleg}@pitt.edu

Denis Sauré

Department of Industrial Engineering, University of Chile, Santiago, Chile dsauré@dii.uchile.cl

We present a framework for a class of sequential decision-making problems in the context of max-min bilevel programming, where a *leader* and a *follower* repeatedly interact. At each period, the leader allocates resources to disrupt the performance of the follower (e.g., as in defender-attacker or interdiction problems), who in turn minimizes some cost function over a set of activities that depends on the leader's decision. While the follower has complete knowledge of his problem, the leader has only partial information, and needs to learn about the cost parameters, available resources, and the follower's activities from the feedback generated by the follower's actions. We measure policy performance in terms of its *time-stability*, defined as the number of periods it takes for the leader to match the actions of an *oracle* with complete information. In particular, we propose a class of *greedy* and *robust* policies and show that these policies are *weakly optimal*, eventually match the oracle's actions, and provide a real-time certificate of optimality. We also study a lower bound on any policy performance based on the notion of a *semi-oracle*. Our numerical experiments demonstrate that the proposed policies consistently outperform reasonable benchmark, and perform fairly close to the semi-oracle.

Key words: bilevel programming, attacker-defender, interdiction, learning, incomplete information, online optimization, robust optimization

1. Introduction

Bilevel optimization deals with problems where a subset of the *lower-level* decisions are constrained to be a solution of another mathematical program that depends on the remaining *upper-level* decisions. This general structure makes bilevel programs useful for modeling hierarchical decision-making problems between multiple, typically, two actors, commonly referred to as the *leader* (an upper-level decision-maker) and the *follower* (a lower-level decision-maker), see Dempe (2002). In this perspective, the leader solves an optimization problem that depends on the optimal solution to the follower's problem, and this latter problem is, in turn, parameterized by the leader's decisions. Bilevel programs are used in several application areas such as law enforcement (Morton et al. 2007), defense (Brown et al. 2006), economics (Sherali et al. 1983), transportation (Lucotte and Nguyen 2013), energy (Bard et al. 2000), revenue management (Côté et al. 2003), among others; see Dempe (2002), Colson et al. (2005, 2007) and the references therein.

An important class of bilevel programs, known as *max-min* problems, deals with settings where the leader and follower are adversaries. More precisely, in these problems the leader’s objective is to maximally degrade the performance of the follower. As an example, consider network flow interdiction problems, which have applications in military and smuggling prevention settings. Here, the follower operates a network with the objective to move between two fixed vertices through a shortest path (Fulkerson and Harding 1977, Corley and Chang 1974, Israeli and Wood 2002), to send the maximum flow possible between two vertices (Wollmer 1964, McMasters and Mustin 1970, Ghare et al. 1971, Corley and Chang 1974, Ratliff et al. 1975, Wood 1993), or more generally, to move flow in the network at minimum cost subject to some demand balance constraints (Chern and Lin 1995, Smith and Lim 2008). The leader, by using the resources at her disposition, can block (either totally or partially) a limited number of arcs and nodes in the network. Her objective is to allocate her resources so as to maximize the length of the follower’s shortest path, minimize the maximum flow, or maximize the minimum cost incurred by the follower, respectively. These types of models are also used in surveillance settings, where the leader places resources (e.g., sensors) in a network so as to minimize the follower’s probability of evasion, see Brown et al. (2006), Morton et al. (2007).

Network flow interdiction models belong to a larger class of Attacker-Defender (AD) or Defender-Attacker (DA) models (Brown et al. 2006, Wood 2011). In a typical AD setting, an attacker (the leader) and a defender (the follower) interact during a war-time confrontation: the attacker allocates her forces so as to disable assets of the defender’s infrastructure; the defender decides how to operate his system at minimum cost given the restrictions set by the leader’s attack. The leader decides her allocation with the objective to maximize the defender’s operational costs. Conversely, in a DA model, a defender (the leader) allocates her limited defensive resources to protect her assets, and an attacker (the follower), for a given defensive configuration, seeks for the most effective attacks. Here, the defender’s objective is to allocate her resources so as to minimize the effectiveness of the attacks. In general, AD and DA models can be casted as max-min bilevel programs to model decisions in a broad range of application areas: see, e.g., Salmeron et al. (2004), Brown et al. (2006), Zenklusen (2010), Shen et al. (2012), Brown et al. (2005).

Typical formulations of max-min bilevel problems in the literature assume a single interaction between the leader and the follower, and that either the leader knows all the parameters of the follower’s problem (as in the references discussed above), or that she knows a probability distribution over the set of problem configurations and parameters (see e.g., Hemmecke et al. (2003), Held et al. (2005), Held and Woodruff (2005), Janjarassuk and Linderoth (2008)). Hence, these models solve a single (possibly stochastic) max-min bilevel problem, assuming that even if the leader and the follower interact across several periods, the leader would implement the resulting *full-information* solution at every time period. In contrast, many applications inherently

involve multiple interactions between the leader and the follower (e.g., as in smuggling interdiction and AD-DA problems). More importantly, in these problems the leader does not always know with certainty the system that the follower operates, and cannot estimate it (a priori) reliably due to the adversarial nature of their confrontation. Consequently, she has *incomplete information* of the problem solved by the follower at each time period, and has to learn about it through time by observing the follower’s reactions to her actions.

Departing from the existing literature, this paper studies *sequential max-min problems with incomplete information* (**SMPI**). In these problems, the leader and follower interact *repeatedly*: at each stage the leader implements a set of actions and then observes the follower’s reaction; from the information, or *feedback*, she gets from the follower’s response, the leader (potentially) updates her knowledge of the follower’s problem, and incorporates this information into her decision-making process. Observe that in **SMPI**, besides determining how to allocate her resources, the leader faces additional questions outside the scope of traditional bilevel models, as she needs to recognize whether a given upper-level solution is the best possible, she needs to force the follower to disclose as much information as possible, and needs to exploit this newly learned information to best re-allocate their resources in future periods. Therefore, given the leader’s limited knowledge of the follower’s problem, at each time period she faces a form of the *exploitation vs. exploration trade-off*: she must choose either to exploit the current information so as to maximize her immediate reward, or to explore solutions that albeit not being maximally rewarding, may reveal new information that can be used to implement better solutions in future periods.

For the reasons above, **SMPI** can be viewed as a class of *online optimization problems*. In said problems, at each period a decision-maker with incomplete knowledge of the problem’s structure has to choose a solution from a fixed action set, incurring a cost. She learns the problem’s structure from the feedback she collects, which depends on the particular implemented solution (Cesa-Bianchi and Lugosi 2006, Zinkevich 2003). Performance is typically measured in terms of *regret*, i.e., the difference between the costs incurred by the decision-maker and those incurred by an *oracle*, i.e., a decision-maker with complete up-front knowledge of the problem. Online optimization is particularly useful to tackle sequential decision-making problems, where the decision-maker has limited knowledge regarding the structure of uncertainty (Hazan 2015), and where there might be no functional assumptions that relate the actions with the costs.

There are several different types of online optimization models differing on their assumptions on the action set, stochasticity of the rewards, and the type of feedback. There are, however, *three major drawbacks* of existing methods that make them unfit to solve **SMPI**:

(i) Current models consider a *fixed amount of possible actions* (with the exception of Kleinberg et al. (2010), that does not adequately specialize to our class of problems). That is, in most

existing models the set of actions available to the decision-maker at each time is always the same, independent of any new information that is discovered. Hence, such models cannot handle the fact that the leader might learn new means to affect the follower’s actions (e.g., new assets or activities in an AD model, new routes in a network flow interdiction setting) from observing the latter.

(ii) Naively using existing online policies, such as those applied to multi-armed bandit settings (see, e.g. Cesa-Bianchi et al. (1997), Auer et al. (2002), Audibert and Bubeck (2009), Cesa-Bianchi and Lugosi (2006)), would result in regret bounds that are *exponential in the primitives of the problem* (i.e., the number of variables and constraints). Specifically, general multi-armed bandit policies provide bounds that are proportional to the number of possible actions available to the decision-maker. In our context the number of possible actions correspond to the number of feasible solutions in the upper-level (leader’s) optimization problem, and it is well-known (see, e.g., Dempe (2002)) that said number is typically exponential in the number of the leader’s variables and constraints.

(iii) Online methods that explicitly consider problems with combinatorial structure (e.g. Cesa-Bianchi and Lugosi (2012), Audibert et al. (2013), Kalai and Vempala (2005), Gyorgy et al. (2007)), or infinitely many solutions, as in online convex optimization (e.g. Zinkevich (2003), Kalai and Vempala (2005), Awerbuch and Kleinberg (2004), Hazan et al. (2007), Hazan (2015)), assume a *single-level relationship* between the decision-maker’s actions and the costs she observes. Consequently, these models cannot address the hierarchical relationship between the leader and follower that is present in bilevel optimization models. To the best of our knowledge only the work by Borrero et al. (2016) consider a combinatorial problem within a hierarchical structure; this work is, however, limited to studying a particular shortest-path interdiction setting.

Given the limitations of current online models, in this paper we develop a general framework for **SMPI**. We represent the leader’s and follower’s decisions in terms of *resources* and *activities*, respectively. Initially, the leader does not know all the follower’s activities and constraints, and as such, she might not know all of her resources or constraints. The leader learns about an unknown follower’s activity as soon as she observes him *performing* it, and at the same time learns about all the lower-level constraints that *restrict* this activity, all the leader’s resources that *interfere* with that activity, and all upper-level constraints associated with the newly learned resources.

From a technical point of view, we first make the assumption that for every activity, resource, and constraint she knows, the leader also knows the corresponding entries in the upper and lower-level constraint matrices and the right-hand side vectors in a typical bilevel programming formulation of the full-information problem. However, we suppose that the leader does not know with certainty the components of the follower’s cost vector for the activities she knows; she only knows that they belong to certain (polyhedral) *uncertainty set*. Furthermore, in Section 4 we analyze a more general uncertainty model, where the uncertainty extends beyond the follower’s cost vector.

Besides learning new activities, resources, and constraints, the leader can also observe additional information of the follower’s problem from his response. In this sense, we introduce the notions of *Standard feedback*, and its specializations, *Value-Perfect* and *Response-Perfect feedbacks*. In Standard feedback, the leader observes the total cost the follower incurs at each time period; in Value-Perfect feedback she also observes the cost coefficient associated with each activity used by the follower at that time, while in Response-Perfect feedback she also observes the value of the decision vector for the activities performed by the follower.

We measure the performance of the leader’s decision-making policy in terms of its *time-stability*, i.e., the first time period by which the costs the follower incurs coincides with the maximum possible cost an *oracle* leader with complete knowledge of the bilevel problem attains. Time-stability is closely related to the *regret* (in particular, any upper bound on the time-stability of a policy implies an upper bound in the regret on that policy), a more common measure of performance in online optimization settings (Bubeck and Cesa-Bianchi 2012, Hazan 2015),

In this paper we analyze a set of *greedy* and *robust* policies, which we denote by Λ . The policies are greedy because at any time they exploit the leader’s information of the follower’s problem so as to maximize the follower’s costs at the current time period, and they are robust because they assume that the follower’s cost vector realizes its worst case for the leader. For these reasons, implementing the policies in Λ involve solving at each time a max-min bilevel problem with lower-level robustness constraints, and as such their computation involve both bilevel and robust optimization techniques: we develop a method that first replaces the lower-level robust optimization problem by its equivalent linear program counterpart (Ben-Tal et al. 2009), and then reformulates the resulting linear bilevel program as a one-level mixed integer program (Audet et al. 1997).

We demonstrate that the time-stability of policies in Λ under Value-Perfect and Response-Perfect feedback is upper bounded by the number of follower’s activities. We show that these policies are *optimal* in the sense that they attain the best possible worst-case time-stability across all possible problem instances. Furthermore, they provide a *certificate of optimality* in real time. We also develop a method to provide a lower bound for the time-stability of any policy based on the concept of a *semi-oracle*. The semi-oracle has full information of the problem beforehand, but cannot signal it through her actions. As such, the semi-oracle combines the knowledge of the standard oracle with the practical limitations of the leader. Our numerical results show that the policies in Λ consistently outperform reasonable benchmark, and perform reasonably close to the semi-oracle.

The remainder of the paper is organized as follows. In Section 2 we provide a mathematical formulation of the problem, and illustrate it with examples of the minimum-cost flow interdiction and the Attacker-Defender knapsack problems. Section 3 discusses greedy and robust policies along with their main properties, while Section 4 extends most of the results of greedy and robust policies

for the case of uncertainty in the lower-level constraint matrix. Section 5 discusses the semi-oracle benchmark and Section 6 presents numerical experiments. In Section 7 we give conclusions and directions for future work. Most proofs and supporting material are relegated to the appendix.

2. Basic Model: Cost Uncertainty

We consider a sequential and adversarial decision-making process where at each time $t \in \mathcal{T} := \{0, 1, \dots, T\}$ a *leader* and a *follower* interact. At the beginning of time $t \in \mathcal{T}$, in the *complete information setting*, the leader can use any *resource* $i \in I$, $|I| < \infty$, and for each $i \in I$ she chooses a value $x_i \geq 0$ such that $x := (x_i : i \in I) \in X$, where X denotes the set of feasible resource levels. We let C_L denote the set of constraints faced by the leader and assume that X is given by

$$X := \{x \in \mathbb{Z}_+^k \times \mathbb{R}_+^{|I|-k} : \mathbf{H}x \leq \mathbf{h}\},$$

where $0 \leq k \leq |I|$, $\mathbf{H} := (H_{di} : d \in C_L, i \in I) \in \mathbb{R}^{|C_L| \times |I|}$ and $\mathbf{h} := (h_d, d \in C_L) \in \mathbb{R}^{|C_L|}$.

The follower, on the other hand, reacts after the leader chooses x . He can pick different levels among his *activities* in a finite set A : we let y_a denote the level by which activity a is performed, and define $y := (y_a : a \in A)$. By performing activity a at level y_a the follower incurs a cost of $c_a y_a$, and hence he desires to select y so as to minimize his total costs. His choices for y are limited, however, as y should satisfy all the constraints in a set C_F and should also be feasible given the leader's decision x . Therefore, at time t the follower selects vector $y(x)$, where for any $x \in X$

$$y(x) \in \arg \min \{\mathbf{c}^\top y : y \in Y(x)\},$$

$\mathbf{c} := (c_a : a \in A) \in \mathbb{R}^{|A|}$, and where for any $x \in \mathbb{Z}_+^k \times \mathbb{R}_+^{|I|-k}$ the set $Y(x)$ denotes the follower's set of feasible actions given the leader decision x . We assume that

$$Y(x) := \{y \in \mathbb{R}_+^{|A|} : \mathbf{F}y + \mathbf{L}x \leq \mathbf{f}\}.$$

In the above, $\mathbf{F} := (F_{da} : d \in C_F, a \in A)$ belongs to $\mathbb{R}^{|C_F| \times |A|}$, $\mathbf{L} := (L_{di} : d \in C_F, i \in I)$ belongs to $\mathbb{R}^{|C_F| \times |I|}$ and $\mathbf{f} := (f_d, d \in C_F) \in \mathbb{R}^{|C_F|}$. The objective of the leader is to choose $x \in X$ so as to maximize the cumulative cost the follower faces through \mathcal{T} . Note that, *had the leader full information about the problem*, at each time $t \in \mathcal{T}$ she would implement a solution to the bilevel problem

$$z^* := \max \{z(x) : x \in X\}, \tag{1}$$

where for any $x \in X$,

$$z(x) := \min \{\mathbf{c}^\top y : y \in Y(x)\}.$$

Throughout this paper, we assume that at all times *the follower has the information needed to compute $y(x)$* , but that *this is not the case for the leader*: we assume that at time $t = 0$ the leader

does not fully know the set of activities A , and hence potentially neither C_F , nor the value of all the data defining region $Y(x)$. Moreover, as some leader's resources might be only available if some of the follower's activities are known, she might have only partial information regarding I , C_L and the set X . Specifically, at the beginning of each time $t \in \mathcal{T}$ the leader is aware of the subset of the follower's activities $A^t \subseteq A$, the subset of the leader's resources $I^t \subseteq I$, the upper-level constraints $C_L^t \subseteq C_L$ and the lower-level constraints $C_F^t \subseteq C_F$. Furthermore, the leader's knowledge of the follower's lower-level problem data is limited, and in this direction we make the following assumptions:

A1: At any time $t \in \mathcal{T}$ the leader knows with certainty the values of $\mathbf{F}^t := (F_{da} : a \in C_F^t, a \in A^t)$ and $\mathbf{f}^t := (f_d : d \in C_F^t)$. In addition, the leader knows with certainty all her data (both *upper-level* and *lower-level*) with respect to the resources in I^t , that is, at time t she knows with certainty $\mathbf{H}^t := (H_{di} : d \in C_L^t, i \in I^t)$, $\mathbf{h}^t := (h_d : d \in C_L^t)$ and $\mathbf{L}^t := (L_{di} : d \in C_F^t, i \in I^t)$.

A2: The leader does not know with certainty all the entries of \mathbf{c} but she knows that $\mathbf{c}^t := (c_a : a \in A^t) \in \mathcal{U}^t$, with

$$\mathcal{U}^t := \{\hat{\mathbf{c}}^t \in \mathbb{R}^{|A^t|} : \mathbf{G}^t \hat{\mathbf{c}}^t \leq \mathbf{g}^t\}.$$

If C_U^t is the set of constraints of polyhedron \mathcal{U}^t , then $\mathbf{G}^t \in \mathbb{R}^{|C_U^t| \times |A^t|}$ and $\mathbf{g}^t \in \mathbb{R}^{|C_U^t|}$. We assume that both \mathbf{G}^t and \mathbf{g}^t are known with certainty to the leader at time t .

A3: The matrix \mathbf{H} and vector \mathbf{h} take non-negative values.

A4: For any $x \in X$, $\mathbf{L}x \leq \mathbf{f}$.

Assumption **A1** implies that, with the *exception of the cost vector*, the leader knows with certainty all the problem data in (1) that is associated with activities in A^t , resources in I^t , and constraints in C_F^t and C_L^t . Particularly, the latter part of this assumption stems from the idea that the leader is always certain about her operational capabilities (hence, she always knows \mathbf{H} and \mathbf{h} for all activities and constraints known to her), and about the effect that her actions have on the follower (hence, she always knows \mathbf{L} for all activities and constraints known to her). We note that the assumption regarding the leader's certain knowledge of the values of \mathbf{F}^t can be relaxed, and most of the results can be extended to this more general setting, see Section 4.

Assumption **A2** states that the leader has a polyhedral uncertainty set for \mathbf{c}^t . Polyhedral sets capture many important classes of uncertainty for the data in \mathbf{c}^t such as lower and upper bounds, linear relationships between the entries, 1-norms, infinity norms, among others, see Ben-Tal et al. (2009). Assumption **A3** reflects the fact that the leader aims to optimally use her assets subject to budgetary constraints. (Note that this assumption holds for broad classes of standard max-min bilevel problems arising in interdiction, AD and DA models.) This follows due to our convention that the upper-level vectors in X are non-negative. Thus, by using resource $i \in I$ at level x_i , the leader consumes $H_{di}x_i$ units of asset d , $d \in C_L$, and the total amount of such asset available to her

at any given time is given by h_d . Finally, assumption **A4** is technical and is made to ensure that the follower's problem is not trivially infeasible.

Given the framework above, at any given time $t \in \mathcal{T}$ the following sequence of events takes place:

1. The leader chooses $x^t \in X^t$, where

$$X^t := \{x \in \mathbb{R}_+^{|I^t|} : \mathbf{H}^t x \leq \mathbf{h}^t, x \geq 0\}. \quad (2)$$

2. The follower solves the linear program $z(\bar{x}^t)$, where \bar{x}^t is defined as $\bar{x}_i^t := x_i^t$ if $i \in I^t$, and $\bar{x}_i^t := 0$ if $i \in I \setminus I^t$. That is, he solves

$$\begin{aligned} z(\bar{x}^t) = \min_{y \geq 0} \mathbf{c}^\top y \\ \text{s.t. } \mathbf{F}y + \sum_{i \in I^t} \mathbf{L}_i x_i^t \leq \mathbf{f}, \end{aligned} \quad (3)$$

where \mathbf{L}_i is the i -th column of L . For notational convenience, we set $y^t := y(\bar{x}^t)$ and $z^t := z(\bar{x}^t)$.

3. The response of the follower generates *feedback* \mathcal{F}^t . The leader observes the information in \mathcal{F}^t and exploits it to update her current knowledge to I^{t+1} , C_L^{t+1} , A^{t+1} , C_F^{t+1} and \mathcal{U}^{t+1} (thus, potentially updating \mathbf{H}^{t+1} , \mathbf{h}^{t+1} , \mathbf{F}^{t+1} , \mathbf{L}^{t+1} and \mathbf{f}^{t+1} as well as c_a for any new activity learned).

The next section elaborates on the information update in \mathcal{F}^t . Before that, we illustrate the assumptions above and the flexibility of the framework through the following examples.

Example 1. Consider a smuggling interdiction problem where a smuggler (the follower) operates over a directed network $G = (V, E)$ and is required to satisfy the demand for illegal goods across different locations. At each time period, the smuggler moves goods from supply vertices $V_S \subseteq V$ to demand vertices $V_D \subseteq V$. Some of the vertices are temporary depots (i.e., transshipment vertices) and we denote them by V_N . Denote by $b(v)$ the amount of goods that vertex v supplies/demands, where $b(v) > 0$ for $v \in V_S$ and $b(v) < 0$ for $v \in V_D$. We assume that $b(v) = 0$ for vertices in V_N , and that $\sum_{v \in V_S} b(v) = \sum_{v \in V_D} b(v)$.

For any $e = (v, w) \in E$, it costs the smuggler c_e to ship one unit of the illegal good from vertex v to vertex w through link e , and due to the transportation limitations (e.g., the fleet or infrastructure size) he can move at most u_e units from v to w at any given time. The smuggler's objective is to ship the goods across the network at each time period in order to minimize the shipment costs, subject to the requirement of supplying all demand.

