

Asymmetric-Information Resource Allocation Games: An LP Approach to Purposeful Deception

Longxu Pan^{1*}, Yue Guan¹, Daigo Shishika², and Panagiotis Tsiotras¹

¹ Georgia Institute of Technology, Atlanta GA 30332, USA

² George Mason University, Fairfax VA 22030, USA

Abstract. In this work, we introduce the Deceptive Resource Allocation Game (DRAG), which studies purposeful deception within a Bayesian game framework. In DRAG, a Defender allocates resources across the true asset along with several decoys to influence an Attacker’s beliefs and actions, with the intent to divert the Attacker away from the true asset. We seek to characterize *purposeful deception*, whereby the Defender deceives only when doing so improves its performance. To this end, we solve for the Perfect Bayesian Nash Equilibrium (PBNE) of the corresponding game. We show that, despite the coupled belief–policy interdependence, the problem admits an efficient, non-iterative linear programming formulation. Numerical results demonstrate that the resulting policies naturally balance performance optimality and belief manipulation, giving rise to purposeful and emergent deceptive behaviors.

Keywords: Multi-Agent System · Game Theory · Decision Making.

1 Introduction

Deception plays a fundamental role in domains such as military operations [9], security [1], and autonomous systems operating under adversarial influence [11]. Existing policies [3] optimize surrogate deception metrics at the expense of performance, leading to “deception for the sake of deception.” A common framework that balances deception and performance is asymmetric-information games, where agents possess unequal knowledge about the underlying environment and strategically manipulate observable signals to influence an opponent’s belief and decision-making.

A well-studied line of work includes *goal recognition* (GR) and *deceptive path planning* (DPP). GR [13, 14, 18] focuses on inferring an agent’s intent from observed trajectories, while DPP [6, 15, 16] studies how an agent can reach its goal while misleading an observer about its intent. Recent DPP works [15, 17] address the limitation of surrogate deception metrics [3] using a Bayesian game formulation. In these settings, an Attacker is privately assigned a goal from a set of candidates, while the Defender seeks to identify the true goal by observing the Attacker’s trajectory and allocate resources accordingly. However, existing approaches are limited to settings where the informed agent fully controls the observable signal and, in some cases, to only two candidate goals.

* Corresponding author.

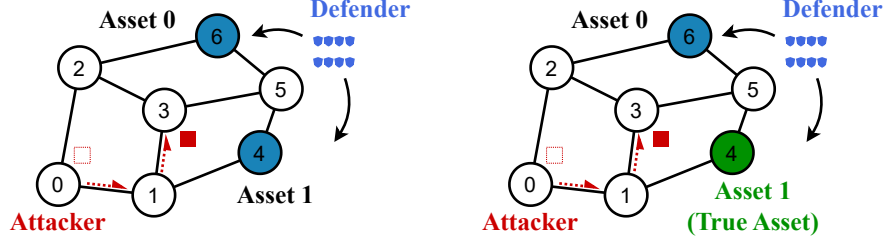


Fig. 1. An example of DRAG from the perspectives of the Attacker (*left*) and Defender (*right*).

In this paper, we study the Deceptive Resource Allocation Game (DRAG), a natural *dual* of DPP, in which a Defender, informed of the true asset, allocates limited resources across candidate assets to mislead the Attacker (see Fig. 1). Similar in spirit to DPP, the Defender in DRAG must balance two competing objectives: allocating resources to the true asset to optimize performance and allocating resources to decoys to manipulate the opponent’s belief. This trade-off gives rise to *purposeful* deception, where deception is employed only when it improves the Defender’s performance, instead of merely inducing uncertainty or bias to the Attacker’s belief.

In DRAG, the Defender influences the Attacker’s belief via its allocation actions, while the Attacker controls the game state and, in turn, influences the Defender’s allocations. Consequently, the observable signal arises from the interaction among agents, thereby coupling belief and strategic behavior in a way absent from prior DPP formulations. Thus, unlike DPP [15, 16], the observable signal in DRAG is *jointly controlled* by both players. To formalize this interaction, we adopt a Bayesian game framework and solve for Perfect Bayesian Nash Equilibria (PBNE), in which agents optimize performance while accounting for belief manipulation. Prior works [5, 8] have proposed linear programming (LP) formulations for computing PBNE when the informed player *fully controls* the observable signal. Despite the jointly-controlled signal, we show that DRAG still admits an LP formulation on the game tree.