Consider the follower's minimum cost flow problem over G . Let \mathbf{M} be the node-arc adjacency matrix of G , so \mathbf{M} is a $|V| \times |E|$ matrix, where for any $v \in V$, $M_{ve} = 1$ for all $e \in E$ such that $e = (v, w)$ for some $w \in V$, and $M_{ve} = -1$ for all $e \in E$ such that $e = (w, v)$ for some $w \in V$. Let \mathbf{b} be the vector given by $b_v = b(v)$ for all $v \in V$, \mathbf{c} and \mathbf{u} be the cost and upper-bound vectors, respectively.

For any $e \in E$ define y_e as the amount of goods the smuggler ships through edge e . Then, *without the leader's intervention* the follower would solve the min-cost flow problem of the form:

$$y^* \in \arg \min_y \{ \mathbf{c}^\top y : \mathbf{M}y \leq \mathbf{b}, -\mathbf{M}y \leq -\mathbf{b}, \mathbf{I}y \leq \mathbf{u}, y \in \mathbb{R}_+^{|E|} \},$$

where \mathbf{I} is a $|E| \times |E|$ identity matrix. Observe that y^* above can be thought as always taking only integer values (as long as \mathbf{u} and \mathbf{b} are integers) as the constraint matrix is totally unimodular, see Wolsey and Nemhauser (2014).

Law enforcement, on the other hand, acts as the leader. She assigns patrolling and interdicting vehicles to links in G . We assume that there are K types of vehicles capable of interdicting any edge. The leader controls r_k units of type k vehicles, each of which reduces the shipment capacity of arc e by d_{ke} units when assigned to the said arc. For any $k \in K$ and $e \in E$, define x_{ke} as the number of vehicles type k the leader sends to edge e . We assume that the values of x should satisfy the constraints $\sum_{e \in E} x_{ke} \leq r_k$ for all $k \in K$, $\sum_{k \in K} d_{ke} x_{ke} \leq u_e$ for all $e \in E$, and $x_{ke} \in \mathbb{Z}_+$ for all $k \in K$ and $e \in E$. Observe that this problem can be viewed as a generalization of the typical minimum cost flow interdiction problem (Chern and Lin 1995, Smith and Lim 2008).

We can model the setting above within our framework as follows. The set of the follower's activities corresponds to E (i.e., $A = E$). For each vertex there are two flow constraints and for each edge of E there is an upper bounding constraint. Thus, $|C_F| = 2|V| + |E|$. Matrix \mathbf{F} is given by $\mathbf{F} = (\mathbf{M}; -\mathbf{M}; \mathbf{I})$, the right-hand side vector is $\mathbf{f} = (\mathbf{b}; -\mathbf{b}; \mathbf{u})$, and the cost vector \mathbf{c} is precisely the cost vector of the network. On the other hand, we model the set of the leader's resources by $I = K \times E$, where each leader resource is represented by a vehicle type and an edge. Note that there is constraint associated with each vehicle type and each edge in the leader's problem, hence $|C_L| = |K| + |E|$. Henceforth, if the leader has all the information of the network, we have $\mathbf{H} = (\mathbf{O}; \mathbf{D})$, where \mathbf{O} is the $|K| \times |K||E|$ matrix given by $O_{k,(k,e)} = 1$ for all $e \in E$, and zero otherwise; and \mathbf{D} is the $|E| \times |K||E|$ matrix defined by $D_{e,(k,e)} = d_{ke}$ for all $k \in K$, and zero otherwise. Vector \mathbf{h} , on the other hand, is given by $\mathbf{h} = (\mathbf{r}; \mathbf{u})$, where $\mathbf{r} = (r_k : k \in K)$. Finally, observe that by the definition of the interdiction activities, matrix \mathbf{L} is given by $\mathbf{L} = (\mathbf{0}; \mathbf{D})$, where $\mathbf{0}$ is a matrix of zeros of size $2|V| \times |K||E|$ as we assume that the leader cannot interdict nodes.

Assume next that at time $t = 0$ the leader does not know all the edges nor all the nodes in G . For each vertex she observes, she knows whether it is supply or demand node, and knows with certainty the value of $b(v)$, while for each edge $e \in A^0$ she knows with certainty its shipment capacity u_e , however she does not know its shipment cost c_e . For each $e \in A^0$ she estimates the cost to be in the interval $[\ell_e, m_e]$, $\ell_e \leq m_e$, and hence $\mathcal{U}^0 = \{ \hat{\mathbf{c}}^0 \in \mathbb{R}^{|A^0|} : \ell_e \leq \hat{c}_e^0 \leq m_e \forall e \in A^0 \}$, so $\mathbf{G}^0 = [\mathbf{I}; -\mathbf{I}]$ and $\mathbf{g}^0 = (\mathbf{m}; \boldsymbol{\ell})$, with $\mathbf{m} = (m_e : e \in A^0)$ and $\boldsymbol{\ell} = (\ell_e : e \in A^0)$. \blacksquare

Example 2. We consider a simple class of the attacker-defender linear models, which can be viewed as an adversarial knapsack problem (DeNegre 2011, Caprara et al. 2013). The defender has $n > 0$ assets; operating asset a during a time period costs him b_a and produces a profit of p_a . He has an operational budget of B per period, and has to decide a level $y_a \in [0, 1]$ at which the operation of asset a is performed for all $a = 1, \dots, n$. Hence, at each period the follower would ideally solve the following knapsack problem absent the actions of the leader

$$y^* \in \arg \max_y \{ \mathbf{p}^\top y : \mathbf{b}^\top y \leq B, 0 \leq y_a \leq 1 \forall a = 1, \dots, n \},$$

where $\mathbf{p} := (p_a : a = 1, \dots, n)$ and $\mathbf{b} := (b_a : a = 1, \dots, n)$.

The attacker, on the other hand, can temporarily disable some of the defender's assets. Disabling asset a during any given period costs her r_a , and the attacker has a budget of R per period. Moreover, if an asset is disabled then the follower cannot operate it. In this setting, $A = I = \{1, \dots, n\}$, C_F consist of $n + 1$ constraints, and hence $\mathbf{F} = (\mathbf{b}^\top; \mathbf{I})$, where \mathbf{I} is a $n \times n$ identity matrix. Here, the lower-level right-hand side vector is given by $\mathbf{f} = (B; \mathbf{1})$ ($\mathbf{1}$ is a vector of ones of size n) and the cost vector satisfies $\mathbf{c} = -\mathbf{p}$. On the other hand, C_L is a singleton that contains the leader budgetary constraint, so $\mathbf{H} = \mathbf{r}^\top$, with $\mathbf{r} = (r_a : a \in I)$, and $\mathbf{h} = (R)$. Observe that matrix \mathbf{L} in this setting is given by $\mathbf{L} = (\mathbf{0}^\top; \mathbf{I})$ where $\mathbf{0}$ is a vector of zeros.

At time $t = 0$, we make the assumption that the attacker does not know all the assets operated by the defender, nor the corresponding profits. For those assets $A^0 \subseteq A$ she knows, she has interval estimates $\ell_e \leq c_e \leq m_e$ for the profits, which implies that $\mathcal{U}^0 = \{\hat{\mathbf{c}}^0 \in \mathbb{R}^{|A^0|} : \ell_e \leq \hat{c}_e \leq m_e \forall e \in A^0\}$. Thus, $\mathbf{G}^0 = [\mathbf{I}; -\mathbf{I}]$ and $\mathbf{g}^0 = (\mathbf{m}; \boldsymbol{\ell})$, with $\mathbf{m} = (m_e : e \in A^0)$ and $\boldsymbol{\ell} = (\ell_e : e \in A^0)$. ■

Example 3. See Section C.3 of the Appendix for an example in assignment interdiction. ■

2.1. Feedback

Depending on the particular application, the feedback $\mathcal{F} := (\mathcal{F}^t, t \in \mathcal{T})$ might include data from the follower's problem as well as from his response y^t , some information regarding the follower's activities and constraints that were unknown to the leader, as well as the leader's resources that were previously unavailable. In order to formalize these notions we introduce the following terminology:

DEFINITION 1. Let time $t \in \mathcal{T}$ be given and consider the bilevel problem (1).

- We say that the follower *performs* activity $a \in A$ (leader *uses* resource $i \in I$) at time t if and only if $y_a^t > 0$ ($x_i^t > 0$).
- We say that a lower-level (upper-level) constraint $d \in C_F$ ($d \in C_L$) *restricts* follower's activity $a \in A$ (leader's resource $i \in I$) if and only if $F_{da} \neq 0$ ($H_{di} \neq 0$), and we denote by $C_F(a)$ ($C_L(i)$) the set of constraints that restrict $a \in A$ ($i \in I$).

- We say that a leader resource $i \in I$ *interferes* with follower activity $a \in A$ if and only if there exist a lower-level constraint $d \in C_F$, such that $d \in C_F(a)$ and $L_{di} \neq 0$. We denote by $I(a)$ the set of all leader's activities that interfere with $a \in A$.

The first of the above definitions reflects the intuitive fact that if the follower's variable y_a takes the value 0 then it does not have incidence in his cost or constraints, and hence this can be interpreted as if activity $a \in A$ is not performed. The second definition is a consequence of the fact that if $F_{da} = 0$ for a given $a \in A$, then y_a can take arbitrarily large values without compromising the satisfiability of constraint d ; the remaining definitions are also inspired by the same observations.

Example 1 (continued). In this example, the follower performs activity $e \in A$ as long as he ships goods through edge e . Similarly, the leader uses resource $(k, e) \in I$ as long as she sends a vehicle type k to interdict edge e . Associated with activity (edge) $e = (v, w) \in A$ there are five constraints in $C_F(e)$. The first four constraints correspond to the supply/demand restrictions at v and w , while the additional constraint corresponds to the maximum shipment capacity constraint of edge e . Additionally, for any resource $(k, e) \in I$ we have that $C_L(k, e)$ consists of two constraints. One of them restricts the amount of vehicles type k that can be used across all edges (i.e., $\sum_{e \in E} x_{ke} \leq r_k$), and the other one corresponds to the maximum interdiction allowed across all vehicle types on edge e (i.e., $\sum_{k \in K} d_{ke} x_{ke} \leq u_e$). Finally, for each $e \in A$, we have $I(e) = \{(1, e), (2, e), \dots, (|K|, e)\}$, that is, $I(e)$ consists $|K|$ leader resources, one for each vehicle type that can interdict edge e . ■

Example 2 (continued). In the AD knapsack example, the follower performs activity $a \in A$ if he operates asset a . The leader uses resource $a \in A$ if she disables asset a (hence, $I = A$). For any $a \in A$, $C_F(a)$ consists of the defender's budget constraint and on the constraint $y_a \leq 1$. On the other hand, for any $a \in I$ it is clear that $C_L(a) = C_L$. Moreover, observe that in this setting, for any asset $a \in A$, we have that $I(a) = \{a\}$. ■

We are now in position to define a *standard feedback*:

DEFINITION 2. We say that feedback \mathcal{F} is *standard* if and only if for any $t \in \mathcal{T}$

S1: The leader observes the total cost z^t incurred by the follower.

S2: The leader observes the activities performed by the follower, that is, she can determine that the follower performed activity $a \in A$ at time t as long as $y_a^t > 0$. If $y_a^t > 0$ and $a \notin A^t$, the leader *learns* about the existence of $a \in A$, and of all the leader resources that can restrict $a \in A$. Therefore,

$$A^{t+1} = A^t \cup \bigcup_{a: y_a^t > 0} \{a\}, \quad I^{t+1} = I^t \cup \bigcup_{a: y_a^t > 0} I(a).$$

S3: For every new follower's activity $a \in A$ learned by the leader, she learns all the lower-level constraints in $C_F(a)$, and all the upper-level constraints $C_L(i)$, for all $i \in I(a)$. Henceforth,

$$C_F^{t+1} = C_F^t \cup \bigcup_{a \in A^{t+1} \setminus A^t} C_F(a), \quad C_L^{t+1} = C_L^t \cup \bigcup_{i \in I^{t+1} \setminus I^t} C_L(i).$$

S4: For any newly learned activity $a \in A$: the leader learns the value of F_{da} for all $d \in C_F(a) \cup C_F^t$; for any $i \in I(a) \cap I^t$ the leader learns the value of H_{di} for all $d \in C_L(i) \setminus C_L^t$ and the value of L_{di} for all $d \in C_F(a) \setminus C_F^t$; for any $i \in I(a) \setminus I^t$ the leader learns the value of H_{di} for all $d \in C_L(i) \cup C_L^t$ and the value of L_{di} for all $d \in C_F(a) \cup C_F^t$. Finally, for any $d \in C_F(a) \setminus C_F^t$ the leader learns the value of f_d , and for any $i \in I(a)$ the leader learns the value of h_d for all $d \in C_L(i) \setminus C_L^t$.

Hereafter, we make the assumption that the feedback is always standard and that the above conditions also hold for the initial information known by the leader before any interaction takes place (see also Section 2.2). Therefore, at any given time $t \in \mathcal{T}$ the matrices \mathbf{F} , \mathbf{L} and \mathbf{H} can be partitioned in submatrices as follows:

$$\mathbf{F} = \begin{matrix} & A^t & A \setminus A^t \\ C_F^t & \begin{pmatrix} \mathbf{F}_1 & \mathbf{F}_2 \end{pmatrix} \\ C_F \setminus C_F^t & \begin{pmatrix} \mathbf{0} & \mathbf{F}_3 \end{pmatrix} \end{matrix}, \quad \mathbf{L} = \begin{matrix} & I^t & I \setminus I^t \\ C_F^t & \begin{pmatrix} \mathbf{L}_1 & \mathbf{0} \end{pmatrix} \\ C_F \setminus C_F^t & \begin{pmatrix} \mathbf{L}_2 & \mathbf{L}_3 \end{pmatrix} \end{matrix}, \quad \mathbf{H} = \begin{matrix} & I^t & I \setminus I^t \\ C_L^t & \begin{pmatrix} \mathbf{H}_1 & \mathbf{H}_2 \end{pmatrix} \\ C_L \setminus C_L^t & \begin{pmatrix} \mathbf{0} & \mathbf{H}_3 \end{pmatrix} \end{matrix}, \quad (4)$$

and it is clear that, in the notation of the above structure, the leader is only aware of \mathbf{F}_1 , \mathbf{L}_1 and \mathbf{H}_1 at the beginning of time $t \in \mathcal{T}$. In particular, note that $\mathbf{F}^t = \mathbf{F}_1$, $\mathbf{L}^t = \mathbf{L}_1$, and $\mathbf{H}^t = \mathbf{H}_1$.

Assumption **S1** on the standard feedback is typical in the online optimization literature (Cesa-Bianchi and Lugosi 2006) and can be seen as a minimum requirement to perform any optimization analysis. The role of the other assumptions, namely, **S2-S4** is to determine what information the leader gains when a new activity is learned; specifically, these assumptions ensure that at any time t the leader has the *structural* information of a version of problem (1). That is: (i) the leader always observes all the constraints associated with the resources/activities she knows, and hence, if she ignores the existence of a constraint (lower-level or upper-level) then she must ignore the existence of all the resources/activities associated with it; (ii) the leader is always aware of all the resources in I that can restrict the follower's activities she knows, and hence, if the leader ignores a resource at any given time, then it must be that said resource cannot interfere with the follower's activities that she already knows.

It is important to note that our assumptions on standard feedback do not rule out the possibility that there might exist resources that the leader knows at time t that might restrict the follower's activities she does not know at time t . In this sense, some of the leader's feasible vectors at time t might 'involuntarily' restrict the follower's activities.

Example 1 (continued). Consider standard feedback in the smuggling example, which implies that the leader observes the total cost incurred by the smuggler at each period. In addition, if the smuggler ships goods through an edge $e = (v, w) \in A$ that the leader was not aware of, then the leader learns about the existence of that edge. Moreover, as she learns $C_F(e)$, she becomes aware

about the existence of vertices v and w as well as the value of the supply/demand b_v and b_w . She also learns the maximum shipment capacity u_e of edge e .

On the other hand, the leader also observes $I(e)$. Thus, she can now send vehicles to interdict edge e . Consequently, she learns about all the vehicle resources/capacity constraints in $C_L(k, e)$ for all $(k, e) \in I(e)$; and as she learns L_{di} for all $d \in C_F(a)$ and all $i \in I(a)$, then she also learns about the value of the capacity restrictions d_{ke} for all vehicle types $k \in K$. ■

Example 2 (continued). In the AD knapsack example, by assuming standard feedback, at each time t the leader observes the profit the follower receives from operating his assets. If the follower uses an asset unknown to the leader, then the leader learns about the existence of this asset, its cost b_a and the operating level upper bound. In addition, she discovers that she can disable the asset and that it costs her r_a to do so. ■

Observe that the assumptions on standard feedback impose no conditions on the *values* that are observed from the follower's response nor on the follower's cost vector. In this sense, stronger assumptions can be made in order to guarantee that the leader learns the follower's data in \mathbf{c} or his response y^t with more accuracy. In this paper we consider the following two cases:

DEFINITION 3. Let \mathcal{F} be standard. We say \mathcal{F} is:

- *Value-Perfect* if and only if at any time $t \in \mathcal{T}$ the leader learns the value of c_a for all $a \in A$ such that $y_a^t > 0$.
- *Response-Perfect* if and only if at any time $t \in \mathcal{T}$ the leader learns the value of y_a^t for all $a \in A$ such that $y_a^t > 0$.

Standard feedback, as well as its Value-Perfect feedback version, can be viewed as adaptations of similar notions in the online optimization literature. For example, suppose that $A = A^0$ (hence, the leader knows all the follower's activities at time $t = 0$). In this case, standard feedback only requires the leader to observe the value of z^t at each $t \in \mathcal{T}$, and thus it parallels to the notion of *bandit feedback* that appears in the online convex and combinatorial optimization (see e.g., Bubeck and Cesa-Bianchi (2012), Hazan (2015) and the references therein). Similarly, Value-Perfect feedback parallels the notion of *semi-bandit feedback* in the online combinatorial optimization (Audibert et al. 2013).

Example 1 (continued). In the smuggling setting, Value-Perfect feedback means that if at period $t \in \mathcal{T}$ the smuggler ships goods through edge e , then the leader learns c_e , the cost of shipping one unit of the illegal good through e . On the other hand, Response-Perfect feedback means that the leader observes y_e^t , the amount of goods shipped by the smuggler through e at time $t \in \mathcal{T}$. ■

Example 2 (continued). In the AD knapsack setting, under Value-Perfect feedback, at each period the leader observes the follower's profit from the assets operated during the period. Under Response-Perfect feedback, she observes the corresponding values of y 's. ■

2.2. Optimality Criteria

In this section we define what constitutes a ‘good’ decision-making policy for the leader. In contrast with most work in online optimization, we measure the performance of a policy in terms of its *time-stability* rather than of its regret. The time-stability of a policy corresponds to the first time period by which the actions prescribed by the policy coincide with the actions of an oracle decision-maker. Recall that the oracle has all the information about the problem and thus, implements the best possible decision starting at time $t = 0$. As it will be seen below, any upper bound on time-stability implies an upper bound on regret.

To formally introduce time-stability and the concept of optimality, we first define what we consider a problem’s instance for the leader. The *initial information* of the problem is the collection \mathcal{D}^0 , where

$$\mathcal{D}^0 := (A^0, I^0, C_F^0, C_L^0, \mathcal{U}^0, \mathbf{H}^0, \mathbf{h}^0, \mathbf{F}^0, \mathbf{L}^0, \mathbf{f}^0).$$

Note that given some initial information \mathcal{D}^0 , there might be several different bilevel problems of the form (1) that agree with the information contained in \mathcal{D}^0 . In view of this, we define $\mathbb{G}(\mathcal{D}^0)$ to be the collection that contains all possible bilevel problems given that the leader knows \mathcal{D}^0 :

$$\mathbb{G}(\mathcal{D}^0) := \{(A, I, C_F, C_L, \mathbf{c}, \mathbf{H}, \mathbf{h}, \mathbf{F}, \mathbf{L}, \mathbf{f}) : \text{conditions C1-C5 below are satisfied}\},$$

where

C1: $A^0 \subseteq A, I^0 \subseteq I, C_F^0 \subseteq C_F, C_L^0 \subseteq C_L$.

C2: $I^0 = \cup_{a \in A^0} I(a), C_L^0 = \cup_{i \in I^0} C_L(i), C_F^0 = \cup_{a \in A^0} C_F(a)$.

C3: \mathcal{U}^0 has valid upper and lower bounds for all $c_a, a \in A^0$.

C4: $(c_a : a \in A^0) \in \mathcal{U}^0$.

C5: $\mathbf{H}^0, \mathbf{h}^0, \mathbf{F}^0, \mathbf{L}^0, \mathbf{f}^0$, are submatrices of $\mathbf{H}, \mathbf{h}, \mathbf{F}, \mathbf{L}, \mathbf{f}$.

Note that conditions **C2-C3** state that the information that the leader initially knows satisfies the standard feedback conditions at time $t = 0$. Using collection $\mathbb{G}(\mathcal{D}^0)$, we define an *instance* of the problem as a pair $(\mathcal{D}^0, \mathcal{D})$, where $\mathcal{D} \in \mathbb{G}(\mathcal{D}^0)$. We denote by \mathbb{G} the set of all possible instances.

A decision-making *policy* π is a sequence of set functions $\pi = (\pi^1, \dots, \pi^T)$, such that $x^t = \pi^t(\mathcal{H}^t(\mathcal{D}^0, \mathcal{D}))$, and $\mathcal{H}^t(\mathcal{D}^0, \mathcal{D})$ denotes the history of both the leader and follower decision-making process up to time t :

$$\mathcal{H}^t(\mathcal{D}^0, \mathcal{D}) := (\mathcal{D}^0, x^0, \mathcal{F}^0, \dots, x^{t-1}, \mathcal{F}^{t-1}), \quad t \geq 1.$$

The set of all policies is denoted by Π . When discussing a particular policy π , we include a superscript π on x^t and in all other other quantities depending on it, and denote them by $x^{t,\pi}, y^{t,\pi}, z^{t,\pi}, I^{t,\pi}, A^{t,\pi}, \mathcal{U}^{t,\pi}$ and $\mathcal{F}^{t,\pi}$.