Contributions: The contributions of this work are threefold. First, we introduce the DRAG problem, a novel dual formulation of the GR and DPP problems. Second, we show that the informed player (the Defender) can influence the uninformed player’s belief (the Attacker) by shaping the distribution of observable allocation histories, thereby affecting the Attacker’s decisions. Third, we present an efficient LP formulation on the DRAG game tree, enabling the computation of PBNE. Our LP approach is general and applies to arbitrary graphs with multiple candidate goals, extending prior DPP work [15, 16], which is restricted to two-goal settings. Numerical results demonstrate that the resulting PBNE policies exhibit purposeful deceptive behaviors by balancing deception and performance.

The notation used in this paper is fairly standard. $\Delta(E)$ denotes the set of probability distributions over a finite set E . For a finite set Θ , $[f(\theta)]_{\theta \in \Theta}$ denotes the vector in $\mathbb{R}^{|\Theta|}$ whose entries are given by $f(\theta)$ for each $\theta \in \Theta$. Unless otherwise

specified, we abbreviate it as $[f(\theta)]_\theta$. We use $[0:T]$ to denote the discrete set $\{0, 1, \dots, T\}$. For a random variable X and its realization x , the probability of the event $\{X = x\}$ is denoted as $\mathcal{P}(X = x)$. When the random variable is clear from the context, we use the shorthand $\mathcal{P}(x)$.

2 Problem Formulation

The Deceptive Resource Allocation Game (DRAG) is a discrete-time dynamic zero-sum game between a Defender and an Attacker over a graph $\mathcal{G} = (\mathcal{S}, \mathcal{E})$, where $\mathcal{S}$ and $\mathcal{E}$ are the node set and the edge set, respectively. Let $\Theta \subseteq \mathcal{S}$ denote a set of candidate asset nodes. At the start of the game, Nature selects an asset $\theta_{\text{true}} \in \Theta$ according to the prior $\mathbf{b}_0 \in \Delta(\Theta)$ and privately reveals it to the Defender, who aims to protect it, without disclosing to the Attacker. The Defender is of type θ if the candidate asset θ is selected by Nature.

At each time step, the Attacker moves on the graph $\mathcal{G}$, while the Defender allocates resources to one candidate asset in Θ . The Defender aims to maximize the resources accumulated at θ_{true} , while the Attacker seeks to minimize the same. By allocating to decoy assets, the Defender can strategically mislead the Attacker's inference of θ_{true} , thereby delaying its progress toward the true asset and gaining additional time to reinforce it. Conversely, the Attacker must infer θ_{true} from the observed allocation history and reach it within a prescribed horizon T , while minimizing the resources accumulated to θ_{true} . Formally, the DRAG problem is defined as

$$\text{DRAG} = \langle \mathcal{S}, \mathcal{E}, \Theta, r, m, s_0, \mathbf{b}_0, T \rangle,$$

whose components are described in detail below.

State and Action Spaces: The game state is given by the Attacker position $s_t \in \mathcal{S}$. At time t , the Attacker selects an edge $u_t = (s_t, s_{t+1}) \in \mathcal{E}$ and attempts to move to the next node s_{t+1} . Thus, the state space is $\mathcal{S}$, and the Attacker action space at node s is $\mathcal{U}(s)$, the set of outward edges at s . The Defender action $v_t \in \mathcal{V}$ specifies the asset node to which resources are allocated at time t . Since the Defender allocates to only one candidate asset at each time step, its action space $\mathcal{V}$ coincides with Θ . The edge weight $w(u_t) > 0$ represents the amount of resources the Defender can allocate while the Attacker traverses edge u_t ³.

Dynamics: At the start of the game, Nature selects the true asset θ_{true} from the candidate set Θ according to a commonly known prior $\mathbf{b}_0 \in \Delta(\Theta)$. The selected asset is privately revealed to the Defender. With θ_{true} fixed throughout the game, the Attacker starts from an initial state $s_0 \in \mathcal{S}$. All assets are initially unallocated.