Let an instance $(\mathcal{D}^0, \mathcal{D})$ be given. We define the *time-stability* of a policy on $(\mathcal{D}^0, \mathcal{D})$, denoted by $\tau^\pi(\mathcal{D}^0, \mathcal{D})$, as the first time in \mathcal{T} such that the z^* is equal to $z^{t,\pi}$ from there on, i.e.,

$$\tau^\pi(\mathcal{D}^0, \mathcal{D}) := \min\{t \in \mathcal{T} : z^{s,\pi} = z^* \text{ for all } s \geq t\}.$$

There is a clear connection between time-stability and regret. Indeed, the *regret* $R_{T_0}^\pi(\mathcal{D}^0, \mathcal{D})$ of policy π on the pair $(\mathcal{D}^0, \mathcal{D})$ until time $T_0 \geq 0$ is defined as

$$R_{T_0}^\pi(\mathcal{D}^0, \mathcal{D}) := \sum_{0 \leq t \leq T_0} (z^* - c^\top y^{t,\pi}).$$

If U is an upper bound on the value of $(z^* - c^\top y^{t,\pi})$ for any $t \in \mathcal{T}$, then it immediately follows that

$$R_{T_0}^\pi(\mathcal{D}^0, \mathcal{D}) \leq U \cdot \tau^\pi(\mathcal{D}^0, \mathcal{D}),$$

for any $T_0 \leq T$ as long as the time-stability is finite. Consequently, any finite upper bound on the time-stability provides a finite upper bound on the regret.

The leader would like to find an “optimal” time-stability policy, i.e., a policy that has a lower time-stability than any other policy across all instances. To this end, let us say that policy π is *absolutely better* than policy π' if and only if $\tau^\pi(\mathcal{D}^0, \mathcal{D}) \leq \tau^{\pi'}(\mathcal{D}^0, \mathcal{D})$ for any instance $(\mathcal{D}^0, \mathcal{D})$, and that π^* is *absolutely optimal* if it is absolutely better than any other policy. Unfortunately, absolute optimality is a very strong notion, and, in general, absolute optimal policies do not exist, see e.g., Remark 1 in Borrero et al. (2016) for the sequential shortest-path interdiction problem with incomplete information, which can be viewed as an example in our general setting.

Henceforth, we study an alternative optimality notion referred to as *weak optimality*. Roughly speaking, π is weakly better than π' if the worst-case time-stability of π across all possible instances is at most the worst-case time-stability of π' across all possible instances, that is, if:

$$\sup_{(\mathcal{D}^0, \mathcal{D}) \in \mathbb{G}} \tau^\pi(\mathcal{D}^0, \mathcal{D}) \leq \sup_{(\mathcal{D}^0, \mathcal{D}) \in \mathbb{G}} \tau^{\pi'}(\mathcal{D}^0, \mathcal{D}). \quad (5)$$

A policy π would be weakly optimal if it is weakly better than any other policy. It turns out, however, that the above definition is not meaningful as the suprema in (5) are infinity. Certainly, for any policy it can be readily checked that there are instances where the time-stability increases linearly with $|A|$, see e.g., Proposition 2 in Section 3.2.

In order to address this issue, we take the suprema in equation (5) over instances of a fixed size, which we assume is given in terms of the follower’s problem. Specifically, we define the *size* of an instance $(\mathcal{D}^0, \mathcal{D})$ as the vector $(|A|, |A^0|)$, and define \mathbb{G}_s as the collection of instances of size $s = (n, n^0)$ (with $n \geq n^0$):

$$\mathbb{G}_s := \{(\mathcal{D}^0, \mathcal{D}) \in \mathbb{G} : (|A|, |A^0|) = s\}.$$

Observe that any direct information on \mathcal{U}^0 in the definition of \mathbf{s} is not included. This follows as, from the worst-case analysis perspective, any reasonable notion of size of \mathcal{U}^0 is likely to be a function of n^0 . Given the above considerations, we say that policy π is weakly better than π' if

$$\max_{(\mathcal{D}^0, \mathcal{D}) \in \mathbb{G}_{\mathbf{s}}} \tau^{\pi}(\mathcal{D}^0, \mathcal{D}) \leq \max_{(\mathcal{D}^0, \mathcal{D}) \in \mathbb{G}_{\mathbf{s}}} \tau^{\pi'}(\mathcal{D}^0, \mathcal{D}) \quad \text{for all } \mathbf{s} \in S,$$

where $S := \{(n, n^0) \in \mathbb{Z}_+^2 : n \geq n^0\}$. We say that π^* is *weakly optimal* if it is weakly better than any other policy, that is, if

$$\pi^* \in \arg \min_{\pi \in \Pi} \max_{(\mathcal{D}^0, \mathcal{D}) \in \mathbb{G}_{\mathbf{s}}} \tau^{\pi}(\mathcal{D}^0, \mathcal{D}) \quad \text{for all } \mathbf{s} \in S.$$

It should be clear that the notion of weak optimality is an adaptation of the notion on min/max optimal policies used in the online optimization literature, specifically, in the multi-armed bandit settings, see Audibert and Bubeck (2009), Audibert et al. (2013).

3. Greedy and Robust Policies

In this section we introduce a set of leader's policies Λ that are greedy and robust. These policies are *greedy* in the sense that at each $t \in \mathcal{T}$ they aim to maximize the immediate cost that the follower faces at time t , and *robust* in the sense that they exploit the cost information in \mathcal{U}^t in a worst-case scenario approach. Under the Value-Perfect and Response-Perfect conditions on the feedback \mathcal{F} , we show that these policies' time-stability are upper-bounded by $|A|$, and moreover, that they are weakly optimal. Also, we show that these policies also have additional features, such as that they can identify the value of time-stability in real time, yielding a *certificate of optimality*. Note that the proposed policies can be viewed, in a sense, as natural generalizations of known results for the shortest-path network interdiction problem, see Borrero et al. (2016). Throughout this section we omit any dependence on the instance $(\mathcal{D}^0, \mathcal{D})$ unless necessary to avoid confusion.

3.1. General Results for Standard Feedback

In order to define the set of greedy and robust policies, Λ , some additional concepts have to be introduced. For any $t \in \mathcal{T}$, and given any $x \in X^t$, define region $Y^t(x)$ as

$$Y^t(x) := \left\{ y \in \mathbb{R}_+^{|A^t|} : \mathbf{F}^t y + \mathbf{L}^t x \leq \mathbf{f}^t \right\}.$$

Observe that, in contrast to $Y(x)$, the leader completely knows $Y^t(x)$ at time t . More importantly, $Y^t(x)$ can be considered as the “best guess” the leader makes about the follower's feasible region, given that she decides x . For any $x \in X^t$, define $z_R^t(x)$ as the value of the robust linear program

$$z_R^t(x) := \min_y \left\{ \max_{\hat{\mathbf{c}}^t \in \mathcal{U}^t} \{ (\hat{\mathbf{c}}^t)^\top y \} : y \in Y^t(x) \right\}.$$

Note that $z_R^t(x)$ is the cost that the leader expects the follower would incur if she chooses vector x and if the follower's worst-case scenario over \mathcal{U}^t is realized. Let $z_R^{t,*}$ be the value that corresponds to the best possible decision the leader can take at time t if she estimates the follower's response using the robust approach above, that is,

$$z_R^{t,*} := \max\{z_R^t(x) : x \in X^t\} \quad \forall t \in \mathcal{T}.$$

Finally, for any policy π define $\xi^\pi := \xi^\pi(\mathcal{D}^0, \mathcal{D})$ as,

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

We define policies in Λ as those policies that greedily optimize in a robust fashion from time $t = 0$ until time ξ^λ . From ξ^λ onwards, policies in Λ repeat the same solution used at time ξ^λ . Formally:

DEFINITION 4. We say that $\lambda \in \Lambda \subseteq \Pi$ if and only if

$$x^{t,\lambda} \in \arg \max\{z_R^t(x) : x \in X^t\} \quad \forall t \leq \xi^\lambda, \quad (6)$$

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

It is important to note that policies in Λ can be computed by standard mixed integer programming (MIP) solvers as robust bilevel problem (6) can be reduced to a single-level MIP, see Appendix C.2 for further details.

The following result lists the main properties of the policies in Λ under the assumption of standard feedback. It establishes a simple relationship between the cost of the optimal oracle solution (z^*), the cost the follower faces at t ($z^{t,\lambda}$), and the cost the leader expects the follower faces if she decides using policy λ ($z_R^{t,*}$). In addition, it reveals the importance that time period ξ^λ has for time-stability.

THEOREM 1. *Let $t \in \mathcal{T}$ be given and let $\lambda \in \Lambda$ be arbitrary. Then, $z^{t,\lambda} \leq z^* \leq z_R^{t,*}$ and $\tau^\lambda \leq \xi^\lambda$.*

Theorem 1 has important practical implications. Note that the leader is always aware of the value of $z_R^{t,*}$, and (by standard feedback) always observes the value of $z^{t,\lambda}$. Therefore, she can determine whether a given period t is equal to ξ^λ . Let $t \in \mathcal{T}$ be given such that $t - 1 < \xi^\lambda$, then at time t exactly one of the following scenarios may occur:

(i) The follower faces the cost the leader expected ($z^{t,\lambda} = z_R^{t,*}$). In this case, $t = \xi^\lambda$, and Theorem 1 implies that the solution implemented by the leader at time t is an optimal solution of the full-information problem.

(ii) The follower faces a cost less than that the leader expects ($z^{t,\lambda} < z_R^{t,*}$). In this case nothing can be said in general by only assuming standard feedback. However, if the stronger notions of either Value-Perfect or Response-Perfect feedback are assumed, it is shown in the following sections that the leader must learn new information of the follower's problem.

Particularly, observation (i) implies that policies in Λ provide certificates of optimality in real-time. That is, as soon as $t = \xi^\lambda$, the leader is sure that the best possible solution has been found. Given the importance of ξ^λ for greedy and robust policies, next we derive a sufficient condition in terms of the uncertainty set \mathcal{U}^t that establishes whether a given time $t \in \mathcal{T}$ corresponds to ξ^λ . The condition is given in terms of the polyhedral dimension $\dim(\mathcal{U}^t)$ of \mathcal{U}^t , which is the maximum number of affine independent points within \mathcal{U}^t . In particular, if $\dim(\mathcal{U}^t) = 0$, then it consists only of one point. That is, if $\dim(\mathcal{U}^t) = 0$, then $\mathcal{U}^t = \{c^t\}$.

PROPOSITION 1. *Suppose $t \in \mathcal{T}$ satisfies that $\dim(\mathcal{U}^t) = 0$ and assume that $y_a^t = 0$ for all $a \notin A^t$. Then $\xi^\lambda \leq t$, and, in particular, $\tau^\lambda \leq t$.*

In other words, whenever there is no uncertainty in \mathcal{U}^t , if the leader decides by using a policy in Λ , and the follower does not reveal any new activity, then the leader can be sure that the best solution has been found. We use this result in the following sections to establish upper bounds on ξ^λ (and hence, on τ^λ) under Value-Perfect and Response-Perfect feedbacks.

3.2. Policies in Λ Under Value-Perfect Feedback

Recall that feedback \mathcal{F} is Value-Perfect if the leader observes the value of c_a for all activities $a \in A$ such that $y_a^t > 0$. Under this feedback the leader should update the uncertainty set \mathcal{U}^t to \mathcal{U}^{t+1} as

$$\mathcal{U}^{t+1} = \{\hat{c} \in \mathbb{R}^{|A^{t+1}|} : (\hat{c}_a)_{a \in A^t} \in \mathcal{U}^t, \hat{c}_a = c_a \text{ for all } a \text{ s.t. } y_a^t > 0\}.$$

For convenience we partition A^t as $A^t = \tilde{A}^t \cup \bar{A}^t$, where for any follower action $a \in \tilde{A}^t$ the leader knows with certainty the value of c_a , that is

$$\tilde{A}^t := \{a \in A^t : \hat{c}_a = c_a \forall \hat{c} \in \mathcal{U}^t\},$$

and $\bar{A}^t := A^t \setminus \tilde{A}^t$. The next lemma establishes that if the cost the follower incurs is different from the one expected by the leader, then the leader must learn the real cost of a follower's activity.

LEMMA 1. *Suppose $\lambda \in \Lambda$ and that feedback \mathcal{F} is Value-Perfect. If $z^{t,\lambda} < z_R^{t,*}$ then $\tilde{A}^{t+1} \setminus \tilde{A}^t \neq \emptyset$. In particular, if $y_a^t = 0$ for all $a \notin A^t$, then $\dim(\mathcal{U}^{t+1}) < \dim(\mathcal{U}^t)$.*

A direct consequence of the above result is that, in conjunction with Proposition 1, it provides an upper bound for the time-stability for any policy in Λ :

THEOREM 2. *Let $\lambda \in \Lambda$ and suppose that \mathcal{F} is Value-Perfect. Then,*

$$\tau^\lambda \leq \xi^\lambda \leq |A \setminus \tilde{A}^0|.$$

Proof. Let $t \in \mathcal{T}$ be given such that $z^{t,\lambda} < z_R^{t,*}$. Lemma 1 implies that $\tilde{A}^{t+1} \setminus \tilde{A}^t \neq \emptyset$. Hence, $\tilde{A}^t \neq A$ can happen at most for $|A \setminus \tilde{A}^0|$ periods. Also, if $t \in \mathcal{T}$ satisfies $\tilde{A}^t = A$, then $\dim(\mathcal{U}^t) = 0$ and Proposition 1 implies that $\xi^\lambda \leq t$. Therefore, $\xi^\lambda \leq |A \setminus \tilde{A}^0|$ and the result follows. \blacksquare

The previous results shed light into the importance of greedy and robust policies for solving the exploitation vs. exploration dilemma. Simply speaking, it states that as long as the leader is being robust with respect to uncertainty, then exploitation (i.e., deciding greedily) always implies exploration (i.e., discovering new information). We emphasize that the key is robustness, as if the leader uses another approach to deal with uncertainty she might not discover any new information, see Remark 7 in Borrero et al. (2016) for an example in the context of shortest path interdiction.

Next, we prove that the upper bound in Theorem 2 is tight across all instances and, more importantly, across all policies. In other words, we establish that policies in Λ are weakly optimal.

PROPOSITION 2. *Consider $\lambda \in \Lambda$ and suppose that \mathcal{F} is Value-Perfect. Then, for any $\mathbf{s} = (n, n^0) \in S$*

$$\max_{(\mathcal{D}^0, \mathcal{D}) \in \mathbb{G}_{\mathbf{s}}} \tau^\lambda(\mathcal{D}^0, \mathcal{D}) \leq n. \quad (7)$$

Moreover, λ is weakly optimal.

Proof. First, observe that equation (7) is an immediate consequence of Theorem 2. In order to prove weak optimality, we show that for any given policy π and any $\mathbf{s} = (n, n^0) \in S$ there exists an instance $(\mathcal{D}^0, \mathcal{D})^\pi$ of size \mathbf{s} such that $\tau^\pi((\mathcal{D}^0, \mathcal{D})^\pi) \geq n$.

Let $A = \{1, 2, \dots, n^0, n^0 + 1, \dots, n\}$, $A^0 = \{1, \dots, n^0\}$ and $I = A$, $I^0 = A^0$. Let X (and hence \mathbf{H} and \mathbf{h}) be given by

$$X = \left\{ x \in \mathbb{Z}_+^n : \sum_{j \in I^0} x_j = n^0 - 1, \sum_{j \in I} x_j \leq n - 1, x_j \leq 1 \ \forall j = 1, \dots, n \right\},$$

and let X^0 (and hence, \mathbf{H}^0 and \mathbf{h}^0) be given by

$$X^0 = \left\{ x \in \mathbb{Z}_+^n : \sum_{j \in I^0} x_j = n^0 - 1, x_j \leq 1 \ \forall j = 1, \dots, n^0 \right\}.$$

On the other hand, for any $x \in X$ define $Y(x)$ as

$$Y(x) := \left\{ y \in \mathbb{R}_+^n : \sum_{j=1}^n y_j \leq 1, y_j + x_j \leq 1 \ \forall j = 1, \dots, n \right\}.$$

That is, $\mathbf{F} = [\mathbf{1}^\top; \mathbf{I}]$ and $\mathbf{L} = [\mathbf{0}^\top; \mathbf{I}]$, where \mathbf{I} is an identity of size n , and \mathbf{f} is a column vector of ones. Define \mathbf{F}^0 , \mathbf{L}^0 and \mathbf{f}^0 as the corresponding submatrices of \mathbf{F} , \mathbf{L} and \mathbf{f} associated with $j = 1, \dots, n^0$. Finally, consider \mathbf{c} to be such that $c_{n^0+q} < c_{n^0+q+1}$, for $q = 1, \dots, n - n^0 - 1$, and for the cost coefficients of the first n^0 activities we assume that the leader knows that they belong to \mathcal{U}^0 , where $\mathcal{U}^0 = \{\hat{\mathbf{c}}^0 \in \mathbb{R}^{n^0} : \ell \leq \hat{c}_j^0 \leq u, j = 1, \dots, n^0\}$, where in addition we assume that $c_n < \ell < u < 0$.

In order to adequately define the instance, a particular $\hat{\mathbf{c}}^0$ in \mathcal{U}^0 has to be fixed. However, independent of which specific $\hat{\mathbf{c}}^0$ is chosen (which will depend on the policy, see below), the above defined data constitutes an instance, i.e., $\mathcal{D}^\pi \in \mathbb{G}((\mathcal{D}^0)^\pi)$, and its size is given by (n, n^0) . Particularly, note that from the leader perspective, the problem consist of blocking those $n - 1$ activities that are most profitable to the follower, constrained to the fact she always need to block exactly $n^0 - 1$ out of the n^0 activities she knows at time $t = 0$. In addition, from the assumptions on \mathbf{c} , the follower's profit from any of the $n - n^0$ activities that the leader does not initially know is better than the profit generated by any activity that the leader initially knows.

From the definition of $(\mathcal{D}^0, \mathcal{D})$ it is clear that if x^* is an optimal oracle decision, then $x_j^* = 1$ for $j = n^0 + 1, \dots, n$, which implies that the leader must learn all those activities before implementing a solution where $z^{t,\pi} = z^*$. Hence, if t_0 denotes the first time after which the leader learns all activities from $A \setminus A^0$, it is clear from the structure of the instance that $t_0 \geq n - n^0$. In addition, note that until t_0 the follower has only used activities in $A \setminus A^0$, so by Value-Perfect feedback, he has not revealed to the leader any of the real costs of the activities in A^0 .

In order to prove weak optimality we show that for any given policy π there is a cost vector $\mathbf{c}^0 \in \mathcal{U}^0$ such that it takes the leader at least another n^0 time periods to consistently implement x^* (this would imply that $\tau^\pi((\mathcal{D}^0, \mathcal{D})^\pi) \geq n$, yielding the desired result). First, assume that π does not repeat any solution from time t_0 , until time $t_n = t_0 + n^0 - 1$. For any $t = t_0, \dots, t_n$, let $j^{\pi,t}$ be the (unique) follower activity in A^0 that $x^{t,\pi}$ does not block at time t , and choose the values of c_1, \dots, c_{n^0} such that

$$\ell < c_{j^{\pi,t_0+1}} < c_{j^{\pi,t_0+2}} < \dots < c_{j^{\pi,t_n}} < c_{j^{\pi,t_0}} < u,$$

and note that the above defined values are admitted by \mathcal{U}^0 . Observe that fixing the costs of the actions in A^0 in this way, we have that x^* satisfies $x_j^* = 1$, for $j \neq j^{\pi,t_0}$ and $x_{j^{\pi,t_0}}^* = 0$, and that $z^* = c_{j^{\pi,t_0}}$. On the other hand, for $t = t_0 + 1, \dots, t_n$,

$$z^{t,\pi} \leq c_{j^{\pi,t}} < z^* \tag{8}$$

(we note the first inequality above is, in general, not an equality, as it is not necessary for $x^{t,\pi}$ to block all the activities j with $j > n^0$). Henceforth, equation (8) implies that $\tau^\pi((\mathcal{D}^0, \mathcal{D})^\pi) > t_n$, and hence, as $t_0 \geq n - n^0$, $\tau^\pi((\mathcal{D}^0, \mathcal{D})^\pi) \geq n$, and the result follows.

Now, suppose that π repeats a solution once between t_0 and t_n , i.e., there exist $t_0 \leq u < v \leq t_n$ such that $x^{u,\pi} = x^{v,\pi}$. In this case $j^{\pi,u} = j^{\pi,v}$, and there exist $1 \leq b \leq n^0$ such that $b \neq j^{t,\pi}$ for all $t = 0, \dots, n$. Let \mathbf{c}^0 satisfy

$$\ell < c_{j^{\pi,t}} < c_{j^{\pi,t+1}} \quad t = t_0, \dots, v-2, \quad c_{j^{\pi,t}} < c_{j^{\pi,t+1}} \quad t = v+1, \dots, t_n,$$

and assume that $c_{j^\pi, t_n} < c_b < u$. Observe that the above defined \mathbf{c}^0 belongs to \mathcal{U}^0 , and hence $(\mathcal{D}^0, \mathcal{D})^\pi$ is a valid instance, and moreover, x^* is given by $x_j^* = 1$ for all $j \neq b$, $x_b^* = 0$, with $z^* = c_b$. In addition, it is seen that for $t = t_0, \dots, t_n$

$$z^{t, \pi} \leq c_{j^\pi, t} < z^*,$$

and hence $\tau^\pi((\mathcal{D}^0, \mathcal{D})^\pi) \geq n$, as desired. Also, note that if π repeats a solution more than once between t_0 and t_n , then the same argument as above yields the result. \blacksquare

3.3. Policies in Λ Under Response-Perfect Feedback

Next, we establish convergence and weak optimality under Response-Perfect feedback. Recall that under this feedback the leader always observe the value of y_a^t for all $a \in A$ such that $y_a^t > 0$. In this setting, the leader should update the uncertainty set \mathcal{U}^t to \mathcal{U}^{t+1} by including the linear equality

$$\sum_{a \in A^{t+1}} y_a^{t, \lambda} \hat{c}_a = z^{t, \lambda}.$$

That is,

$$\mathcal{U}^{t+1} = \left\{ \hat{\mathbf{c}} \in \mathbb{R}^{|A^{t+1}|} : (\hat{c}_a)_{a \in A^t} \in \mathcal{U}^t, \sum_{a \in A^{t+1}} y_a^{t, \lambda} \hat{c}_a = z^{t, \lambda} \right\}. \quad (9)$$

Observe that if $A^{t+1} = A^t$, i.e., if the leader does not learn any new activity at time t , then \mathcal{U}^{t+1} has the same number of variables as \mathcal{U}^t , and moreover, equation (9) implies that $\mathcal{U}^{t+1} \subseteq \mathcal{U}^t$.