At time t , both players act *simultaneously*. After the Attacker takes action $u_t \in \mathcal{U}(s_t)$ at s_t , it transitions to a neighboring node s_{t+1} with probability

³ If edge weights represent travel time, then $w(u_t)$ corresponds to the amount of resources allocated by the Defender with a unit allocation rate.

$\mathcal{P}(s_{t+1} | s_t, u_t)$. Concurrently, the Defender takes action $v_t \in \mathcal{V}$, thereby allocating resources to asset v_t . The game terminates when the Attacker reaches *any* asset in Θ or when the horizon T is reached. For notational simplicity, we use the shorthand $\mathcal{P}_{s_t, u_t}^{s_{t+1}}$ to denote $\mathcal{P}(s_{t+1} | s_t, u_t)$.

Information Structure: All game parameters are common information to both players. However, the selected asset θ_{true} is only revealed to the Defender, leading to an information asymmetry as illustrated in Fig 1. The Attacker therefore forms a belief $\mathbf{b}_t \in \Delta(\Theta)$ about the true asset from the observed game history. At time t , both players possess *perfect recall* of the game history $h_t = (s_0, u_0, v_0, \dots, s_{t-1}, u_{t-1}, v_{t-1}, s_t)$, containing the Attacker's state-action history and the Defender's action history [4]. Let $\mathcal{H}_t$ denote the admissible game history set, and let $|\mathcal{H}_t|$ denote its cardinality.

Behavioral Strategies: Since DRAG is a finite extensive-form game with perfect recall, Nash's theorem and Kuhn's theorem guarantee the existence of a Nash equilibrium in behavioral strategies [10]. Let $\sigma \in \Sigma$ and $\tau \in \mathcal{T}$ denote the behavioral strategies of the Attacker and Defender, respectively. Because the Defender observes its type θ , its behavioral strategy is type-dependent and can be written as $\tau = \{\tau^\theta\}_{\theta \in \Theta}$. At time t , the Defender's policy τ_t maps the history h_t and its type θ to an action distribution $\tau_{h_t}^\theta \in \Delta(\mathcal{V})$, where $\tau_{h_t}^\theta(v_t)$ denotes the probability that Defender selects action v_t given its type θ and the observed history h_t . Since the Defender's type (i.e., the selected asset) is unobservable to the Attacker, the Attacker's policy σ_t only takes the history h_t as input and maps to an action distribution $\sigma_{h_t} \in \Delta(\mathcal{U}(s_t))$, where $\sigma_{h_t}(u_t)$ represents the probability that Attacker selects action u_t given the observed history h_t .

Reward Structure: We consider a general reward function $r(h_t, \theta, u_t, v_t)$ that assigns a running reward at time t based on the history, the game type, and the actions of both agents. The reward is not revealed to the agents during the game and is used only for performance evaluation. Therefore, it cannot serve as a signal for inferring the true asset. For notational simplicity, we introduce the reward matrix $R_{h_t}^\theta \in \mathbb{R}^{|\mathcal{U}(s_t)| \times |\mathcal{V}|}$, which captures the rewards corresponding to different action pairs.

2.1 Optimization Objectives

For each type $\theta \in \Theta$, the Defender optimizes its (type-dependent) strategy $\tau^\theta = \{\tau_{h_t}^\theta\}_{t=0}^T$ to maximize the following expected accumulated payoff of the type- θ game,

$$V^\theta(h_0, \sigma, \tau^\theta) = \mathbb{E}_{s_0, \theta}^{\sigma, \tau^\theta} \left[\sum_{t=0}^T r(h_t, \theta, u_t, v_t) \right], \quad (1)$$

where the superscript indicates that the trajectory distribution is induced by (σ, τ^θ) , while the subscript indicates that the game is initialized at s_0 with the type fixed at θ .