In Response-Perfect feedback, as in the Value-Perfect setting, by using a policy in Λ the follower must be forced to reveal new information whenever $z^{t, \lambda} < z_R^{t, *}$. Specifically, if $y_a^t = 0$ for all $a \notin A^t$, then it must be the case that $\dim(\mathcal{U}^{t+1}) < \dim(\mathcal{U}^t)$. This inequality follows because in this case $\dim(\mathcal{U}^t)$ cannot increase (since $\mathcal{U}^{t+1} \subseteq \mathcal{U}^t$), and, more importantly, from the fact that the linear equality $\sum_{a \in A^{t+1}} y_a^{t, \lambda} \hat{c}_a = z^{t, \lambda}$ is linearly independent from all the linear equalities in \mathcal{U}^t . These observations are formalized in the following result, which can be considered analogous to Lemma 1:

LEMMA 2. *Let $\lambda \in \Lambda$ and suppose feedback \mathcal{F} is Response-Perfect. If $z^{t, \lambda} < z_R^{t, *}$ and $y_a^t = 0$ for all $a \notin A^t$ then*

$$\dim(\mathcal{U}^{t+1}) < \dim(\mathcal{U}^t).$$

On the other hand, if the leader learns new activities at t , then \mathcal{U}^{t+1} has $|A^{t+1} \setminus A^t|$ more variables than \mathcal{U}^t . The addition of the corresponding new variables potentially increases the dimension of \mathcal{U}^{t+1} with respect to \mathcal{U}^t by $|A^{t+1} \setminus A^t|$. However, it is readily seen that the linear equality $\sum_{a \in A^{t+1}} y_a^{t, \lambda} \hat{c}_a = z^{t, \lambda}$ is trivially linearly independent of previous inequalities in \mathcal{U}^t , and as such if the leader learns new activities at t it can be concluded that $\dim(\mathcal{U}^{t+1}) \leq \dim(\mathcal{U}^t) + |A^{t+1} \setminus A^t| - 1$. This observation, in conjunction with Lemma 2 immediately provides the following upper bound:

THEOREM 3. *Let $\lambda \in \Lambda$ be given. Then, under Response-Perfect feedback,*

$$\tau^\lambda \leq \xi^\lambda \leq \dim(\mathcal{U}^0) + |A \setminus A^0|.$$

The above results, as in the case of Value-Perfect feedback, have the same implications regarding the exploitation vs. exploration dilemma. That is, exploitation always implies exploration as long as the leader decides robustly. In addition, for Response-Perfect feedback weak optimality also holds. The proof of this fact applies the same arguments as in Proposition 2. Thus, its proof is omitted.

PROPOSITION 3. *Let $\lambda \in \Lambda$ be given and suppose that \mathcal{F} is Response-Perfect. Then, for any $\mathbf{s} \in S$*

$$\max_{(\mathcal{D}^0, \mathcal{D}) \in \mathbb{G}_s} \tau^\lambda(\mathcal{D}^0, \mathcal{D}) \leq n.$$

Moreover, λ is weakly optimal.

4. Model for Matrix Uncertainty

In this section we consider a more general model referred to as the *matrix model* for the uncertainty of the leader regarding the data of the follower's problem. We assume that she knows with certainty the value of \mathbf{c}^t at the beginning of time t , but that she does not know with certainty the values of matrix \mathbf{F}^t . We emphasize the generality of this model: if in a given problem \mathbf{c}^t is uncertain as well, then it can be included in \mathbf{F}^t w.l.o.g., see Remark 1 below. In this setup, and under the appropriate extensions of certain assumptions and feedback definitions, we show that the results for standard feedback for the basic model of Section 2 (which, in view of the current discussion, can be referred to as simply the *cost model*) are also valid. Moreover, we show that for the Value-Perfect feedback case, the time-stability upper bound of Theorem 2 also holds, while for Response-Perfect feedback, an extension of the upper bound in Theorem 3 holds under certain assumptions.

REMARK 1. Consider the case where there is uncertainty regarding the cost function c . In this case the problem $\min\{\mathbf{c}^\top \mathbf{y} : \mathbf{y} \in Y(x)\}$ can be equivalently posed as $\min\{y_0 : (y_0, \mathbf{y}) \in Y'(x)\}$ where

$$Y'(x) := \{(y_0, \mathbf{y}) \in \mathbb{R}^{|A|+1} : \mathbf{F}'(y_0; \mathbf{y}) + \mathbf{L}'x \leq (0; \mathbf{f})\},$$

$$\mathbf{F}' := \begin{pmatrix} -1 & \mathbf{c}^\top \\ \mathbf{0} & \mathbf{F} \end{pmatrix}, \quad \mathbf{L}' := \begin{pmatrix} \mathbf{0}^\top \\ \mathbf{L} \end{pmatrix},$$

and in each case $\mathbf{0}$ is a column vector of zeros of appropriate dimensions. Observe that this new formulation has an additional variable but there is no uncertainty regarding any cost coefficient (it is always one for the new variable and zero for the rest).

4.1. Assumptions and Feedback in the Matrix Model

In this model we assume that the leader knows \mathbf{c}^t with certainty, but only knows that \mathbf{F}^t belongs to an uncertainty set \mathcal{U}^t . For any $d \in C_F^t$ let us denote by n_d^t the number of the follower's activities in A^t that d restricts, that is

$$n_d^t := |\{a \in A^t : d \in C_F(a)\}|.$$

We replace assumption **A2** from Section 2 with the following:

A2E: The leader does not know with certainty all entries of \mathbf{F} but she knows that $\mathbf{F}^t \in \mathcal{U}^t$, with

$$\mathcal{U}^t = \{\hat{\mathbf{F}}^t \in \mathbb{R}^{\sum_{d \in C_F^t} n_d^t} : \mathbf{G}^t \hat{\mathbf{F}}^t \leq \mathbf{g}^t\},$$

where we make the convention that

$$\hat{\mathbf{F}}^t = (F_{11}, \dots, F_{1n_1^t}, F_{21}, \dots, F_{2n_2^t}, \dots, F_{|C_F^t|1}, \dots, F_{|C_F^t|n_{|C_F^t|}^t})^\top.$$

If C_U^t is the set of constraints of polyhedron \mathcal{U}^t , then $\mathbf{G}^t \in \mathbb{R}^{|C_U^t| \times \sum_{d \in C_F^t} n_d^t}$ and $\mathbf{g}^t \in \mathbb{R}^{|C_U^t|}$. We assume that both \mathbf{G}^t and \mathbf{g}^t are known by the leader at time t .

We also modify the definition of standard feedback; specifically we replace **S4** by **S4E**:

S4E: For any new learned activity $a \in A$, the leader learns the value of c_a (instead of learning the value of F_{da} for all $d \in C_F(a) \cup C_F^t$). The rest of the assumption is as S4.

Moreover, in this setting Value-Perfect feedback is extended to account for the values of the constraint matrix. That is, we refine the concept of Value-Perfect feedback as follows

DEFINITION 5. In the context of the matrix model, standard feedback \mathcal{F} is called *Value-Perfect* if and only if at any time $t \in \mathcal{T}$ the leader learns the value of F_{da} for all a such that $y_a^t > 0$ and $d \in C_F^t \cup C_F(a)$.

Note that the definition of Value-Perfect feedback in the previous sections is a particular case of the above. On the other hand, we do not make additional assumptions on Response-Perfect feedback.

Finally, we modify the definition of an instance. The initial information in this setting consists of the vector $\mathcal{D}^0 := (A^0, I^0, C_F^0, C_L^0, \mathcal{U}^0, \mathbf{H}^0, \mathbf{h}^0, \mathbf{L}^0, \mathbf{f}^0, \mathbf{c}^0)$, and $\mathbb{G}(\mathcal{D}^0)$ becomes

$$\mathbb{G}(\mathcal{D}^0) := \{(A, I, C_F, C_L, \mathbf{F}, \mathbf{H}, \mathbf{h}, \mathbf{L}, \mathbf{f}, \mathbf{c}) : \text{conditions } \mathbf{C1}, \mathbf{C2} \text{ and } \mathbf{C3E-C5E} \text{ below hold}\},$$

where

C3E: \mathcal{U}^0 has valid upper and lower bounds for all F_{da} , $d \in C_F^0$, $a \in A^0$.

C4E: $(F_{da} : d \in C_F^0, a \in A^0) \in \mathcal{U}^0$.

C5E: $\mathbf{H}^0, \mathbf{h}^0, \mathbf{L}^0, \mathbf{f}^0, \mathbf{c}^0$ are submatrices and subvectors of $\mathbf{H}, \mathbf{h}, \mathbf{L}, \mathbf{f}, \mathbf{c}$.

The above definitions are straightforward extensions of the assumptions and definitions of the basic cost model in Section 2. Using them, we extend most of the results in the next sections.

4.2. Extended Greedy and Robust Policies

In what follows we generalize the greedy and robust policies in Λ to the matrix model which we denote by Λ_E . Policies in Λ_E are greedy because they maximize the follower's costs at the next time period, and they are robust because they consider all possible realizations of $\hat{\mathbf{F}}^t$ over \mathcal{U}^t . As shown below, these policies share most properties of the policies in Λ under the different modes of feedback.

For any $t \in \mathcal{T}$, and given any $x \in X^t$ define the ‘‘robust’’ region $Y_E^t(x)$ as

$$Y_E^t(x) := \left\{ y \in \mathbb{R}_+^{|A^t|} : \hat{\mathbf{F}}^t y + \mathbf{L}^t x \leq \mathbf{f}^t \ \forall \hat{\mathbf{F}}^t \in \mathcal{U}^t \right\}.$$

The robustness of $Y_E^t(x)$ follows from the fact that any element of this set must be feasible for any possible realization of the uncertain data in \mathcal{U}^t . Define

$$z_E^t(x) := \min\{(\mathbf{c}^t)^\top y : y \in Y_E^t(x)\}, \quad x \in X^t \quad \text{and} \quad z_E^{t,*} := \max\{z_E^t(x) : x \in X^t\} \quad t \in \mathcal{T}.$$

Additionally, for any policy π define $\xi_E^\pi := \xi^\pi(\mathcal{D}^0, \mathcal{D})$ as,

$$\xi_E^\pi := \min\{t \in \mathcal{T} : z_E^{t,*} = z^{t,\pi}\}.$$

DEFINITION 6. We say that $\lambda \in \Lambda_E \subseteq \Pi$ if and only if

$$x^{t,\lambda} \in \arg \max\{z_E^t(x) : x \in X^t\} \quad \forall t \leq \xi^\lambda,$$

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

As before, ξ_E^λ is the first time period when the follower uses a solution with the cost expected by the leader. Finally, from ξ_E^λ onwards, policies in Λ_E repeat the same solution used at time ξ_E^λ .

4.2.1. Policies in Λ_E under Standard and Value-Perfect Feedback The following proposition states that the standard feedback results that hold for Λ in Section 3.1, (i.e., Theorem 1 and Proposition 1) also hold for Λ_E .

PROPOSITION 4. *Let $\lambda \in \Lambda_E$ be given and assume that \mathcal{F} is standard. Then,*

- (i) *For any given $t \in \mathcal{T}$ it follows that $z^{t,\lambda} \leq z^* \leq z_E^{t,*}$.*
- (ii) *$\tau^\lambda \leq \xi_E^\lambda$.*
- (iii) *Given $t \in \mathcal{T}$, if $\dim(\mathcal{U}^t) = 0$ and $y_a^t = 0$ for all $a \notin A^t$, then $\xi_E^\lambda \leq t$, and, in particular, $\tau^\lambda \leq t$.*

In addition, given the extended definition of Value-Perfect feedback, Lemma 1 and Theorem 2 can be generalized in a straightforward fashion for the policies in Λ_E . Indeed, define \tilde{A}_E^t as the set of the follower's activities for which the leader knows (with certainty) the values of the columns of A associated with them, that is,

$$\tilde{A}_E^t := \{a \in A^t : \forall \hat{\mathbf{F}} \in \mathcal{U}^t \ \hat{F}_{da} = F_{da} \ \forall d \in C_F^t\}.$$

PROPOSITION 5. *Suppose $\lambda \in \Lambda_E$ and that feedback \mathcal{F} is Value-Perfect. Then,*

- (i) *If $z^{t,\lambda} < z_E^{t,*}$ then $\tilde{A}_E^{t+1} \setminus \tilde{A}_E^t \neq \emptyset$.*
- (ii) *$\tau^\lambda \leq \xi_E^\lambda \leq |A \setminus \tilde{A}_E^0|$.*

4.2.2. Policies in Λ_E under Response-Perfect Feedback In this section we establish convergence under Response-Perfect feedback for policies in Λ_E . In contrast with the Value-Perfect case, the extended results are more involved. We begin with the following observation.

LEMMA 3. *Let $\lambda \in \Lambda_E$, and suppose that $z^{t,\lambda} < z_E^{t,*}$ and that $y_a^t = 0$ for all $a \notin A^t$. Then there exist a $\tilde{\mathbf{F}}^t \in \mathcal{U}^t$ and a lower-level constraint $d \in C_F^t$ such that*

$$\left(\tilde{\mathbf{F}}_d^t\right)^\top y^{t,\lambda} > f_d - (\mathbf{L}_d^t)^\top x^{t,\lambda}. \quad (10)$$

The above result implies that the leader can remove matrix $\tilde{\mathbf{F}}^t$ from the uncertainty set at time t , as equation (10) means that $\tilde{\mathbf{F}}^t \neq \mathbf{F}^t$. For any given $t \in \mathcal{T}$ and $\lambda \in \Lambda_E$, let us define $D^{t,\lambda}$ as the set of constraints for which equation (10) holds at time t , that is

$$D^{t,\lambda} := \{d \in C_F^t : \exists \tilde{\mathbf{F}}^t \in \mathcal{U}^t \text{ s.t. } \left(\tilde{\mathbf{F}}_d^t\right)^\top y^{t,\lambda} > f_d - (\mathbf{L}_d^t)^\top x^{t,\lambda}\}.$$

Suppose that $z^{t,\lambda} < z_E^{t,*}$ and $y_a^t = 0$ for all $a \notin A^t$. Under the assumption of Response-Perfect feedback, one direct way to remove those elements of \mathcal{U}^t that satisfy equation (10) is to define \mathcal{U}^{t+1} as

$$\mathcal{U}^{t+1} = \{\hat{\mathbf{F}}^t \in \mathcal{U}^t : \left(\hat{\mathbf{F}}_d^t\right)^\top y^{t,\lambda} \leq f_d - (\mathbf{L}_d^t)^\top x^{t,\lambda} \forall d \in D^{t,\lambda}\}, \quad (11)$$

where we note that $\mathcal{U}^{t+1} \subset \mathcal{U}^t$ by Lemma 3. On the other hand, if $y_a^t > 0$ for some $a \notin A^t$, then, in general, the existence of a $\tilde{\mathbf{F}}^t$ such that (10) holds cannot be guaranteed, and hence the update in equation (11) can be vacuous (i.e., $\mathcal{U}^{t+1} = \mathcal{U}^t$).

From the above discussion it is clear that whenever the leader does not learn a new follower activity, then her uncertainty set reduces its size. However, the update defined by (11) does not necessarily reduce the dimension of \mathcal{U}^t , and hence an upper bound similar to that of Theorem 3 cannot be proved in this setting by using the polyhedral dimension arguments. However, if we make additional assumptions about the lower-level problem or about the leader's ability to observe the said problem, a finite upper bound can be established. These assumptions guarantee that the uncertainty update reduces the dimension of the uncertainty polyhedron at least by one.

PROPOSITION 6. *Let $\lambda \in \Lambda_E$ and suppose that \mathcal{F} is Response-Perfect.*

(i) *If all constraints of the lower-level problem are equalities, then*

$$\tau^\lambda \leq \xi^\lambda \leq \dim(\mathcal{U}^0) + \sum_{a \in A \setminus A^0} |C_F(a)|.$$

(ii) *If for any period $t \in \mathcal{T}$ such that $y_a^t = 0$ for all $a \notin A^t$ the leader observes the slack associated with at least one of the constraints in $D^{t,\lambda}$, then*

$$\tau^\lambda \leq \xi^\lambda \leq \dim(\mathcal{U}^0) + \sum_{a \in A \setminus A^0} (|C_F(a)| + 1).$$

Observe that all of the upper-bound results for policies in Λ (or Λ_E) proved so far rely on the fact that whenever the leader does not learn a new activity, then the dimension of \mathcal{U}^{t+1} can be made strictly less than the dimension of \mathcal{U}^t . For the matrix model and under Response-Perfect feedback, if no additional assumptions are made, then this reduction in dimension cannot be guaranteed. In this general setting, however, we can prove that every time \mathcal{U}^t is updated, the difference in ‘size’ between \mathcal{U}^{t+1} and \mathcal{U}^t is sufficiently large.

PROPOSITION 7. *Let $\epsilon > 0$, $\lambda \in \Lambda_E$, and $t \in \mathcal{T}$ be given. Assume that $y_a^t = 0$ for all $a \notin A^t$ and define $\Delta^t := \mathcal{U}^t \setminus \mathcal{U}^{t+1}$. If $z_E^* - z^{t,\lambda} > \epsilon$, then there exist $K > 0$ (independent of ϵ) such that*

$$\text{diam}(\Delta^t) > \frac{-\|y^{t,\lambda}\| + \sqrt{\|y^{t,\lambda}\|^2 + 4K\epsilon^2\|\mathbf{c}^t\|^{-2}}}{2K},$$

where $\text{diam}(\Delta^t)$ denotes the diameter of polyhedron Δ^t , i.e., $\text{diam}(\Delta^t) = \max_{\mathbf{F}', \mathbf{F}'' \in \Delta^t} \|\mathbf{F}' - \mathbf{F}''\|$.

5. Semi-Oracle Lower Bounds

In online optimization, the performance of a policy is compared against that of an *oracle*, who represents an ideal decision-maker who has all information of the problem beforehand, see Cesa-Bianchi and Lugosi (2006). Such an oracle faces no uncertainty and is able to make the best possible decision. In our problem setting, the oracle solves problem (1) at every period, and thus always attains a time-stability of zero. Unfortunately, such a lower bound is rather trivial and of not particular interest.

Consider instead a *weaker* oracle that, albeit knowing all the information of the problem in advance, has restrictions in the way she can use this information. Specifically, at any period such a weaker oracle can only use resources that she initially knows at time $t = 0$, or that have been revealed to her by the follower in previous periods. Hence, this *semi-oracle*, see Borrero et al. (2016), represents a decision-maker that combines both the practical limitations of the leader, with all the knowledge of the traditional oracle. Specifically, the semi-oracle solves:

$$\min \sum_{t \in \mathcal{T}} \mathbb{1}_{\{\mathbf{c}^\top y^t < z^*\}} \tag{12a}$$

$$\text{s.t. } x^t \in X \quad t \in \mathcal{T} \tag{12b}$$

$$y^t \in \arg \min \{\mathbf{c}^\top y : y \in Y(x^t)\} \quad t \in \mathcal{T} \tag{12c}$$

$$x_i^t = 0 \quad i \in I \setminus I^t, t \in \mathcal{T} \tag{12d}$$

$$I^{t+1} = I^t \cup \bigcup_{a: y_a^t > 0} I(a) \quad t \in \mathcal{T} \setminus \{T\}, \tag{12e}$$

where constraint (12d) prevents the semi-oracle from using activities which she does not know by time t . Observe that absent this constraint, the formulation corresponds to what the oracle (with

full information) would solve. Constraints (12b) and (12c), on the other hand, imply that the semi-oracle has all the information of the problem. As a consequence, the leader cannot be expected to formulate nor optimally solve (at least, consistently) the problem given by (12) in practice.

There are two main advantages of using the notion of the semi-oracle, rather than the oracle, as a benchmark. First, it yields a more informative lower bound on the performance of any policy: the time-stability attained by the semi-oracle is not always zero; moreover, by using it we can evaluate the effect that the initial information has on the performance of any policy. Second, for any given instance, there is always a policy that attains the time-stability of the semi-oracle policy. Specifically, for any policy, any interaction between the leader and the follower can be mapped into a feasible solution of (12), and more importantly, given a *fixed* instance, there must exist a policy that yields the same values of x^t and y^t as an optimal solution of (12).

It is important to note, however, that the semi-oracle decision process does not constitute a feasible policy: given a same history $\mathcal{H}^t(\mathcal{D}^0, \mathcal{D})$, the semi-oracle might determine two different values for x^t for different instances, see an example for the sequential shortest path interdiction in Borrero et al. (2016). This, because problem (12) is a function of the instance $(\mathcal{D}^0, \mathcal{D})$, rather than a function of the history (as it is the case with any admissible policies; recall their definition in Section 2.2).