Under the zero-sum game assumption and the asymmetric information structure, the Attacker optimizes a *single* strategy $\sigma = \{\sigma_{h_t}\}_{t=0}^T$ that minimizes the ex-ante value,

$$V(h_0, \mathbf{b}_0, \sigma, \tau) = \mathbb{E}_{s_0, \mathbf{b}_0}^{\sigma, \tau} \left[\sum_{t=0}^T r(h_t, \theta, u_t, v_t) \right]. \quad (2)$$

Here, $\mathbb{E}_{s_0, \mathbf{b}_0}^{\sigma, \tau}$ denotes expectation under the strategy profile (σ, τ) , with initial state s_0 and θ drawn according to the prior $\mathbf{b}_0$. We denote the type- θ and the ex-ante continuation values as $V^\theta(h_t, \sigma, \tau^\theta)$ and $V(h_t, \mathbf{b}_t, \sigma, \tau)$, respectively.

2.2 Perfect Bayesian Nash Equilibrium

We now formally introduce the *Perfect Bayesian Nash Equilibrium* (PBNE), where the Defender maximizes its type-dependent values, and the Attacker minimizes the ex-ante value. The PBNE solution concept consists of two parts: *belief consistency* and *policy sequential rationality* [12, 19]. To formalize belief consistency, we first define the Attacker's belief about the true asset.

Definition 1 (Belief). For $\theta \in \Theta$ and $t \in [0 : T]$, the belief $b_t(\theta)$ is the posterior probability that θ is the true asset given the observed history h_t , i.e.,

$$b_t(\theta) := \mathcal{P}^\tau(\theta \mid h_t),$$

where the superscript τ emphasizes the dependence on the Defender's strategy.

At each time t , the uninformed Attacker updates its belief according to the Bayesian update rule

$$b_{t-1}(\theta) \tau_{h_{t-1}}^\theta(v_{t-1}) / \sum_{\theta'} b_{t-1}(\theta') \tau_{h_{t-1}}^{\theta'}(v_{t-1}). \quad (3)$$

Since the Defender's action v_{t-1} is sampled from the type-dependent policy $\tau_{h_{t-1}}^\theta$, it serves as the signal for inferring the true asset. The numerator of (3) is the joint probability of the Defender being type θ and selecting the observed action v_{t-1} , while the denominator is the total probability of observing v_{t-1} across all types. Thus, (3) is the posterior that the Defender is of type θ given v_{t-1} .

The following assumption and the resulting proposition establish that (3) yields the posterior belief $b_t(\theta)$ as defined in Definition 1.

Assumption 1 (Conditional Independence) For $t \in [1 : T]$, conditioned on the observed history h_t , the belief over the true asset is independent of the Attacker's immediate action-state pair (u_{t-1}, s_t) , i.e.,

$$b_t(\theta) := \mathcal{P}^\tau(\theta \mid \underbrace{h_{t-1}, u_{t-1}, v_{t-1}, s_t}_{h_t}) = \mathcal{P}^\tau(\theta \mid h_{t-1}, v_{t-1}),$$

for all $(u_{t-1}, s_t) \in \mathcal{U}(s_{t-1}) \times \mathcal{S}$.

Proposition 1. *Under Assumption 1, the Bayesian update rule (3) coincides with $b_t(\theta)$, i.e.,*

$$b_t(\theta) = \frac{b_{t-1}(\theta) \tau_{h_{t-1}}^\theta(v_{t-1})}{\sum_{\theta'} b_{t-1}(\theta') \tau_{h_{t-1}}^{\theta'}(v_{t-1})}.$$

Proof. The result follows from the interpretation of (3) as $\mathcal{P}^\tau(\theta \mid h_{t-1}, v_{t-1})$, which equals to $b_t(\theta)$ under Assumption 1. See Appendix A.1 for details.

The PBNE is formally defined as follows.

Definition 2. *A strategy profile (σ^*, τ^*) constitutes a PBNE if the following holds for all $t \in [0 : T]$ and all histories $h_t \in \mathcal{H}_t$ with non-zero probability,*

1. *Sequential rationality:*

$$\begin{aligned} V(h_t, \mathbf{b}_t, \sigma^*, \tau^*) &\leq V(h_t, \mathbf{b}_t, \sigma, \tau^*), & \forall \sigma \in \Sigma, \\ V^\theta(h_t, \sigma^*, \tau^{\theta*}) &\geq V^\theta(h_t, \sigma, \tau^\theta), & \forall \theta \in \Theta, \tau^\theta \in \mathcal{T}^\theta. \end{aligned}$$

2. *Belief consistency: the belief $\mathbf{b}_t$ is propagated according to the Bayesian update rule in (3).*

For brevity, we use $V^*(h_t, \mathbf{b}_t)$ to denote $V(h_t, \mathbf{b}_t, \sigma^*, \tau^{\theta*})$. Similarly, $V^{\theta*}(h_t)$ denotes $V^\theta(h_t, \sigma^*, \tau^{\theta*})$.