It can be readily seen that the semi-oracle optimization problem (12) is *NP*-hard. Small and moderately sized instances of the problem, however, can be tackled by state-of-the-art MIP solvers. Indeed, a single-level MIP reformulation of (12) is given by:

$$\min_{u,v,w,y,x,\theta} \sum_{t \in \mathcal{T}} w^t \quad (13a)$$

$$\text{s.t. } \mathbf{H}x^t \leq \mathbf{h} \quad t \in \mathcal{T} \quad (13b)$$

$$\mathbf{F}y^t + \mathbf{L}x^t \leq \mathbf{f}, \quad -\mathbf{F}^\top \theta^t \leq \mathbf{c} \quad t \in \mathcal{T} \quad (13c)$$

$$\theta^t \leq \mathbf{M}^{\theta^t} u^t, \quad y^t \leq \mathbf{M}^{y^t} v^t \quad t \in \mathcal{T} \quad (13d)$$

$$\mathbf{f} - \mathbf{F}y^t - \mathbf{L}x^t \leq \mathbf{M}^{p^t} (\mathbf{1} - u^t) \quad t \in \mathcal{T} \quad (13e)$$

$$\mathbf{c} + \mathbf{F}^\top \theta^t \leq \mathbf{M}^{q^t} (\mathbf{1} - v^t) \quad t \in \mathcal{T} \quad (13f)$$

$$x_i^t \leq M^{x_i} \sum_{s=0}^{t-1} \sum_{a \in A(i) \setminus A^0} y_a^s \quad t \in \mathcal{T}, \quad i \in I \setminus I^0 \quad (13g)$$

$$z^*(\mathbf{1} - M^w w^t) \leq \mathbf{c}^\top y^t \quad t \in \mathcal{T} \quad (13h)$$

$$u^t \in \{0, 1\}^{|C_F|}, v^t \in \{0, 1\}^{|A|}, w^t \in \{0, 1\} \quad t \in \mathcal{T} \quad (13i)$$

$$y^t \in \mathbb{R}_+^{|A|}, x^t \in \mathbb{R}_+^{|I|-k} \times \mathbb{Z}_+^k, \theta^t \in \mathbb{R}_+^{|C_F|} \quad t \in \mathcal{T}, \quad (13j)$$

where x^t is the solution of the semi-oracle at time t , and y^t is the solution of the follower at time $t \in \mathcal{T}$. The fact that $y^t \in \arg \min\{\mathbf{c}^\top y : y \in Y(x^t)\}$ is represented by its linear programming (LP)

optimality conditions via constraints (13c) (primal and dual feasibility) and (13d), (13e), and (13f) (the linearized complementary slackness conditions). In these constraints, \mathbf{M}^{θ^t} , \mathbf{M}^{p^t} , \mathbf{M}^{y^t} , and \mathbf{M}^{q^t} are diagonal matrices that are upper bounds on θ^t , $\mathbf{f} - \mathbf{F}y^t - \mathbf{L}x^t$, y^t , and $\mathbf{c} + \mathbf{F}^\top \theta^t$, respectively. We refer the reader to Audet et al. (1997) for more details on single-level MIP reformulations of bilevel problems with the lower-level problem given by an LP.

Variable w^t is binary and takes the value of zero if $\mathbf{c}^\top y^t = z^*$, i.e., if the optimal semi-oracle solution is used at time t , see constraint (13h). Here, $M^w = (z^* - \ell)/z^*$ and ℓ is a valid lower bound on the value of $\mathbf{c}^\top y$ for any feasible y . Finally, constraint (13g) implies that a resource cannot be used if it has not been revealed by the follower or if it is not in I^0 . In this constraint, $A(i)$ is the set of follower activities that i interferes with, i.e., $A(i) = \{a \in A : i \in I(a)\}$, and $M^{x_i} = u^i/\ell_i$, where u^i is an upper bound on the value of the i -th entry of any $x \in X$, and ℓ_i is a strictly positive lower bound on the value that any y_a , $a \in A(i)$, can take whenever $y_a > 0$. In general, the computation of these lower bounds can be highly involved, but for specific applications they can be computed rather efficiently from the problem’s data, see Section 6 for an example.

We close this section by noting that although MIP problem (13) can be solved directly for moderately sized instances, it might require lengthy computational times due to the large number of variables and constraints, particularly if T is large. It turns out, however, that this problem can be made somewhat less “dependent” on the time horizon T by feeding to the solver an initial feasible solution. This approach can drastically reduce the size of the resulting MIP, and thus lead to shorter computational times; see the discussion on this approach in Appendix C.1.

6. Numerical Illustration

In this section we demonstrate the numerical performance of the policies in Λ . For this, we use the AD Knapsack problem of Example 2. We consider both Value-Perfect and Response-Perfect feedbacks as well as two different models for the initial uncertainty set. In order to provide a broader picture of the performance of the policies in Λ , we compare them against reasonable benchmark policies in the context of **SMPI**, and with respect to the semi-oracle lower-bounding procedure of the previous section. Our results show that the policies in Λ outperform the benchmark, and compare rather favorably with respect to the semi-oracle lower bound.

The decisions generated by the policies in Λ are computed by solving a one-level MIP reformulation of the bilevel problem (6), see Section C.2 of the Appendix for further details. Generally speaking, the transformation of optimization problem (6) into an MIP involves application of methods from bilevel optimization (to transform the hierarchical problem into a single-level problem) and robust optimization (to adequately optimize over the uncertainty set \mathcal{U}^t) areas. We note that, in general, problem (6) is *NP*-hard, as bilevel linear optimization is its special case.

Test Instances. We consider the AD knapsack problem from Example 2, where the defender has $n = 12$ assets, $\mathbf{b} = \mathbf{r} = \mathbf{1}$, and $B = R = 4$. We consider two models of initial uncertainty sets, namely, *hypercube* uncertainty and *general* uncertainty:

- In the hypercube model the defender's profits satisfy $p_a \in [\ell_a, \ell_a + m_a]$, $a \in A$, where ℓ_a is drawn at random from uniform discrete $U(1, 5)$ distribution and m_a is drawn from $U(1, 15)$ distribution.
- For the general uncertainty model we generate a non-negative polytope with $C_U = 3$ inequalities. The polytope is given by $\mathcal{P} := \{p : Gp \leq g, p \geq 0\}$, where $G_{u,j}$ is drawn at random from $U(1, 10)$ distribution for $j \in \{4(u-1) + 1, 4u\}$, and $G_{u,j} = 0$ otherwise, while g_u is drawn at random from $U(1, 20)$ distribution, for $u = 1, 2, 3$.

Given a polytope \mathcal{P} , we generate the follower's profit vector \mathbf{p} by using the following approach. First, we compute the barycenter (or analytical center) of the polytope by solving the following convex problem (see, e.g., Bertsimas and Tsitsiklis (1997)):

$$\mathbf{p}_b \in \arg \min_{(\hat{p}, \hat{q}) \geq 0} \left\{ - \sum_{j=1}^n \log(\hat{p}_j) - \sum_{u=1}^{|C_U|} \log(\hat{q}_u) : \mathbf{G}\hat{p} + \hat{q} = \mathbf{g}, \hat{p} \geq 0 \right\}.$$

Next, we randomly construct an extreme point of \mathcal{P} by first generating a vector $\boldsymbol{\ell}$ of size n (where each entry is zero or one with the same probability) and then solving an LP of the form:

$$\mathbf{p}_e \in \arg \max_{\hat{p}} \{ \boldsymbol{\ell}^\top \hat{p} : \mathbf{G}\hat{p} \leq \mathbf{g}, \hat{p} \geq 0 \}.$$

Finally, we combine the barycenter with the obtained extreme point by $\mathbf{p} = (\mathbf{p}_b + 7\mathbf{p}_e)/8$, to generate an interior point of the polytope \mathcal{P} , and hence ensuring that $p_a > 0$ for all $a \in A$.

For each uncertainty model, we generated at random $N = 30$ instances, considering both Value-Perfect and Response-Perfect feedbacks. We consider three sets of initial information A^0 : in the first, the leader knows four activities of the follower; in the second, she knows eight activities; and in the last, she knows all activities. Finally, we set $T = 24$.

Benchmark Policies. In addition to policies in Λ , we consider the following benchmarks:

- The *barycenter policy* π_b : At each time $t \in \mathcal{T}$ the policy computes x^{t, π_b} by solving the deterministic bilevel problem

$$x^{t, \pi_b} \in \arg \max_{x \in X^t} \{ (\mathbf{c}^t)_b^\top y : y \in Y^t(x) \}, \quad (14)$$

where $\mathbf{c}_b^t = -\mathbf{p}_b^t$, and \mathbf{p}_b^t is the barycenter of the polytope \mathcal{U}^t .

- The *random policy* π_r : At each time $t \in \mathcal{T}$ the policy computes x^{t, π_r} by solving problem (14) with \mathbf{c}_r^t used instead of \mathbf{c}_b^t . We have $\mathbf{c}_r^t = -\mathbf{p}_r^t$, and \mathbf{p}_r^t is a randomly generated extreme point of \mathcal{U}^t that is obtained by solving the linear program $\mathbf{p}_r^t \in \arg \max \{ \boldsymbol{\ell}^{t, T} \hat{p} : \hat{p} \in \mathcal{U}^t \}$. In this problem, at each time $t \in \mathcal{T}$ each entry of vector $\boldsymbol{\ell}^t$ is drawn at random from a Bernoulli distribution with parameter $1/2$, i.e., each entry is zero or one with equal probability.

- The “stopped” random policy π_s : At each time $t \in \mathcal{T}$ the policy computes x^{t,π_s} in the same manner as policy π_r . However, whenever the leader observes a follower’s response that she has observed in the earlier time periods, then the policy keeps using the same solution thereafter. That is, if time t' is the earliest period such that $z^{t',\pi_s} = z^{t,\pi_s}$ for some $t < t'$, then $x^{t,\pi_s} = x^{t',\pi_s}$ for all $t \geq t'$.

- We also consider the *lower bound* provided by the semi-oracle approach discussed in Section 5. While it is not an admissible policy, with a slight abuse of notation we denote it by π^* hereafter.

Results and Discussion. For each uncertainty model we compute its time-stability across $N = 30$ replications by using each of the policies described above. Tables 1 and 2 report the mean time-stability and mean absolute deviation (MAD) for the hypercube uncertainty model under Value-Perfect and Response-Perfect feedbacks, respectively. Similarly, Tables 3 and 4 show the same results for the general uncertainty model. For the sake of reporting averages, *policies that do not find an optimal solution within the first 24 periods of an instance are assigned the value $\tau^\pi = T = 24$.*

(a) Value-Perfect: time-stability mean						(b) Value-Perfect: time-stability MAD					
A^0	λ	π_b	π_r	π_s	π^*	A^0	λ	π_b	π_r	π_s	π^*
$\{1, \dots, 4\}$	2.13	20.17	21.70	21.70	1	$\{1, \dots, 4\}$	0.29	6.39	4.14	4.14	0.00
$\{1, \dots, 8\}$	2.93	21.73	23.20	23.20	0.93	$\{1, \dots, 8\}$	0.50	4.08	1.55	1.55	0.12
A	3.03	21.77	24.00	24.00	0	A	0.59	4.02	0.00	0.00	0

Table 1 Time-stability mean and MAD for the hypercube uncertainty model and Value-Perfect feedback.

(a) Response-Perfect: time-stability mean						(b) Response-Perfect: time-stability MAD					
A^0	λ	π_b	π_r	π_s	π^*	A^0	λ	π_b	π_r	π_s	π^*
$\{1, \dots, 4\}$	7.77	23.23	21.10	21.87	1	$\{1, \dots, 4\}$	2.21	1.48	4.64	3.84	0.00
$\{1, \dots, 8\}$	8.00	23.20	23.20	23.20	0.93	$\{1, \dots, 8\}$	1.87	1.55	1.55	1.55	0.12
A	7.93	24.00	24.00	24.00	0	A	1.74	0.00	0.00	0.00	0

Table 2 Time-stability mean and MAD for the hypercube uncertainty model and Response-Perfect feedback.

(a) Value-Perfect: time-stability mean						(b) Value-Perfect: time-stability MAD					
A^0	λ	π_b	π_r	π_s	π^*	A^0	λ	π_b	π_r	π_s	π^*
$\{1, \dots, 4\}$	2	14.20	4.43	3.07	1	$\{1, \dots, 4\}$	0	11.11	2.77	1.52	0.00
$\{1, \dots, 8\}$	1	10.27	5.20	4.30	0.97	$\{1, \dots, 8\}$	0	10.99	3.73	3.94	0.06
A	1	9.37	7.67	6.67	0	A	0	10.73	6.56	6.93	0

Table 3 Time-stability mean and MAD for the general uncertainty model and Value-Perfect feedback.

(a) Response-Perfect: time-stability mean						(b) Response-Perfect: time-stability MAD					
A^0	λ	π_b	π_r	π_s	π^*	A^0	λ	π_b	π_r	π_s	π^*
$\{1, \dots, 4\}$	6.90	14.23	9.10	11.37	1	$\{1, \dots, 4\}$	0.30	11.07	3.87	8.42	0.00
$\{1, \dots, 8\}$	7.50	17.30	14.53	12.97	0.97	$\{1, \dots, 8\}$	0.53	9.38	7.37	9.56	0.06
A	7.73	16.77	17.73	11.93	0	A	0.41	9.64	7.14	8.85	0

Table 4 Time-stability mean and MAD for the general uncertainty model and Response-Perfect feedback

We observe that the proposed policies $\lambda \in \Lambda$ consistently outperform the benchmark except for the semi-oracle lower bound π^* , which is expected. For most instances in the hypercube model, policies π_b , π_r and π_s yield very poor time-stability results, not being able to find an optimal solution for most cases within the time horizon. Their performance improves, however, for the general uncertainty model. These results reflect one of the key advantages of the greedy and robust nature of the policies in Λ , namely, the fact that the leader is guaranteed to eventually find an optimal solution to the full information problem.

Furthermore, we observe that the performance of the proposed policies is better for the case of Value-Perfect feedback when compared to Response-Perfect feedback, by a factor of at least two, under both uncertainty models. This is to be expected: under Value-Perfect feedback more linearly independent equations are added on average to \mathcal{U}^t at each time period. It is also noticeable that the amount of initial information does not seem to have any significant impact on policy performance under both feedback types and uncertainty models. Although this behavior is rather counter-intuitive, it might stem from the fact that in this particular bilevel setting, the follower's activities are fairly independent of each other (they are interrelated only through the follower's budget constraint). Hence, partial knowledge of the follower activities does not implicitly reveal much information about the remaining unknown activities. We note that in more complex bilevel settings the amount of initial information does have a very important effect (see, e.g., discussion in Borrero et al. (2016) for an example in the context of the shortest path interdiction).

An important feature of the policies in Λ is their low variability. This is especially true in the general uncertainty model, where the policies yield no variability in the Value-Perfect setting, and a very low variability in the Response-Perfect setting. In contrast, the benchmark policies are orders of magnitude more variable in the general uncertainty setting. While these policies have low MAD values in Tables 1(b) and 2(b) for the hypercube uncertainty, this is due to the fact that for most instances their time-stability is infinity (recall our earlier remark that policies that do not find an optimal solution within the first 24 periods of an instance are assigned the value $\tau^\pi = T = 24$).

7. Conclusions

This paper presents a framework for addressing **SMPI** where at each period a leader allocates a series of resources so as to degrade the performance of a follower, who in turn aims at minimizing a cost function by performing a series of activities. The interaction at each time period is modeled as a bilevel program. We assume that, unlike the follower, the leader has incomplete information about the variables, constraints, and cost function of the follower's problem and has to learn them by observing the feedback generated by the follower's actions. Such feedback includes the total cost incurred by the follower, the activities performed, and any resource that might interfere with

the said activities. Such settings naturally arise in military and law enforcement applications, e.g., attacker-defender and interdiction problems, which are often modeled as max-min bilevel problems.

We propose a class of policies Λ that are both greedy and robust, as they optimize the immediate performance considering worst-case realizations of the instance among those that are consistent with the information at hand. Under reasonable assumptions on the information that the leader collects from the follower’s response, our theoretical results show that in **SMPI** exploitation always implies exploration as long as the leader is using policies in Λ , and moreover, their greediness and robustness are sufficient to guarantee weak optimality. Particularly, we show that the time-stability of policies in Λ is upper-bounded by the number of the follower’s activities and the dimension of the cost’s uncertainty polyhedron, which implies that they are guaranteed to eventually match the actions of the oracle with prior knowledge of the instance. Moreover, we show that these policies provide the leader with a real-time certificate of optimality.

We also consider a more general setting where the leader has uncertainty regarding the follower’s constraint matrix. We demonstrate that the extension of greedy and robust policies preserves most of the attractive features of their cost-model counterparts. Particularly, no extra assumptions are required to extend the time-stability upper bounds under Value-Perfect feedback, while only mild assumptions are required to preserve the upper bounds under Response-Perfect feedback.

Implementation of the proposed policies requires solving a linear MIP in each period: these problems can be solved by available commercial solvers. We also present a lower bound on the best possible achievable performance based on the actions of a *semi-oracle* that possess full information about the setting, but cannot signal it through her actions. We show that the said bound can also be computed via an MIP. Our theoretical results are supported by a series of numerical experiments that show that the proposed policies consistently outperform reasonable benchmark.

Several questions remain open at this point with regard to sequential bilevel problems with incomplete information. Possibly the more relevant is to study up to what point the results in this work can be extended to general (i.e., not necessarily max-min) bilevel programs. Also, models with more general assumptions on uncertainty, where, for instance, the leader is not certain about her upper-level data, provide an attractive avenue of future research. Regarding **SMPI**, the question of determining whether finite time-stability upper bounds can be proved for the matrix model under Response-Perfect feedback with no extra assumptions remains open, as well as to determine alternative feedback settings where finite bounds, and weak optimality, can be also be attained.

References

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

- Audibert, J.-Y. and Bubeck, S. (2009), Minimax policies for adversarial and stochastic bandits, in S. Dasgupta and A. Klivans, eds, ‘Proceedings of the 21st Annual Conference on Learning Theory (COLT)’, Omnipress, pp. 217–226.
- Audibert, J.-Y., Bubeck, S. and Lugosi, G. (2013), ‘Regret in online combinatorial optimization’, *Mathematics of Operations Research* **39**(1), 31–45.
- Auer, P., Cesa-Bianchi, N., Freund, Y. and Schapire, R. E. (2002), ‘The nonstochastic multiarmed bandit problem’, *SIAM Journal on Computing* **32**(1), 48–77.
- Awerbuch, B. and Kleinberg, R. D. (2004), Adaptive routing with end-to-end feedback: Distributed learning and geometric approaches, in L. Babai, ed., ‘Proceedings of the Thirty-Sixth Annual ACM Symposium on Theory of Computing’, ACM, pp. 45–53.
- Bard, J. F., Plummer, J. and Sourie, J. C. (2000), ‘A bilevel programming approach to determining tax credits for biofuel production’, *European Journal of Operational Research* **120**(1), 30–46.
- Ben-Tal, A., El Ghaoui, L. and Nemirovski, A. (2009), *Robust optimization*, Princeton University Press.
- Bertsimas, D. and Tsitsiklis, J. N. (1997), *Introduction to linear optimization*, Vol. 6, Athena Scientific Belmont, MA.
- Borrero, J. S., Prokopyev, O. A. and Sauré, D. (2016), ‘Sequential shortest path interdiction with incomplete information’, *Decision Analysis* **13**(1), 68–98.
- Brown, G., Carlyle, M., Diehl, D., Kline, J. and Wood, K. (2005), ‘A two-sided optimization for theater ballistic missile defense’, *Operations Research* **53**(5), 745–763.
- Brown, G., Carlyle, M., Salmerón, J. and Wood, K. (2006), ‘Defending critical infrastructure’, *Interfaces* **36**(6), 530–544.
- Bubeck, S. and Cesa-Bianchi, N. (2012), ‘Regret analysis of stochastic and nonstochastic multi-armed bandit problems’, *CoRR* **abs/1204.5721**.
URL: <http://arxiv.org/abs/1204.5721>
- Caprara, A., Carvalho, M., Lodi, A. and Woeginger, G. J. (2013), A complexity and approximability study of the bilevel knapsack problem, in ‘Integer programming and combinatorial optimization’, Springer, pp. 98–109.
- Cesa-Bianchi, N., Freund, Y., Haussler, D., Helmbold, D. P., Schapire, R. E. and Warmuth, M. K. (1997), ‘How to use expert advice’, *Journal of the ACM* **44**(3), 427–485.
- Cesa-Bianchi, N. and Lugosi, G. (2006), *Prediction, learning, and games*, Cambridge University Press.
- Cesa-Bianchi, N. and Lugosi, G. (2012), ‘Combinatorial bandits’, *Journal of Computer and System Sciences* **78**(5), 1404–1422.
- Chern, M. and Lin, K. (1995), ‘Interdicting the activities of a linear program: a parametric analysis’, *European Journal of Operational Research* **86**(3), 580–591.