3 Dynamic Programming for Value Functions

A key challenge in DRAG is the coupling between belief updates and strategic behaviors. To address this, we first establish a dynamic programming (DP) representation of the ex-ante value $V(h_t, \mathbf{b}_t, \sigma, \tau)$. Building on this structure, we show that the equilibrium game value $V^*(h_t, \mathbf{b}_t)$ admits a min-max DP characterization. These results enable a sequence of duality-based reformulations on $V^*(h_t, \mathbf{b}_t)$, through which we construct an LP over the game tree [7].

Given the history $h_t \in \mathcal{H}_t$ and belief $\mathbf{b}_t$ at time t , the strategy profile (σ, τ) and the transition dynamics induce a probability distribution over future game trajectories. Under this induced distribution, the type- θ continuation value $V^\theta(h_t, \sigma, \tau^\theta)$ and the corresponding ex-ante value $V(h_t, \mathbf{b}_t, \sigma, \tau)$ admit the following DP representations

$$V^\theta(h_t, \sigma, \tau^\theta) = \sigma_{h_t}^\top R_{h_t}^\theta \tau_{h_t}^\theta + \sum_{u_t, v_t, s_{t+1}} \sigma_{h_t}(u_t) \mathcal{P}_{s_t, u_t}^{s_{t+1}} \tau_{h_t}^\theta(v_t) V^\theta(h_{t+1}, \sigma, \tau^\theta), \quad (4)$$

$$\begin{aligned} V(h_t, \mathbf{b}_t, \sigma, \tau) &= \sum_{\theta} b_t(\theta) \sigma_{h_t}^\top R_{h_t}^\theta \tau_{h_t}^\theta \\ &\quad + \underbrace{\sum_{u_t, v_t, s_{t+1}} \sigma_{h_t}(u_t) \sum_{\theta} \mathcal{P}_{s_t, u_t}^{s_{t+1}} [b_t(\theta) \tau_{h_t}^\theta(v_t) V^\theta(h_{t+1}, \sigma, \tau^\theta)]}_{A}. \end{aligned} \quad (5)$$

In particular, the continuation term $V^\theta(h_{t+1}, \sigma, \tau^\theta)$ in (4) and (5) denotes the expected accumulated payoff from stage $t+1$ to horizon T under type θ .

The terminal values are given by

$$V^\theta(h_T, \sigma, \tau^\theta) = \sigma_{h_T}^\top R_{h_T}^\theta \tau_{h_T}^\theta, \quad (6)$$

$$V(h_T, \mathbf{b}_T, \sigma, \tau) = \sum_\theta b_T(\theta) \sigma_{h_T}^\top R_{h_T}^\theta \tau_{h_T}^\theta. \quad (7)$$

Note that (5) is not a DP in terms of the ex-ante value V alone due to the presence of V^θ in A . To resolve this, we leverage the following properties of ex-ante values from (4)-(7).

Remark 1. The ex-ante value is the expectation of the type- θ values, that is,

$$V(h_t, \mathbf{b}_t, \sigma, \tau) = \sum_\theta b_t(\theta) V^\theta(h_t, \sigma, \tau^\theta). \quad (8)$$

Remark 2. The ex-ante value is positively homogeneous in the belief, i.e., for $\alpha \geq 0$,

$$V(h_t, \alpha \mathbf{b}_t, \sigma, \tau) = \alpha V(h_t, \mathbf{b}_t, \sigma, \tau). \quad (9)$$

Rearranging the coefficients of the term A in (5), we have

$$\begin{aligned} A &= \sum_\theta \mathcal{P}_{s_t, u_t}^{s_{t+1}} \xi \left[\frac{b_t(\theta) \tau_{h_t}^\theta(v_t)}{\xi} V^\theta(h_{t+1}, \sigma, \tau^\theta) \right] \\ &= \sum_\theta \mathcal{P}_{s_t, u_t}^{s_{t+1}} \xi [b_{t+1}(\theta) V^\theta(h_{t+1}, \sigma, \tau^\theta)], \end{aligned}$$

where $\xi = \sum_{\theta'} b_t(\theta') \tau_{h_t}^{\theta'}(v_t)$ is the denominator in (3). We use Remarks 1-2 to relate the type- θ value in term A with the ex-ante value, yielding the following DP in terms of the ex-ante value V alone:

$$V(h_t, \mathbf{b}_t, \sigma, \tau) = \sum_\theta b_t(\theta) \sigma_{h_t}^\top R_{h_t}^\theta \tau_{h_t}^\theta + \sum_{u_t, v_t, s_{t+1}} \left[\sigma_{h_t}(u_t) V(h_{t+1}, \alpha \mathbf{b}_{t+1}, \sigma, \tau) \right], \quad (10)$$

where $\alpha = \mathcal{P}_{s_t, u_t}^{s_{t+1}} \xi$. Since (10) is derived under the Bayesian belief update rule, the belief in the recursion is consistent with the underlying (σ, τ) . Thus, the PBNE characterization reduces to enforcing sequential rationality on (10), yielding the min-max DP characterization in (11),

$$V^*(h_t, \mathbf{b}_t) = \min_{\sigma_{h_t}} \max_{\tau_{h_t}} \left\{ \sum_\theta b_t(\theta) \sigma_{h_t}^\top R_{h_t}^\theta \tau_{h_t}^\theta + \sum_{u_t, v_t, s_{t+1}} \sigma_{h_t}(u_t) V^*\left(h_{t+1}, \underbrace{\left[\mathcal{P}_{s_t, u_t}^{s_{t+1}} b_t(\theta') \tau_{h_t}^{\theta'}(v_t) \right]_{\theta'}}_{\alpha \mathbf{b}_{t+1}} \right) \right\} \quad (11a)$$

$$= \min_{\sigma_{h_t}} \left\{ \max_{\tau_{h_t}} \underbrace{\left\{ \sigma_{h_t}^\top \left\{ \sum_\theta b_t(\theta) R_{h_t}^\theta \tau_{h_t}^\theta + \left[\sum_{v_t, s_{t+1}} V^*(h_{t+1}, [\mathcal{P}_{s_t, u_t}^{s_{t+1}} b_t(\theta') \tau_{h_t}^{\theta'}(v_t)]_{\theta'}) \right]_{u_t} \right\}}_{B(\sigma_{h_t})} \right\} \right\}, \quad (11b)$$

where the scaled belief $\alpha \mathbf{b}_{t+1}$ is written explicitly in (11a) for clarity. The max-min DP follows by symmetry.

4 Defender's Linear Program

We now construct the LP that computes the Defender's equilibrium strategy τ^* . The key idea is to introduce variables that encode the joint effect of belief and Defender policies [8], thereby eliminating the nonlinear coupling between belief updates and policy choices.

4.1 Single-stage LP

We begin with the single-stage case. Given $(h_T, \mathbf{b}_T)$, the DRAG reduces to a static Bayesian game whose value $V^*(h_T, \mathbf{b}_T)$ is given by the following LP:

$$\begin{aligned} V^*(h_T, \mathbf{b}_T) &= \max_{\boldsymbol{\tau}_{h_T}, \ell_{h_T}} \ell_{h_T} \\ \text{s.t. } \quad &\mathbf{1}^\top \boldsymbol{\tau}_{h_T}^\theta = 1, \quad \boldsymbol{\tau}_{h_T}^\theta \geq \mathbf{0}, \quad \forall \theta \in \Theta, \end{aligned} \quad (12a)$$

$$\ell_{h_T} \mathbf{1} \leq \sum_{\theta} R_{h_T}^\theta b_T(\theta) \boldsymbol{\tau}_{h_T}^\theta, \quad (12b)$$

where the scalar ℓ_{h_T} represents the worst-case ex-ante value that the Defender can achieve under belief $\mathbf{b}_T$. Note that if $\mathbf{b}_T$ is scaled by a nonnegative scalar α , the LP remains valid, with $b_T(\theta)$ replaced by the scaled belief $\alpha b_T(\theta)$ in (12b).