- Colson, B., Marcotte, P. and Savard, G. (2005), ‘Bilevel programming: A survey’, *4OR* **3**(2), 87–107.
- Colson, B., Marcotte, P. and Savard, G. (2007), ‘An overview of bilevel optimization’, *Annals of Operations Research* **153**(1), 235–256.
- Corley, H. and Chang, H. (1974), ‘Finding the n most vital nodes in a flow network’, *Management Science* **21**(3), 362–364.
- Côté, J.-P., Marcotte, P. and Savard, G. (2003), ‘A bilevel modelling approach to pricing and fare optimisation in the airline industry’, *Journal of Revenue and Pricing Management* **2**(1), 23–36.
- Dempe, S. (2002), *Foundations of bilevel programming*, Kluwer Academic Publishers.
- DeNegre, S. (2011), Interdiction and discrete bilevel linear programming, PhD thesis, Lehigh University.
- Diaz, J. B. and Metcalf, F. T. (1964), ‘Complementary inequalities i: Inequalities complementary to cauchy’s inequality for sums of real numbers’, *Journal of Mathematical Analysis and Applications* **9**(1), 59–74.
- Fulkerson, D. and Harding, G. (1977), ‘Maximizing the minimum source-sink path subject to a budget constraint’, *Mathematical Programming* **13**(1), 116–118.
- Ghare, P., Montgomery, D. and Turner, W. (1971), ‘Optimal interdiction policy for a flow network’, *Naval Research Logistics Quarterly* **18**(1), 37–45.
- Gyorgy, A., Linder, T., Lugosi, G. and Ottucsak, G. (2007), ‘The on-line shortest path problem under partial monitoring’, *Journal of Machine Learning Research* **8**(10), 2369–2403.
- Hazan, E. (2015), ‘Introduction to online convex optimization (Draft)’, Foundations and Trends in Optimization.
- URL:** <http://ocobook.cs.princeton.edu/OCObook.pdf>
- Hazan, E., Agarwal, A. and Kale, S. (2007), ‘Logarithmic regret algorithms for online convex optimization’, *Machine Learning* **69**(2), 169–192.
- Held, H., Hemmecke, R. and Woodruff, D. (2005), ‘A decomposition algorithm applied to planning the interdiction of stochastic networks’, *Naval Research Logistics* **52**(4), 321–328.
- Held, H. and Woodruff, D. (2005), ‘Heuristics for multi-stage interdiction of stochastic networks’, *Journal of Heuristics* **11**(5-6), 483–500.
- Hemmecke, R., Schultz, R. and Woodruff, D. L. (2003), Interdicting stochastic networks with binary interdiction effort, in ‘Network Interdiction and Stochastic Integer Programming’, Springer, pp. 69–84.
- Israeli, E. and Wood, R. (2002), ‘Shortest-path network interdiction’, *Networks* **40**(2), 97–111.
- Janjarassuk, U. and Linderoth, J. (2008), ‘Reformulation and sampling to solve a stochastic network interdiction problem’, *Networks* **52**(3), 120–132.
- Kalai, A. and Vempala, S. (2005), ‘Efficient algorithms for online decision problems’, *Journal of Computer and System Sciences* **71**(3), 291–307.

- Kleinberg, R., Niculescu-Mizil, A. and Sharma, Y. (2010), ‘Regret bounds for sleeping experts and bandits’, *Machine Learning* **80**(2-3), 245–272.
- Lucotte, M. and Nguyen, S. (2013), *Equilibrium and advanced transportation modelling*, Springer Science & Business Media.
- McMasters, A. and Mustin, T. (1970), ‘Optimal interdiction of a supply network’, *Naval Research Logistics Quarterly* **17**(3), 261–268.
- Morton, D., Pan, F. and Saeger, K. (2007), ‘Models for nuclear smuggling interdiction’, *IIE Transactions* **39**(1), 3–14.
- Ratliff, H., Sicilia, G. and Lubore, S. (1975), ‘Finding the n most vital links in flow networks’, *Management Science* **21**(5), 531–539.
- Salmeron, J., Wood, K. and Baldick, R. (2004), ‘Analysis of electric grid security under terrorist threat’, *IEEE Transactions on Power Systems* **19**(2), 905–912.
- Shen, S., Smith, J. and Goli, R. (2012), ‘Exact interdiction models and algorithms for disconnecting networks via node deletions’, *Discrete Optimization* **9**(3), 172–188.
- Sherali, H. D., Soyster, A. L. and Murphy, F. H. (1983), ‘Stackelberg-Nash-Cournot equilibria: characterizations and computations’, *Operations Research* **31**(2), 253–276.
- Smith, J. C. and Lim, C. (2008), Algorithms for network interdiction and fortification games, in ‘Pareto optimality, game theory and equilibria’, Springer, pp. 609–644.
- Wollmer, R. (1964), ‘Removing arcs from a network’, *Operations Research* **12**(6), 934–940.
- Wolsey, L. A. and Nemhauser, G. L. (2014), *Integer and combinatorial optimization*, John Wiley & Sons.
- Wood, R. K. (1993), ‘Deterministic network interdiction’, *Mathematical and Computer Modelling* **17**(2), 1–18.
- Wood, R. K. (2011), ‘Bilevel network interdiction models: Formulations and solutions’, *Wiley Encyclopedia of Operations Research and Management Science* .
- Zenklusen, R. (2010), ‘Matching interdiction’, *Discrete Applied Mathematics* **158**(15), 1676–1690.
- Zinkevich, M. (2003), Online convex programming and generalized infinitesimal gradient ascent, in T. Fawcett and N. Mishra, eds, ‘Proceedings of the Twentieth International Conference on Machine Learning’, AAAI, pp. 928–936.

Appendix A: Proofs of the results for the basic model from Sections 2 and 3

We first introduce some auxiliary results. We have the following basic observation:

LEMMA 4. For any $t \in \mathcal{T}$ and $x \in X^t$,

$$z_R^t(x) = \min_{y'} \{(\mathbf{d}^t)^\top y' : y' \in Y_R^t(x)\} \quad (\text{A-1})$$

where $\mathbf{d}^t = (1, 0, \dots, 0)^\top$ and

$$Y_R^t(x) := \{(y_0, y) \in \mathbb{R} \times \mathbb{R}_+^{|A^t|} : -y_0 + (\mathbf{c}^t)^\top y \leq 0 \ \forall \hat{\mathbf{c}}^t \in \mathcal{U}^t, y \in Y^t(x)\}.$$

Proof. Note that $z_R^t(x)$ can be equivalently posed as

$$\begin{aligned} & \min_{y_0, y} y_0 \\ & \text{s.t. } y_0 \geq \max_{\hat{\mathbf{c}}^t \in \mathcal{U}^t} (\hat{\mathbf{c}}^t)^\top y \\ & \mathbf{F}^t y + \mathbf{L}^t x \leq \mathbf{f}^t \\ & y \geq 0. \end{aligned}$$

The result follows after noting that (y_0, y) satisfies the first of the above constraints if and only if $y_0 \geq (\mathbf{c}^t)^\top y$ for all $\hat{\mathbf{c}}^t \in \mathcal{U}^t$. \blacksquare

In all the proofs that follow use the representation of $z_R^t(x)$ given by equation (A-1) instead of the representation given by the original definition.

Proof of Theorem 1. (i) We first proof the statement that $z^{t,\lambda} \leq z^* \leq z_R^{t,*}$. For the leftmost inequality, the result follows from the definition of both z^* and $z^{t,\lambda}$ (see equations (1) and (3)) because $\bar{x}^{t,\lambda} \in X$, the feedback is standard and Assumption **A3** holds. For the rightmost inequality, let x^* be an element of X that attains z^* . Partition x^* as $x^* = (\hat{x}, \tilde{x})$, where $\hat{x} = (x_i^*)_{i \in I^t}$ and $\tilde{x} = (x_i^*)_{i \in I \setminus I^t}$. Recall the definition of the partition of matrices given by (4). Therefore, because $x^* \in X$ and **A3** holds, one has that $\hat{x} \in X^t$.

Now, suppose that $Y_R^t(\hat{x})$ is non-empty (if it is empty then it must be the case that $z_R^{t,*} = +\infty$ and the result holds) and let (y_0, \hat{y}) be such that

$$(y_0, \hat{y}) \in \arg \min \{(\mathbf{d}^t)^\top y' : y' \in Y_R^t(\hat{x})\},$$

hence $(\mathbf{c}^t)^\top \hat{y} = z_R^t(\hat{x})$. By the definition of $z_R^{t,*}$ we have that

$$(\mathbf{c}^t)^\top \hat{y} \leq z_R^{t,*}. \tag{A-2}$$

Define \bar{y} as $\bar{y}_a := \hat{y}_a$ if $a \in A^t$, and $\bar{y}_a := 0$ if $a \in A \setminus A^t$. Because \mathcal{F} is standard, Assumption **A4** holds, and $\hat{y} \in Y^t(\hat{x})$, thus it follows that $\bar{y} \in Y(x^*)$. Therefore,

$$z^* \leq \mathbf{c}^\top \bar{y}. \tag{A-3}$$

As $\mathbf{c}^\top \bar{y} = (\mathbf{c}^t)^\top \hat{y}$, equations (A-2) and (A-3) yield the desired result.

(ii) Next, we show that $\tau^\lambda \leq \xi^\lambda$. Recall the definition of $\bar{x}^{t,\lambda}$ (see equations (2) and (3)), i.e., $\bar{x}_i^{t,\lambda} = x_i^{t,\lambda}$ if $i \in I^t$ and $\bar{x}_i^{t,\lambda} = 0$ if $i \notin I^t$. For notational convenience, in the proof, let $\xi = \xi^\lambda$. We claim

that $\bar{x}^{\xi,\lambda} \in \arg \max\{z(x) : x \in X\}$. Indeed, the fact that the feedback is standard (recall equation (4)) implies that $\bar{x}^{\xi,\lambda} \in X$. Because by definition of ξ we have that $z^{\xi,\lambda} = z_R^{\xi,*}$, part (i) implies that (recall that by definition we have $z^{t,\lambda} = z(\bar{x}^{t,\lambda})$ for any t)

$$z(\bar{x}^{\xi,\lambda}) = z^*,$$

and therefore the claim follows. Now, by definition of λ , for all $s \geq t$ it must be the case that $x^{s,\lambda} = x^{\xi,\lambda}$. We claim that this implies that $z^{s,\lambda} = z^*$ for all $s \geq t$, and hence that $\tau^\lambda \leq \xi^\lambda$. In order to arrive at a contradiction, assume that $z^{s,\lambda} < z^*$ for $s > \xi^\lambda$; because $x^{s,\lambda} = x^{\xi,\lambda}$, one has that $y^{s,\lambda} \in Y(x^{\xi,\lambda})$, and by the definition of $y^{\xi,\lambda}$ it would follow that $z^{\xi,\lambda} \leq z^{s,\lambda} < z^*$, which contradicts the fact that $z^{\xi,\lambda} = z^*$. Therefore, the desired claim follows. ■

Proof of Proposition 1. As $y^{t,\lambda} \in Y(x^{t,\lambda})$ and $y_a^t = 0$ for all $a \notin A^t$, it follows that

$$\sum_{a \in A^t} F_{da} y_a^{t,\lambda} + \sum_{i \in I^t} L_{di} x_i^{t,\lambda} \leq f_d \quad \forall d \in C_F^t,$$

which implies that $(y_a^{t,\lambda})_{a \in A^t} \in Y^t(x^{t,\lambda})$. On the other hand, as \mathcal{U}^t has dimension zero, the set $Y_R^t(x^{t,\lambda})$ becomes

$$Y_R^t(x^{t,\lambda}) = \{(y_0, y) \in \mathbb{R}_+^{|A^t|} : -y_0 + (\mathbf{c}^t)^\top y \leq 0, y \in Y(x^{t,\lambda})\},$$

and hence, $z_R^t(x^{t,\lambda}) \leq (\mathbf{c}^t)^\top (y^{t,\lambda})_{a \in A^t}$. Therefore, from the first set of inequalities of Theorem 1 (see part (i) in its proof above) and as $z_R^t(x^{t,\lambda}) = z_R^{t,*}$ by definition of $x^{t,\lambda}$, we have that

$$z^{t,\lambda} \leq z_R^{t,*} \leq (\mathbf{c}^t)^\top (y^{t,\lambda})_{a \in A^t} \tag{A-4}$$

but on the other hand, from the definition of $y^{t,\lambda}$ we have that

$$z^{t,\lambda} = (\mathbf{c}^t)^\top (y^{t,\lambda})_{a \in A^t}. \tag{A-5}$$

Equations (A-4) and (A-5) imply that $z^{t,\lambda} = z_R^{t,*}$, and hence $\xi^\lambda \leq t$ as desired. The later part of the proposition is a consequence of the above result and the second set of inequalities of Theorem 1 (see part (ii) in its proof above). ■

Proof of Lemma 1. First, note that if $y_a^{t,\lambda} > 0$ for some $a \notin A^t$, then the result follows from the assumptions of Value-Perfect feedback. Therefore, suppose that $y_a^{t,\lambda} = 0$ for all $a \notin A^t$. We claim that there exists an activity $a \in A^t \setminus \tilde{A}^t$ such that $y_a^{t,\lambda} > 0$; the existence of such an activity implies the desired result from the assumptions of Value-Perfect feedback. Indeed, to proceed by contradiction, suppose that this is not the case, i.e., $y_a^{t,\lambda} = 0$ for all $a \in A^t \setminus \tilde{A}^t$. As $y^{t,\lambda} \in Y(\bar{x}^{t,\lambda})$

and $y_a^t = 0$ for all $a \notin A^t$ it must be that $(y_a^{t,\lambda})_{a \in A^t} \in Y^t(x^{t,\lambda})$. Now, because $\hat{c}_a = c_a$ for all $a \in \tilde{A}^t$, one has that for all $\hat{c}^t \in \mathcal{U}^t$

$$(\hat{c}^t)^\top (y_a^{t,\lambda})_{a \in A^t} = (c^t)^\top (y_a^{t,\lambda})_{a \in A^t},$$

and therefore $((c^t)^\top (y_a^{t,\lambda})_{a \in A^t}, (y_a^{t,\lambda})_{a \in A^t}) \in Y_R^t(x^{t,\lambda})$. Thus, by the definition of $x^{t,\lambda}$ we have that

$$z_R^{t,*} \leq (c^t)^\top (y_a^{t,\lambda})_{a \in A^t}. \quad (\text{A-6})$$

On the other hand, because $y_a^t = 0$ for all $a \notin A^t$, one has that $z^{t,\lambda} = (c^t)^\top (y_a^{t,\lambda})_{a \in A^t}$, and henceforth, by Theorem 1 along with (A-6), we have that $z^{t,\lambda} = z_R^{t,*}$, yielding the desired contradiction. ■

Proof of Lemma 2. As $z^{t,\lambda} < z_R^{t,*}$ there must exist $\tilde{c}^t \in \mathcal{U}^t$ such that

$$z^{t,\lambda} < (\tilde{c}^t)^\top (y_a^{t,\lambda})_{a \in A^t}.$$

Because $A^{t+1} = A^t$ we have that

$$\mathcal{U}^{t+1} = \{\hat{c}^t \in \mathbb{R}^{|A^t|} : (\hat{c}^t)^\top (y_a^{t,\lambda})_{a \in A^t} = z^{t,\lambda}, \hat{c}^t \in \mathcal{U}^t\},$$

and therefore $\tilde{c}^t \notin \mathcal{U}^{t+1}$.

Now, in view of equation above, $\mathbf{G}^{t+1} = (\mathbf{G}^t; (y^{t,\lambda})^\top)$ and $\mathbf{g}^{t+1} = (\mathbf{g}^t; z^{t,\lambda})$. For any $t \in \mathcal{T}$ let us denote by $C_U^{t,=}$ those inequalities in the definition of \mathcal{U}^t that must be satisfied as strict equalities, i.e.,

$$j \in C_U^{t,=} \Leftrightarrow \mathbf{G}_j^t \hat{c}^t = g_j \quad \forall \hat{c}^t \in \mathcal{U}^t,$$

where \mathbf{G}_j^t denotes j -th row of \mathbf{G}^t . Let us denote by $\mathbf{G}^{t,=}$ and $\mathbf{g}^{t,=}$ the corresponding submatrix and subvector of \mathbf{G}^t and \mathbf{g}^t associated with those elements in $C_U^{t,=}$. We have that (see e.g., Wolsey and Nemhauser (2014))

$$\dim(\mathcal{U}^t) = |A^t| - \text{rank}(\mathbf{G}^{t,=}, \mathbf{g}^{t,=}). \quad (\text{A-7})$$

We claim that $\text{rank}(\mathbf{G}^{t+1,=}, \mathbf{g}^{t+1,=}) \geq \text{rank}(\mathbf{G}^{t,=}, \mathbf{g}^{t,=}) + 1$, and the desired result then follows from equation (A-7). Indeed, arguing by contradiction, suppose that $\text{rank}(\mathbf{G}^{t+1,=}, \mathbf{g}^{t+1,=}) = \text{rank}(\mathbf{G}^{t,=}, \mathbf{g}^{t,=})$. This implies that $((y^{t,\lambda})_{a \in A^t}; z^{t,\lambda})^\top$ can be written as a linear combination of the rows of $(\mathbf{G}^{t,=}, \mathbf{g}^{t,=})$, and thus it is readily seen that

$$\{\hat{c}^t : \mathbf{G}^{t+1,=} \hat{c}^t = \mathbf{g}^{t+1,=}\} = \{\hat{c}^t : \mathbf{G}^{t,=} \hat{c}^t = \mathbf{g}^{t,=}\}.$$

Because $\tilde{c}^t \in \mathcal{U}^t$, it belongs to $\{\hat{c}^t : \mathbf{G}^{t,=} \hat{c}^t = \mathbf{g}^{t,=}\}$, which by the above equation implies that it also belongs to $\{\hat{c}^t : \mathbf{G}^{t+1,=} \hat{c}^t = \mathbf{g}^{t+1,=}\}$ and thus to \mathcal{U}^{t+1} , which yields the desired contradiction. ■

Appendix B: Proofs of the results for the matrix model from Section 4

Proof of Proposition 4. For (i), the leftmost inequality follows from the definition of both z^* and $z^{t,\lambda}$ (see equations (1) and (3)) and the fact that the feedback standard and Assumption **A3** holds. For the rightmost inequality, let x^* be an element of X that attains z^* . Partition x^* as $x^* = (\hat{x}, \tilde{x})$, where $\hat{x} = (x_i^*)_{i \in I^t}$ and $\tilde{x} = (x_i^*)_{i \in I \setminus I^t}$. Recall the partition of matrices given by (4). Therefore, because $x^* \in X$ and **A3** holds, one has that $\hat{x} \in X^t$. Now, suppose that $Y_E^t(\hat{x})$ is non-empty (if it is empty then it must be the case that $z_R^{t,*} = \infty$ and the result holds) and let \hat{y} be such that

$$\hat{y} \in \arg \min \{(\mathbf{c}^t)^\top y : y \in Y_E^t(\hat{x})\},$$

hence $(\mathbf{c}^t)^\top \hat{y} = z_E^t(\hat{x})$. By definition of $z_E^{t,*}$ we have that

$$(\mathbf{c}^t)^\top \hat{y} \leq z_E^{t,*}. \quad (\text{B-1})$$

Define \bar{y} as $\bar{y}_a := \hat{y}_a$ if $a \in A^t$, and $\bar{y}_a := 0$ if $a \in A \setminus A^t$. Because \mathcal{F} is standard, Assumption **A4** hold, and $\hat{y} \in Y_E^t(\hat{x})$, it follows that $\bar{y} \in Y(x^*)$. Therefore,

$$z^* \leq \mathbf{c}^\top \bar{y}. \quad (\text{B-2})$$

As $\mathbf{c}^\top \bar{y} = (\mathbf{c}^t)^\top \hat{y}$, equations (B-1) and (B-2) yield the desired result.

The proof of (ii) is the same as the proof of (ii) of Theorem 1. For (iii) note that $y^{t,\lambda} \in Y(x^{t,\lambda})$ and $y_a^t = 0$ for all $a \notin A^t$, hence it follows that

$$\sum_{a \in A^t} F_{da} y_a^{t,\lambda} + \sum_{i \in I^t} L_{di} x_i^{t,\lambda} \leq f_d \quad \forall d \in C_F^t. \quad (\text{B-3})$$

On the other hand, as \mathcal{U}^t has dimension zero, the set $Y_E^t(x^{t,\lambda})$ becomes

$$Y_E^t(x^{t,\lambda}) = \{y \in \mathbb{R}_+^{|A^t|} : \mathbf{F}^t y + \mathbf{L}^t x^{t,\lambda} \leq \mathbf{f}^t\},$$

and hence, from equation (B-3) it follows that $(y^{t,\lambda})_{a \in A^t} \in Y_E^t(x^{t,\lambda})$. Therefore, from part (i) and the definition of $z_E^{t,*}$ we have that

$$z^{t,\lambda} \leq z_E^{t,*} \leq (\mathbf{c}^t)^\top (y^{t,\lambda})_{a \in A^t}, \quad (\text{B-4})$$

but on the other hand, from the definition of $y^{t,\lambda}$ we have that

$$z^{t,\lambda} = (\mathbf{c}^t)^\top (y^{t,\lambda})_{a \in A^t}. \quad (\text{B-5})$$

Equations (B-4) and (B-5) imply that $z^{t,\lambda} = z_E^{t,*}$, and hence $\xi^\lambda \leq t$ as desired. The later part of the proposition is a consequence of the above result and part (ii). \blacksquare

Proof of Proposition 5. For (i), note that if $y_a^{t,\lambda} > 0$ for some $a \notin A^t$, then the result follows from the assumptions of Value-Perfect feedback. Therefore, suppose that $y_a^{t,\lambda} = 0$ for all $a \notin A^t$. We claim that there exists an activity $a \in A^t \setminus \tilde{A}^t$ such that $y_a^{t,\lambda} > 0$; the existence of such an activity implies the desired result. Indeed, to proceed by contradiction, suppose that this is not the case, i.e., $y_a^{t,\lambda} = 0$ for all $a \in A^t \setminus \tilde{A}^t$. As $y^{t,\lambda} \in Y(\bar{x}^{t,\lambda})$. Recalling the definition of $\bar{x}^{t,\lambda}$ in (3), this assumption implies that

$$\sum_{a \in \tilde{A}^t} F_{da} y_a^{t,\lambda} + \sum_{i \in I^t} L_{di} x_i^{t,\lambda} \leq f_d \quad \forall d \in C_F^t. \quad (\text{B-6})$$

Define $y' \in \mathbb{R}_+^{|A^t|}$ as $y'_a := y_a^{t,\lambda}$ for all $a \in A^t$. For any $\hat{\mathbf{F}} \in \mathcal{U}^t$, the vector y' satisfies

$$\sum_{a \in \tilde{A}^t} \hat{F}_{da} y'_a + \sum_{i \in I^t} L_{di} x_i^{t,\lambda} \leq f_d \quad \forall d \in C_F^t. \quad (\text{B-7})$$

Now, from the definition of \tilde{A}^t , $\hat{F}_{da} = F_{da}$ for all $a \in \tilde{A}^t$ and $\hat{\mathbf{F}} \in \mathcal{U}^t$, hence equations (B-6) and (B-7) imply that $y' \in Y_E^t(x^{t,\lambda})$. Therefore, from the definition of $z_E^{t,*}$ we have that

$$z_E^{t,*} \leq (\mathbf{c}^t)^\top y', \quad (\text{B-8})$$

but because it is readily checked that $z^{t,\lambda} = (\mathbf{c}^t)^\top y'$, and, moreover, that $z^{t,\lambda} \leq z_E^{t,*}$ by part (i) of Proposition 4, equation (B-8) implies that $z^{t,\lambda} = z_E^{t,*}$, yielding the desired contradiction.