4.2 Two-stage LP

We now consider the two-stage DRAG. Notice in (11b) that the maximization over the Defender's policy can be isolated as an inner problem for a fixed Attacker policy. The term $B(\boldsymbol{\sigma}_{h_{T-1}})$ can be reformulated as an LP, dualized, and then embedded back into the overall optimization.

Under (11), the equilibrium value $V^*(h_{T-1}, \mathbf{b}_{T-1})$ is given by a min-max problem that involves the continuation value $V^*(h_T, \alpha \mathbf{b}_T)$ at time T . By the single-stage result (12), $V^*(h_T, \alpha \mathbf{b}_T)$ admits the following optimization:

$$\begin{aligned} V^*(h_T, \alpha \mathbf{b}_T) &= \max_{\boldsymbol{\tau}_{h_T}, \ell_{h_T}} \ell_{h_T} \\ \text{s.t. } \quad &\mathbf{1}^\top \boldsymbol{\tau}_{h_T}^\theta = 1, \quad \boldsymbol{\tau}_{h_T}^\theta \geq \mathbf{0}, \quad \forall \theta \in \Theta, \\ &\ell_{h_T} \mathbf{1} \leq \sum_{\theta} R_{h_T}^\theta \underbrace{\mathcal{P}_{s_{T-1}, u_{T-1}}^{s_T} b_{T-1}(\theta) \tau_{h_{T-1}}^\theta(v_{T-1})}_{\alpha b_T(\theta)} \boldsymbol{\tau}_{h_T}^\theta. \end{aligned} \quad (13)$$

However, (13) cannot be solved independently from the optimization at $T-1$, since α contains $\tau_{h_{T-1}}^\theta(v_{T-1})$ which is a decision variable at $T-1$. Given a fixed $\boldsymbol{\sigma}_{h_{T-1}}$, the term $B(\boldsymbol{\sigma}_{h_{T-1}})$ in (11b) can be reformulated as the following optimization problem by incorporating (13)

$$\begin{aligned} B(\boldsymbol{\sigma}_{h_{T-1}}) &= \max_{\substack{\boldsymbol{\tau}_{h_{T-1}}, \ell_{h_{T-1}}, \forall h_T}} \boldsymbol{\sigma}_{h_{T-1}}^\top \left\{ \sum_{\theta} R_{h_{T-1}}^\theta b_{T-1}(\theta) \boldsymbol{\tau}_{h_{T-1}}^\theta + \left[\sum_{v_{T-1}, s_T} \ell_{h_T} \right]_{u_{T-1}} \right\} \\ \text{s.t. } \quad &\mathbf{1}^\top \boldsymbol{\tau}_{h_{T-1}}^\theta = 1, \quad \boldsymbol{\tau}_{h_{T-1}}^\theta \geq \mathbf{0}, \quad \forall \theta \in \Theta, \end{aligned} \quad (14a)$$

$$\forall h_T \in \mathcal{H}_T,$$

$$\mathbf{1}^\top \boldsymbol{\tau}_{h_T}^\theta = 1, \quad \boldsymbol{\tau}_{h_T}^\theta \geq \mathbf{0}, \quad \forall \theta \in \Theta, \quad (14b)$$

$$\ell_{h_T} \mathbf{1} \leq \sum_{\theta} R_{h_T}^\theta \mathcal{P}_{s_{T-1}, u_{T-1}}^{s_T} b_{T-1}(\theta) \tau_{h_{T-1}}^\theta(v_{T-1}) \boldsymbol{\tau}_{h_T}^\theta. \quad (14c)$$