The proof of (ii) is the same as the proof of Theorem 2. ■

Proof of Lemma 3. It is clear that because $z^{t,\lambda} < z_E^{t,*}$, it must be that $y^{t,\lambda} \notin Y_E^t(x^{t,\lambda})$. By definition of $Y_E^t(x^{t,\lambda})$, this means that there exist $\tilde{\mathbf{F}}^t \in \mathcal{U}^t$ and $d \in C_F^t$ such that $(\tilde{\mathbf{F}}^t)^\top_d y^{t,\lambda} > f_d - (\mathbf{L}^t)^\top_d x^{t,\lambda}$, as desired. ■

Before proceeding with the proof of Proposition 6, additional notation, concepts and results need to be introduced. In the discussion that follows, let us suppose that in Response-Perfect feedback, besides observing the values of y_a^t the leader is also able to observe the value of the left-hand side (or, equivalently, the slack q_d^t) for all constraints $d \in C_F^{t+1}$. For simplicity, let us denote $r_d^t := \sum_{a: y_a^t > 0} F_{da} y_a^t = f_d - q_d^t - \mathbf{L}_d^\top x^t$; by using the information from the feedback the leader updates \mathcal{U}^t by including the linear equalities

$$\sum_{a: y_a^t > 0} y_a^t \hat{F}_{da} = r_d^t \quad \text{for all } d \in C_F^{t+1}, \quad (\text{B-9})$$

in the definition of polyhedron \mathcal{U}^{t+1} . Recall that for any $d \in C_F^t$, n_d^t denotes the number of the follower's activities in A^t that d restricts, that is

$$n_d^t := |\{a \in A^t : d \in C_F(a)\}|.$$

For any given time $t \in \mathcal{T}$ we have that

$$\mathcal{U}^t \subseteq \mathbb{R}^{\sum_{d \in C_F^t} n_d^t}.$$

Suppose that $m^t = |C_F^t|$ and let us write $C_F^t = \{d_1, \dots, d_{m^t}\}$. We organize the elements of \mathcal{U}^t into blocks, so that $\hat{\mathbf{F}} \in \mathcal{U}^t$ is given by

$$\hat{\mathbf{F}} = [\hat{\mathbf{F}}^{d_1}; \hat{\mathbf{F}}^{d_2}; \dots; \hat{\mathbf{F}}^{d_{m^t}}],$$

where $\hat{\mathbf{F}}^d \in \mathbb{R}^{n_d^t}$ for all $d \in C_F^t$. We also assume that the columns of matrix \mathbf{G}^t are organized in this way. Using the conventions above, for any $d \in C_F^{t+1}$, constraint (B-9) can be rewritten as

$$\mathbf{v}_d^\top \hat{\mathbf{F}} = r_d^t, \quad (\text{B-10})$$

where vector \mathbf{v}_d is divided in subvectors as $\mathbf{v}_d := [\mathbf{v}_d^{d_1}; \mathbf{v}_d^{d_2}; \dots; \mathbf{v}_d^{d_{m^t+1}}]$, and each subvector $\mathbf{v}_d^{d_j} \in \mathbb{R}^{n_{d_j}^{t+1}}$. If $d \neq d_j$, then $\mathbf{v}_d^{d_j}$ is a vector of zeros, i.e., $\mathbf{v}_d^{d_j} = \mathbf{0}_{n_{d_j}^{t+1}}$. Otherwise, if $d = d_j$, then it has the information of $y^{t,\lambda}$ for those activities in A^{t+1} that are restricted by d , i.e., $(\mathbf{v}_d^d)_a = y_a^{t,\lambda}$ for all $a \in A^{t+1}$ such that $d \in C_F(a)$.

Let \mathcal{D}^0 and $\mathcal{D} \in \mathbb{G}(\mathcal{D}^0)$ be given, and suppose that T is sufficiently large. For any π , define $\mathcal{S}^\pi(\mathcal{D}^0, \mathcal{D}) := \{t \in \mathcal{T} : \exists a \notin A^t \text{ s.t. } y_a^t > 0\}$, that is, $\mathcal{S}^\pi(\mathcal{D}^0, \mathcal{D})$ is the set of time periods when at least a new activity is learned by the leader (who is using policy π). Suppose that $\mathcal{S}^\pi(\mathcal{D}^0, \mathcal{D}) = \{s_1, s_2, \dots, s_p\}$, where w.l.o.g. we suppose that $s_k < s_{k+1}$ for all $k \leq p-1$ (observe p depends on π , we drop it for the notation for simplicity). In addition, for any $k = 1, \dots, p$, define $N^k := \{a \in A \setminus A^{s_k} : y_a^{s_k} > 0\}$, i.e., N^k is the set of activities the leader learns by the end of time period s_k .

LEMMA 5. *Let $\lambda \in \Lambda$, suppose that feedback \mathcal{F} is Response-Perfect and that the leader observes the values of all the slack variables of the follower problem at any time $t \in \mathcal{T}$. If $\xi^\lambda > s_p$ then,*

$$\dim(\mathcal{U}^{t+1}) - \dim(\mathcal{U}^t) \leq \begin{cases} \sum_{a \in N^k} |C_F(a)| - \left| \bigcup_{a \in N^k} C_F(a) \right|, & \text{if } t = s_k \text{ for some } k \leq p, \\ -1, & \text{otherwise.} \end{cases} \quad (\text{B-11})$$

Proof. Let $k < p$ be given. Observe that at the end of period s_k the leader learns all the activities in N^k , and as such introduces a new variable \hat{F}_{da} into \mathcal{U}^{s_k+1} for all $d \in C_F(a)$ and $a \in N^k$; henceforth, \mathcal{U}^{s_k+1} has $\sum_{a \in N^k} |C_F(a)|$ more variables (columns) than \mathcal{U}^{s_k} (observe that there is no new variable \hat{F}_{da} for $a \in A^t$ from the standard feedback assumption). On the other hand, for every $d \in \bigcup_{a \in N^k} C_F(a)$ the leader includes the linear equality constraint (B-10) into \mathcal{U}^{s_k+1} (in addition to the potentially new constraints associated with each $d \in C_F^t$).

From the definition of \mathbf{v}_d in equation (B-10), it is readily seen that if $d \neq d'$, and both $d, d' \in \bigcup_{a \in N^k} C_F(a)$, then $(\mathbf{v}_d; r_d^{s_k})$ and $(\mathbf{v}_{d'}; r_{d'}^{s_k})$ are linearly independent. Moreover, it is also readily

observed that these vectors are also linearly independent of all the other (expanded) vectors that give equality constraints in \mathcal{U}^{s_k} .

The above analysis implies that, with respect to $\dim(\mathcal{U}^{s_k})$, $\dim(\mathcal{U}^{s_k+1})$ increases by $\sum_{a \in N^k} |C_F(a)|$ because of the new variables, but $\dim(\mathcal{U}^{s_k+1})$ reduces by (at least) $\left| \bigcup_{a \in N^k} C_F(a) \right|$, because of the newly introduced linearly independent equality constraints. In other words,

$$\dim(\mathcal{U}^{s_k+1}) \leq \dim(\mathcal{U}^{s_k}) + \sum_{a \in N^k} |C_F(a)| - \left| \bigcup_{a \in N^k} C_F(a) \right|. \quad (\text{B-12})$$

On the other hand, let $t < \xi^\lambda$ such that $t \notin \mathcal{S}^\lambda$, that is $y_a^t = 0$ for all $a \notin A^t$. Note that because $\xi^\lambda > t$ one has that $\mathbf{c}^\top y^{t,\lambda} < z_R^t$ (by part (i) of Proposition 4). We claim that (recall from the proof of Lemma 2 the definition of $\mathbf{G}^{t,=}$ and $\mathbf{g}^{t,=}$)

$$\text{rank}([\mathbf{G}^{t+1,=}, \mathbf{g}^{t,=}]) > \text{rank}([\mathbf{G}^{t,=}, \mathbf{g}^{t,=}]).$$

Indeed, because the assumptions of Lemma 3 hold, let $\tilde{\mathbf{F}}^t$ such that

$$\left(\tilde{\mathbf{F}}^t \right)_d^\top y^{t,\lambda} > f_d - (\mathbf{L}^t)_d^\top x^{t,\lambda}.$$

Now consider \mathcal{U}^t after adding the equation $\mathbf{v}_d^\top \hat{\mathbf{F}} = r_d^t$. Because $q_d^t \geq 0$, one has that $\tilde{\mathbf{F}}_d^\top y^{t,\lambda} > f_d - (\mathbf{L}^t)_d^\top x^t - q_d^t$ and hence $\tilde{\mathbf{F}} \notin \mathcal{U}^{t+1}$. Therefore, $\tilde{\mathbf{F}} \in \mathcal{U}^t \setminus \mathcal{U}^{t+1}$ and by the same arguments of Lemma 2 the vector $(\mathbf{v}_d; k_d)$ must be linearly independent from all the rows of $(\mathbf{G}^t, \mathbf{g}^t)$, thus the desired claim follows, and, we can conclude that

$$\dim(\mathcal{U}^{t+1}) \leq \dim(\mathcal{U}^t) - 1,$$

as desired. ■

LEMMA 6. *Let $\lambda \in \Lambda$ be given, suppose that feedback \mathcal{F} is Response-Perfect and that the leader observes the values of all the slack variables of the follower problem at any time $t \in \mathcal{T}$. With this, one has that $s_1 + \dim(\mathcal{U}^{s_1}) \leq \dim(\mathcal{U}^0)$, and*

$$s_{k+1} + \dim(\mathcal{U}^{s_{k+1}}) \leq s_k + \dim(\mathcal{U}^{s_k}) + 1 + \sum_{a \in N^k} |C_F(a)| - \left| \bigcup_{a \in N^k} C_F(a) \right| \quad k = 1, \dots, p-1.$$

Proof. By definition of s_1 , at periods $t = 0, 1, 2, \dots, s_1 - 1$ we have that the leader does not learn any activity and hence by Lemma 5 $\dim(\mathcal{U}^t) - \dim(\mathcal{U}^{t-1}) \leq -1$ for $t = 1, \dots, s_1$. This implies that $\dim(\mathcal{U}^{s_1}) \leq \dim(\mathcal{U}^0) - s_1$ and the result follows. Suppose that $k = 1, \dots, p-1$ is given. By definition of s_{k+1} , from $t = s_k + 1, \dots, s_{k+1} - 1$ the leader does not learn any activity and Lemma 5 again implies that $\dim(\mathcal{U}^t) - \dim(\mathcal{U}^{t-1}) \leq -1$, $t = s_k + 2, \dots, s_{k+1}$. This observation implies that

$$\dim(\mathcal{U}^{s_{k+1}}) \leq \dim(\mathcal{U}^{s_k+1}) - (s_{k+1} - s_k - 1).$$

Now, the above equation along with equation (B-12) imply that

$$\dim(\mathcal{U}^{s_{k+1}}) \leq \dim(\mathcal{U}^{s_k}) + \sum_{a \in N^k} |C_F(a)| - \left| \bigcup_{a \in N^k} C_F(a) \right| - s_{k+1} + s_k + 1,$$

which yields the desired result. \blacksquare

Using the above Lemma 6 we have the following important result.

LEMMA 7. *Let $\lambda \in \Lambda$ be given, suppose that feedback \mathcal{F} is Response-Perfect and that the leader observes the values of all the slack variables of the follower problem at any time $t \in \mathcal{T}$. One has that*

$$\tau^\lambda \leq \xi^\lambda \leq \dim(\mathcal{U}^0) + p + \sum_{k=1}^p \left(\sum_{a \in N^k} |C_F(a)| - \left| \bigcup_{a \in N^k} C_F(a) \right| \right). \quad (\text{B-13})$$

Proof. By repeated application of Lemma 6 it is verified that

$$s_p + \dim(\mathcal{U}^{s_p}) \leq \dim(\mathcal{U}^0) + (p-1) + \sum_{k=1}^{p-1} \left(\sum_{a \in N^k} |C_F(a)| - \left| \bigcup_{a \in N^k} C_F(a) \right| \right). \quad (\text{B-14})$$

Because by definition no new action is learned after s_p , $\dim(\mathcal{U}^t) - \dim(\mathcal{U}^{t-1}) \leq -1$ for $t \geq s_p + 2$. This implies that at most by time $s_p + \tilde{t}$, where $\tilde{t} := \sum_{a \in N^p} |C_F(a)| - \left| \bigcup_{a \in N^p} C_F(a) \right| + 1$, it must be the case that $\dim(\mathcal{U}^{s_p + \tilde{t}}) = 0$. Henceforth, part (iii) of Proposition 4 implies that $\xi^\lambda \leq s_p + \tilde{t}$, and hence equation (B-14) and the selection of t yield the desired result. \blacksquare

Proof of Proposition 6. Suppose first that (i) holds, i.e., that all the constraints are equality constraints, thus the leader always knows that their slack is zero. Hence, a direct application of Lemma 7 implies that

$$\tau^\lambda \leq \xi^\lambda \leq \dim(\mathcal{U}^0) + p + \sum_{k=1}^p \left(\sum_{a \in N^k} |C_F(a)| - \left| \bigcup_{a \in N^k} C_F(a) \right| \right).$$

The desired result follows by noting that $\sum_{k=1}^p \sum_{a \in N^k} |C_F(a)| = \sum_{a \in A \setminus A^0} |C_F(a)|$ and that $\left| \bigcup_{a \in N^k} C_F(a) \right| \geq 1$. On the other hand, consider (ii), i.e., that the leader observes the slack of one of the constraints in $D^{t,\lambda}$ at every period $t \in \mathcal{T}$ such that $y_a^t = 0$ for all $a \notin A^t$. In this case, following the same arguments as in Lemma 5, equation (B-11) can be simplified to:

$$\dim(\mathcal{U}^{t+1}) - \dim(\mathcal{U}^t) \leq \begin{cases} \sum_{a \in N^k} |C_F(a)|, & \text{if } t = s_k, \text{ for some } k \leq p, \\ -1, & \text{otherwise.} \end{cases}$$

The result follows from Lemma 7, after mimicking the proofs of the previous results, as in this case equation (B-13) becomes

$$\tau^\lambda \leq \xi^\lambda \leq \dim(\mathcal{U}^0) + p + \sum_{k=1}^p \sum_{a \in N^k} |C_F(a)|.$$

Proof of Proposition 7. Observe that as $z_R^t - \mathbf{c}^\top y^{t,\lambda} > 0$, thus $D^t \neq \emptyset$. Pick an arbitrary $d \in D^t$, and let $y^{t,*}$ be

$$y^{t,*} \in \arg \min \{ (\mathbf{c}^t)^\top y : y \in Y_E^t(x^{t,\lambda}) \},$$

hence $y^{t,*}$ is (one of) the solution(s) the leader expects from the follower after deciding $x^{t,\lambda}$. Now, let $\tilde{\mathbf{F}}$ be given by

$$\tilde{\mathbf{F}} \in \arg \max \{ (y^{t,*})^\top \hat{\mathbf{F}}_d : \hat{\mathbf{F}}_d \in \mathcal{U}^t \}$$

thus, $\tilde{\mathbf{F}}$ is (one of) the value(s) the leader assigns to the row d of the follower lower-level matrix by deciding robustly (i.e., by using policy λ). Observe that because $d \in D^t$ it follows that

$$\tilde{\mathbf{F}}_d^\top y^{t,\lambda} > f_d - \mathbf{L}_d^\top x^{t,\lambda},$$

and, henceforth, $\tilde{\mathbf{F}} \in \Delta^t$. Now, define the hyperplane \mathcal{P} as

$$\mathcal{P} := \{ \hat{\mathbf{F}} \in \mathbb{R}^{\sum_{d \in C_P^t} n_d} : \mathbf{v}_d^\top \hat{\mathbf{F}} = f_d - \mathbf{L}_d^\top x^{t,\lambda} \},$$

observe that \mathcal{U}^{t+1} is at one side of \mathcal{P} while $\tilde{\mathbf{F}}$ is at the other side, that is, \mathcal{P} separates \mathcal{U}^{t+1} and $\tilde{\mathbf{F}}$. Let $\mathbf{F}' \in \mathcal{U}^{t+1}$ be the closest point of \mathcal{U}^{t+1} to $\tilde{\mathbf{F}}$. It is clear that $\mathbf{F}' \in \mathcal{P}$, hence $\|\tilde{\mathbf{F}} - \mathbf{F}'\|$ is the *distance* from $\tilde{\mathbf{F}}$ to \mathcal{P} , which by standard linear algebra is given by the projection of $\tilde{\mathbf{F}} - \mathbf{F}'$ onto the vector \mathbf{v}_d ; that is

$$\|\tilde{\mathbf{F}} - \mathbf{F}'\| = \frac{(\tilde{\mathbf{F}} - \mathbf{F}')^\top \mathbf{v}_d}{\|\mathbf{v}_d\|} = \frac{(\tilde{\mathbf{F}}_d - \mathbf{F}'_d)^\top y^{t,\lambda}}{\|y^{t,\lambda}\|}.$$

Observe that

$$\begin{aligned} \|\tilde{\mathbf{F}} - \mathbf{F}'\| &= \frac{(\tilde{\mathbf{F}}_d - \mathbf{F}'_d)^\top y^{t,\lambda}}{\|y^{t,\lambda}\|} = \frac{(\tilde{\mathbf{F}}_d - \mathbf{F}'_d)^\top y^{t,\lambda} + \mathbf{F}'_d{}^\top y^{t,*} - \mathbf{F}'_d{}^\top y^{t,*}}{\|y^{t,\lambda}\|} \\ &= \frac{\mathbf{F}'_d{}^\top (y^{t,*} - y^{t,\lambda}) + \tilde{\mathbf{F}}_d^\top y^{t,\lambda} - \mathbf{F}'_d{}^\top y^{t,*}}{\|y^{t,\lambda}\|} \\ &\geq \frac{\mathbf{F}'_d{}^\top (y^{t,*} - y^{t,\lambda}) + \tilde{\mathbf{F}}_d^\top (y^{t,\lambda} - y^{t,*})}{\|y^{t,\lambda}\|} = \frac{(y^{t,*} - y^{t,\lambda})^\top (\mathbf{F}'_d - \tilde{\mathbf{F}}_d)}{\|y^{t,\lambda}\|}, \end{aligned} \quad (\text{B-15})$$

where the inequality follows because $\mathbf{F}' \in \mathcal{U}^t$. From the definition of $\tilde{\mathbf{F}}$ we have that

$$\mathbf{F}'_d{}^\top y^{t,*} \leq \tilde{\mathbf{F}}_d^\top y^{t,*}.$$

Now, let U be given by

$$U := \max_{\{a \in A^t : \mathbf{F}'_{d,a} - \tilde{\mathbf{F}}_{d,a} \neq 0\}} \left| \frac{y_a^{t,*} - y_a^{t,\lambda}}{\mathbf{F}'_{d,a} - \tilde{\mathbf{F}}_{d,a}} \right|,$$

therefore, if $K = (U + 1)U$, then the Diaz-Metcalf inequality (see e.g Diaz and Metcalf (1964)) implies that

$$(y^{t,*} - y^{t,\lambda})^\top (\mathbf{F}'_d - \tilde{\mathbf{F}}_d) \geq \|y^{t,*} - y^{t,\lambda}\|^2 - K \|\mathbf{F}'_d - \tilde{\mathbf{F}}_d\|^2. \quad (\text{B-16})$$

Inequalities in (B-15) and (B-16) imply that

$$\|\tilde{\mathbf{F}} - \mathbf{F}'\| \geq \frac{\|y^{t,*} - y^{t,\lambda}\|^2 - K\|\mathbf{F}'_d - \tilde{\mathbf{F}}_d\|^2}{\|y^{t,\lambda}\|},$$

and henceforth

$$K\|\tilde{\mathbf{F}} - \mathbf{F}'\|^2 + \|y^{t,\lambda}\|\|\tilde{\mathbf{F}} - \mathbf{F}'\| \geq \|y^{t,*} - y^{t,\lambda}\|^2.$$

Now, because $z_R^t - \mathbf{c}^\top y^{t,\lambda} = \mathbf{c}^\top (y^{t,*} - y^{t,\lambda}) > \epsilon$, the Cauchy-Schwartz inequality implies that

$$\|y^{t,*} - y^{t,\lambda}\| > \frac{\epsilon}{\|\mathbf{c}\|},$$

and hence

$$K\|\tilde{\mathbf{F}} - \mathbf{F}'\|^2 + \|y^{t,\lambda}\|\|\tilde{\mathbf{F}} - \mathbf{F}'\| \geq \epsilon^2 \|\mathbf{c}\|^{-2}.$$

Because $K \geq 0$, the above inequality implies the desired result. ■

Appendix C: Additional Results and Complementary Material

C.1. Semi-Oracle Algorithm

In this section we discuss an algorithm that speeds-up the solution of problem (13) and that is particularly useful to determine the semi-oracle decisions for instances where T is large. The algorithm works by computing a time-stability upper bound, which is constructed by forcing the follower to reveal an ‘optimal’ set of resources I^* as soon as possible. Once this upper bound is computed, MIP (13) is solved by truncating the time to T^0 , which, as it will be seen, can be bounded by the cardinality of I^* . Then, the optimal solution of the original MIP is obtained by extending the truncated solution until time T .

Before proceeding, we introduce some additional notation. Let x^* be an optimal solution of the full-information problem, and let $I^* := \{i \in I : x_i^* > 0\}$ be the set of resources that x^* uses. For any $J \subseteq I^*$ define $x^{*,J}$ as $x_i^{*,J} := x_i^*$ if $i \in J$ and zero otherwise, thus $x^{*,J}$ is the restriction of x^* to the resources in J . In addition, for any y define (with a slight abuse of notation)

$$I(y) := \bigcup_{a: y_a > 0} I(a)$$

i.e., $I(y)$ is the set of resources that interfere with the activities that y performs.

The computation of the upper bound T^0 is based on two following observations: (i) as soon as the semi-oracle enforces the follower to reveal all the resources in I^* , then she can implement the optimal solution x^* ; (ii) if for a given $J \subset I^*$ the semi-oracle implements $x^{*,J}$, then the response of the follower must reveal a new resource in $I^* \setminus J$, or else the response yields the optimal value z^* . While the proof of the first observation is straightforward, the proof of the second is a consequence of the following lemma.

LEMMA 8. *Let $J \subseteq I^*$ and suppose that $y^J \in \arg \min\{\mathbf{c}^\top y : y \in Y(x^{*,J})\}$. If $I(y^J) \cap I^* \subseteq J$, then $z^* \leq \mathbf{c}^\top y^J$.*

Proof. We proceed to prove that $y^J \in Y(x^*)$, note that if this holds, then $z^* \leq \mathbf{c}^\top y^J$ by the definition of z^* . Indeed, let $d \in C_F$ and note that

$$\begin{aligned} \sum_{a \in A} F_{da} y_a^J + \sum_{i \in I^*} L_{di} x_i^* &= \sum_{a \in A} F_{da} y_a^J + \sum_{i \in J} L_{di} x_i^* + \sum_{i \in I^* \setminus J} L_{di} x_i^* \\ &= \sum_{a \in A} F_{da} y_a^J + \sum_{i \in J} L_{di} x_i^* + \sum_{i \in K_1} L_{di} x_i^* + \sum_{i \in K_2} L_{di} x_i^*, \end{aligned} \quad (\text{C-1})$$

where in the last equation $K_1 = (I^* \setminus J) \cap I(y^J)$ and $K_2 = (I^* \setminus J) \setminus I(y^J)$. Our objective is to prove that the expression in equation (C-1) is at most f_d for all $d \in C_F$, from this the desired result follows.

First, suppose that $d \in C_F$ satisfies that $\sum_{a \in A} F_{da} y_a^J = 0$; then (C-1) is at most f_d by Assumption **A4**. Hence, suppose that $d \in C_F$ satisfies that $\sum_{a \in A} F_{da} y_a^J \neq 0$. Note that $K_1 = I^* \cap (I \setminus J) \cap I(y^J) = (I \setminus J) \cap (I(y^J) \cap I^*) = \emptyset$, because by hypothesis $I(y^J) \cap I^* \subseteq J$; therefore, $\sum_{i \in K_1} L_{di} x_i^* = 0$. On the other hand, suppose that $i \in K_2$. Then $i \notin I(y^J)$ and, since $\sum_{a \in A} F_{da} y_a^J \neq 0$, it must be the case that $L_{di} = 0$. As this holds for any $i \in K_2$, we have that $\sum_{i \in K_2} L_{di} x_i^* = 0$.

From the above observations, it follows that if $\sum_{a \in A} F_{da} y_a^J \neq 0$ then

$$\sum_{a \in A} F_{da} y_a^J + \sum_{i \in I^*} L_{di} x_i^* = \sum_{a \in A} F_{da} y_a^J + \sum_{i \in J} L_{di} x_i^* \leq f_d,$$

where the inequality in the above expression follows from the assumption that $y^J \in Y(x^{*,J})$. Thus, (C-1) is at most f_d for any $d \in C_F$ and hence $y^J \in Y(x^*)$, as desired. \blacksquare

Supported by the observations above, Algorithm 1 outputs an initial feasible solution. It starts by computing x^* and z^* . At any time t , it implements the solution x^{*,J^t} , with $J^t = I^* \cap I^t$. If the follower's solution at t yields a value less than z^* , then, per observation (ii), the semi-oracle can use a new resource in I^* at the next time period; otherwise, the solution implemented at t is optimal. The value of T^0 is set to be the first time that z^* is equal to the follower's cost. We note that T^0 is upper-bounded by $|I^*|$ since in at most $|I^*|$ periods the semi-oracle discovers all the resources in I^* , and once these resources are available, the solution of the semi-oracle is optimal, per observation (i). The above considerations are formalized in Lemma 9.

LEMMA 9. *Let T^0 be as computed by Algorithm 1. Then, T^0 is an upper bound on the optimal value of problem (13), and if $|I^* \setminus I^0| \leq T$, then $T^0 \leq |I^* \setminus I^0|$.*

Proof. First, if the algorithm outputs $T^0 = \infty$, the results holds trivially. Hence, suppose $T^0 < \infty$. In this case, it is readily checked that T^0 is an upper bound as the solution $\{(x^{*,J^t}, y^t) : t \in \mathcal{T}\}$ output by Algorithm 1 is feasible in (13) and yields an objective value of T^0 .

Algorithm 1 Finding an initial feasible solution to (13).

Require: $(\mathcal{D}^0, \mathcal{D}), T$ Compute x^* and z^*

$$J^0 = I^0 \cap I^*, y^0 \in \arg \min\{\mathbf{c}^\top y : y \in Y(x^*, J^0)\}, z^0 = \mathbf{c}^\top y^0$$

 $t = 0$ **while** $z^* > z^t$ and $t \leq T$ **do**

$$J^{t+1} = J^t \cup (I(y^t) \cap I^*)$$

$$y^{t+1} \in \arg \min\{\mathbf{c}^\top y : y \in Y(x^*, J^{t+1})\}, z^{t+1} = \mathbf{c}^\top y^{t+1}$$

 $t = t + 1$ **end while****if** $z^* = z^t$ **then**

$$T^0 = t, z^s = z^*, x^{*, J^s} = x^*, y^s = y^t \text{ for } s = t + 1, \dots, T$$

else

$$T^0 = \infty$$

end if**return** $T^0, z^*, \{(x^{*, J^t}, y^t) : t \in \mathcal{T}\}$

On the other hand, suppose that $|I^* \setminus I^0| \leq T$ and let $s \in \mathcal{T} \setminus \{0\}$ be given such that $z^* > z^r$ for all $r \leq s$. Because $J^s \subseteq I^*$, $y^s \in \arg \min\{\mathbf{c}^\top y : y \in Y(x^*, J^s)\}$, and $z^s = \mathbf{c}^\top y^s$, Lemma 8 implies that there exist $i \in I(y^s) \cap I^*$ such that $i \notin J^s$. Henceforth, $|J^{s+1} \setminus J^s| \geq 1$.

In order to arrive at a contradiction, suppose that $T^0 > |I^* \setminus I^0|$. This implies that if we let $t = |I^* \setminus I^0|$, then $z^* > z^s$ for all $s \leq t$, and,

$$|J^t| = |J^0| + \sum_{s=1}^{|I^* \setminus I^0|} |J^s \setminus J^{s-1}| \geq |J^0| + |I^* \setminus I^0| = |I^* \cap I^0| + |I^* \setminus I^0| = |I^*|. \quad (\text{C-2})$$

where the inequality follows as $|J^s \setminus J^{s-1}| \geq 1$ for all $s \leq t$. By construction, we have that $J^t \subseteq I^*$ for any t , thus inequality (C-2) implies that $J^t = I^*$, and hence, by observation (i) that $z^t = z^*$; which yields the desired contradiction. \blacksquare

By using Algorithm 1, an optimal solution of (13) can be readily computed via Algorithm 2. The correctness of Algorithm 2 follows from noting that T^0 is an upper bound for the time-stability. Hence, we have the following result, which we state without proof.

PROPOSITION 8. *Algorithm 2 correctly solves program (13).*

C.2. Numerical Computation of Policies in Λ

The following result establishes that $x^{t, \lambda}$ and $z_R^{t, *}$ can be computed by solving a system of linear equalities and inequalities with integer and continuous variables as long as $z_R^{t, *}$ is bounded.

Algorithm 2 Finding an optimal solution to (13)**Require:** $(\mathcal{D}^0, \mathcal{D}), T$ Compute $(T^0, z^*, \{(x^t, y^t) : t \in \mathcal{T}\})$ by calling **Algorithm 1** using $((\mathcal{D}^0, \mathcal{D}), T)$ **if** $T^0 \leq T$ **then**Solve program (13) until time T^0 passing $\{(x^t, y^t) : t = 0, \dots, T^0\}$ as an initial feasible solution,
let τ^* be the objective value**else**Solve program (13) until time T passing $\{(x^t, y^t) : t = 0, \dots, T\}$ as an initial feasible solution,
let τ^* be the objective value**if** $\tau^* = T + 1$ **then** $\tau^* = \infty$ **end if****end if****return** τ^*

LEMMA 10. Let $t \in \mathcal{T}$ be given and suppose that for all $x \in X^t$ problem $z_R^t(x)$ is bounded. Assume that there exist $y \in \mathbb{R}_+^{|A^t|}$, $p \in \mathbb{R}_+^{|C_U^t|}$, $\beta \in \mathbb{R}_+^{|C_F^t|}$, $\gamma \in \mathbb{R}^{|A^t|}$, $y_0 \in \mathbb{R}$, $u \in \{0, 1\}^{|C_F^t|}$, $v \in \{0, 1\}^{|A^t|}$, $w \in \{0, 1\}^{|C_U^t|}$ such that $x^* \in \mathbb{Z}_+^k \times \mathbb{R}_+^{|I^t| - k}$ satisfies

$$\mathbf{H}^t x^* \leq \mathbf{h}^t \quad \beta \leq \mathbf{M}^\beta u \quad (\text{C-3a})$$

$$y_0 - (\mathbf{g}^t)^\top p = 0 \quad \mathbf{f} - \mathbf{F}^t y - \mathbf{L}^t x^* \leq \mathbf{M}^{s(\beta)} (\mathbf{1} - u) \quad (\text{C-3b})$$

$$-y + (\mathbf{G}^t)^\top p = 0 \quad y \leq \mathbf{M}^y v \quad (\text{C-3c})$$

$$\mathbf{F}^t y + \mathbf{L}^t x^* \leq \mathbf{f}^t \quad \gamma + (\mathbf{F}^t)^\top \beta \leq \mathbf{M}^{s(y)} (\mathbf{1} - v) \quad (\text{C-3d})$$

$$-(\mathbf{F}^t)^\top \beta - \gamma \leq 0 \quad p \leq \mathbf{M}^p w \quad (\text{C-3e})$$

$$\mathbf{G}^t \gamma \leq \mathbf{g}^t \quad \mathbf{g}^t - \mathbf{G}^t \gamma \leq \mathbf{M}^{s(p)} (\mathbf{1} - w), \quad (\text{C-3f})$$

where in the above equations \mathbf{M}^β , $\mathbf{M}^{s(\beta)}$, \mathbf{M}^y , $\mathbf{M}^{s(y)}$, \mathbf{M}^p , $\mathbf{M}^{s(p)}$ are diagonal matrices whose elements are large enough numbers. Specifically, if $\mathcal{P} = \{(y, \beta, p) : \text{the constraints in the left column of (C-3a)-(C-3f) hold}\}$, then \mathbf{M}^β and $\mathbf{M}^{s(\beta)}$ satisfy that $\beta_d \leq \mathbf{M}_{dd}^\beta$ and $\mathbf{f}_d - \mathbf{F}_d^{t,T} y - \mathbf{L}_d^{t,T} x^* \leq \mathbf{M}_{dd}^{s(\beta)}$ for all $d \in C_F^t$ and $(y, \beta, p) \in \mathcal{P}$ (\mathbf{M}^y , $\mathbf{M}^{s(y)}$, \mathbf{M}^p and $\mathbf{M}^{s(p)}$ are defined analogously). Then, we have that $x^* \in \arg \max\{z_R^t(x) : x \in X^t\}$ and $y_0 = z_R^{t,*}$.

Proof. The optimization problem $\max\{z_R^t(x) : x \in X^t\}$ can be written as

$$\max_{x \in X^t} \min_{y_0, y} y_0 \quad (\text{C-4a})$$

$$\text{s.t. } (\hat{\mathbf{c}}^t)^\top y \leq y_0 \quad \forall \hat{\mathbf{c}}^t \in \mathcal{U}^t \quad (\text{C-4b})$$

$$-\mathbf{F}^t \mathbf{y} \geq \mathbf{L}^t x - \mathbf{f}^t \quad (\text{C-4c})$$

$$\mathbf{y} \geq 0 \quad (\text{C-4d})$$

Recall that $\mathcal{U}^t = \{\hat{\mathbf{c}}^t : \mathbf{G}^t \hat{\mathbf{c}}^t \leq \mathbf{g}^t\}$. The vector \mathbf{y} satisfies the robust constraint $(\hat{\mathbf{c}}^t)^\top \mathbf{y} \leq y_0 \forall \hat{\mathbf{c}}^t \in \mathcal{U}^t$ if and only if there exist $p \in \mathbb{R}_+^{|\mathcal{U}^t|}$ such that

$$(\mathbf{g}^t)^\top p \leq y_0 \text{ and } (\mathbf{G}^t)^\top p = \mathbf{y}$$

(see e.g., Ben-Tal et al. (2009)). Moreover, due to the objective function and to the fact that there are no other constraints on y_0 , in the optimal solution it must be the case that $(\mathbf{g}^t)^\top p = y_0$. Henceforth, problem (C-4) is equivalent to

$$\begin{aligned} & \max_{x \in X^t} \min_{y_0, \mathbf{y}} y_0 \\ & \text{s.t. } y_0 - (\mathbf{g}^t)^\top p = 0 \\ & \quad -\mathbf{y} + (\mathbf{G}^t)^\top p = 0 \\ & \quad -\mathbf{F}^t \mathbf{y} \geq \mathbf{L}^t x - \mathbf{f}^t \\ & \quad \mathbf{y} \geq 0 \end{aligned}$$

Because $z_R^t(x)$ is bounded for any $x \in X^t$, any optimal solution (y_0, \mathbf{y}) of the inner minimization problem satisfies its Karush-Kuhn-Tucker (KKT) optimality conditions (and vice-versa). Henceforth, replacing the minimization problem by the KKT conditions yields

$$\max_{x \in X^t} 0 \quad (\text{C-5a})$$

$$\text{s.t. } y_0 - (\mathbf{g}^t)^\top p = 0 \quad \mathbf{D}(\beta) \mathbf{D}(\mathbf{f} - \mathbf{F}^t \mathbf{y} - \mathbf{L}^t x) = \mathbf{0} \quad (\text{C-5b})$$

$$-\mathbf{y} + (\mathbf{G}^t)^\top p = 0 \quad \mathbf{D}(\mathbf{y}) \mathbf{D}(\gamma + (\mathbf{F}^t)^\top \beta) = \mathbf{0} \quad (\text{C-5c})$$

$$-\mathbf{F}^t \mathbf{y} \geq \mathbf{L}^t x - \mathbf{f}^t \quad \mathbf{D}(p) \mathbf{D}(\mathbf{g}^t - \mathbf{G}^t \gamma) = \mathbf{0} \quad (\text{C-5d})$$

$$-(\mathbf{F}^t)^\top \beta - \gamma \leq 0 \quad (\text{C-5e})$$

$$\mathbf{G}^t \gamma \leq \mathbf{g}^t \quad (\text{C-5f})$$

$$\mathbf{y} \geq 0, \beta \geq 0, p \geq 0, \gamma, y_0 \text{ free,} \quad (\text{C-5g})$$

where for any vector ℓ of n entries, $\mathbf{D}(\ell)$ is a diagonal matrix of size $n \times n$ whose j -th diagonal entry is given by ℓ_j . Observe that problem (C-5) is a non-linear feasibility problem (due to the non-linear complementary slackness constraints and because the objective function is constant). However, it can be linearized by introducing 0-1 variables. Indeed, β , \mathbf{y} and x satisfy the constraint

$D(\beta)D(\mathbf{f} - \mathbf{F}^t \mathbf{y} - \mathbf{L}^t \mathbf{x}) = \mathbf{0}$ if and only if there exists $u \in \{0, 1\}^{|C_F^t|}$ such that (see Audet et al. (1997))

$$\beta \leq \mathbf{M}^\beta u \text{ and } \mathbf{f} - \mathbf{F}^t \mathbf{y} - \mathbf{L}^t \mathbf{x} \leq \mathbf{M}^{s(\beta)}(\mathbf{1} - u).$$

A similar equivalence exists between the other two set of complementary slackness constraints and the remaining 0-1 linear constraints in (C-3a)-(C-3f). The desired result follows. \blacksquare

Lemma 10 implies that the decisions of policies in Λ can be computed by solving a mixed-integer linear program. Specifically, $x^{t,\lambda} = x$, and $z_R^{t,*} = y_0$, where x and y_0 are solutions of the system of equations (C-3a)-(C-3f).

C.3. Sequential Assignment Interdiction

We complement the example applications presented in Section 2 by modeling an interdiction assignment problem. Consider the problem discussed in Zenklusen (2010); here the enemy is the follower, who at each time has to assign each agent in a set V to exactly one job in a set W at minimum cost; assigning agent $v \in V$ to job $w \in W$ costs the follower c_{vw} . Define y_{vw} as 1 if v is assigned to w , and zero otherwise. The follower, absent the interventions of the leader, solves the following minimum-weighted matching (assignment) problem on the bipartite graph $G = (V \cup W, E)$, with $E := V \times W$:

$$y^* \in \arg \min_y \{ \mathbf{c}^\top y : \mathbf{M}^V y \leq \mathbf{1}, \mathbf{M}^W y \leq \mathbf{1}, -\mathbf{M}^W y \leq -\mathbf{1}, y \in \{0, 1\}^{|E|} \}.$$

In this formulation \mathbf{M}^V is a $|V| \times |E|$ (undirected) vertex-edge adjacency matrix, where $M_{v,(v,w)} = 1$ for all $v \in V$, and zero otherwise; similarly, \mathbf{M}^W is a $|W| \times |E|$ matrix where $M_{w,(v,w)} = 1$ for all $w \in W$, and $\mathbf{1}$ is a vector of ones. Observe that the constraints enforce $\mathbf{M}^W y = \mathbf{1}$, which means that each job must be taken by some agent. Also, note that while the above program is binary, the binary restrictions can be relaxed as the constraint matrix is totally unimodular, and hence it can be replaced by its linear programming relaxation.

The leader, on the other hand, has the ability to disable agents in V (the settings where she can disable assignments in E or jobs in W follow similar lines). Disabling agent v during each time period costs her b_v and she has a total budget of B at each period. Thus, if we let x_v take the value 1 if the leader disables v and zero otherwise, she faces the constraints $\sum_{v \in V} b_v x_v \leq B$, and $x_v \in \{0, 1\}$ for all $v \in V$ at each time period.

The above problem can be modeled within our framework as follows: the set of follower activities is E , thus $A = E$, and the set of follower constraints C_F consist of the restrictions regarding the assignment at each vertex, so $|C_F| = |V| + 2|W|$. It is readily seen that $\mathbf{F} = [\mathbf{M}^V; \mathbf{M}^W; -\mathbf{M}^W]$, $\mathbf{f} = [\mathbf{1}; \mathbf{1}; -\mathbf{1}]$, and that the cost vector \mathbf{c} is given by the assignment costs, thus $\mathbf{c} = (c_e : e \in E)$.

The leader resources are given by $I = V$ and C_L is a singleton consisting on the budgetary constraint, hence $\mathbf{H} = \mathbf{b}^\top$, where $\mathbf{b} = (b_v : v \in V)$ and $\mathbf{h} = B$. Matrix \mathbf{L} , on the other hand, has the agent-disabling constraints. Thus, $\mathbf{L} = (\mathbf{I}; \mathbf{0})$, where \mathbf{I} is a $|V| \times |V|$ identity matrix and $\mathbf{0}$ is a $2|W| \times |V|$ matrix of zeros.

Initially, the leader has knowledge about all the jobs W , but potentially ignores all possible agents as well as some of the possible assignments and their corresponding cost. For those assignments $A^0 \subseteq E$ she knows, she has interval estimates $\ell_e \leq c_e \leq m_e$, hence $\mathcal{U}^0 = \{\hat{\mathbf{c}}^0 \in \mathbb{R}^{A^0} : \ell_e \leq \hat{c}_e \leq m_e \forall e \in A^0\}$, thus $\mathbf{G}^0 = [\mathbf{I}; -\mathbf{I}]$ and $\mathbf{g}^0 = (\mathbf{m}; \boldsymbol{\ell})$, with $\mathbf{m} = (m_e : e \in A^0)$ and $\boldsymbol{\ell} = (\ell_e : e \in A^0)$.

Note that in this example the follower performs activity $e = (u, w) \in A$ whenever agent u is assigned to job w . The leader uses resource v if she disables agent v . The set $C_F(v, w)$ consist of the three constraints $\sum_{(v, w') \in E} y_{(v, w')} \leq 1$, $\sum_{(v', w) \in E} y_{(v', w)} \leq 1$ and $-\sum_{(v', w) \in E} y_{(v', w)} \leq -1$, while for any $v \in I$ there is only one constraint in $C_L(v)$ which corresponds to the budgetary constraint, thus $C_L(v) = C_L$ for all $v \in I$. For any $(v, w) \in A$, the set $I(v, w)$ corresponds to that agent whose disabling stops assignment of agent v to job w , i.e., $I(v, w) = \{v\}$.

In this example, standard feedback implies that at each time $t \in \mathcal{T}$ the leader always observes the cost incurred by the follower at time t . If the follower makes an assignment $(v, w) \in A$ that the leader did not observe before, then the leader learns that the assignment between agent v and job w is possible. Moreover, she learns $C_F(v, w)$, and as such, if agent v was never used before by the follower, she also learns about the existence of agent v . Also, the leader learns $I(v, w)$ and L_{di} for all $d \in C_F(a)$ and all $i \in I(a)$, and as such, she learns that by disabling agent v she can disable the assignment (v, w) . Finally, she also learns that disabling v costs her b_v .

Finally, in this setting, as the follower responses are binary, then by assumption **S2**, standard feedback is automatically Response-Perfect. On the other hand, in Value-Perfect feedback, the leader, besides observing the assignments, also observes the costs incurred by the follower when performing each of the assignments made at time $t \in \mathcal{T}$.