Notice that the constraint (14c) involves products between $\tau_{h_{T-1}}^\theta(v_{T-1})$ and $\tau_{h_T}^\theta$, which results in a bilinear constraint. To restore linearity in (14), we introduce auxiliary variables $\mathbf{z}$ that encode the joint effect of belief and policies across stages. Specifically, for all $\theta \in \Theta$ and histories $h_T \in \mathcal{H}_T$, we define

$$\mathbf{z}_{h_{T-1}}^\theta := b_{T-1}(\theta) \tau_{h_{T-1}}^\theta, \quad \mathbf{z}_{h_T}^\theta := \mathcal{P}_{s_{T-1}, u_{T-1}}^{s_T} z_{h_{T-1}}^\theta(v_{T-1}) \tau_{h_T}^\theta. \quad (15)$$

The variable $z_{h_{T-1}}^\theta(v_{T-1})$ represents the joint probability of the Defender being type- θ and selecting action v_{T-1} at history h_{T-1} . For each successor history $h_T = (h_{T-1}, u_{T-1}, v_{T-1}, s_T)$, the variable $z_{h_T}^\theta(v_T)$ represents the joint probability of the Defender being type- θ and transitioning h_{T-1} to h_T and then select action v_T . Finally, the policy can be recovered as

$$\tau_{h_t}^\theta(v_t) = \frac{z_{h_t}^\theta(v_t)}{\sum_{v'} z_{h_t}^\theta(v')}. \quad (16)$$

Using the auxiliary $\mathbf{z}$ variables, (14) can be transformed to the equivalent LP under fixed Attacker policy $\sigma_{h_{T-1}}$,

$$B(\sigma_{h_{T-1}}) = \max_{\substack{\mathbf{z}_{h_{T-1}} \\ \mathbf{z}_{h_T}, \ell_{h_T}, \forall h_T}} \sigma_{h_{T-1}}^\top \left\{ \sum_{\theta} R_{h_{T-1}}^\theta \mathbf{z}_{h_{T-1}}^\theta + \left[\sum_{v_{T-1}, s_T} \ell_{h_T} \right]_{u_{T-1}} \right\} \quad (17a)$$

$$\text{s.t.} \quad \mathbf{1}^\top \mathbf{z}_{h_{T-1}}^\theta = b_{T-1}(\theta), \quad \mathbf{z}_{h_{T-1}}^\theta \geq \mathbf{0}, \quad \forall \theta \in \Theta, \quad (17b)$$

$$\forall h_T \in \mathcal{H}_T,$$

$$\mathbf{1}^\top \mathbf{z}_{h_T}^\theta = \mathcal{P}_{s_{T-1}, u_{T-1}}^{s_T} z_{h_{T-1}}^\theta(v_{T-1}), \quad \mathbf{z}_{h_T}^\theta \geq \mathbf{0}, \quad \forall \theta \in \Theta, \quad (17c)$$

$$\ell_{h_T} \mathbf{1} \leq \sum_{\theta} R_{h_T}^\theta \mathbf{z}_{h_T}^\theta. \quad (17d)$$

Note that (17) yields $B(\sigma_{h_{T-1}})$ in (11b) for a fixed $\sigma_{h_{T-1}}$. Since a feasible LP with its objective bounded in the direction of optimization admits a finite optimal solution and satisfies strong duality [2], it suffices to show that (17) under fixed $\sigma_{h_{T-1}}$ is feasible and upper bounded. Feasibility follows because (17b) and (17c) construct non-empty scaled simplices, while ℓ_{h_T} can be chosen sufficiently small to satisfy (17d). One can easily verify the upper-boundedness of its objective (17a) (see Appendix A.2). Dualizing (17) and combining with the outer minimization over $\sigma_{h_{T-1}}$ in (11) leads to the following LP for $V^*(h_{T-1}, \mathbf{b}_{T-1})$

$$V^*(h_{T-1}, \mathbf{b}_{T-1}) = \min_{\substack{\sigma_{h_{T-1}}, \mathbf{q}_{h_{T-1}} \\ \boldsymbol{\eta}_{h_T}, \mathbf{q}_{h_T}, \forall h_T}} \sum_{\theta} b_{T-1}(\theta) q_{h_{T-1}}(\theta) \quad (18a)$$

$$\text{s.t.} \quad \forall \theta \in \Theta, \forall v_{T-1} \in \mathcal{V},$$

$$\mathbf{1}^\top \sigma_{h_{T-1}} = 1, \quad \sigma_{h_{T-1}} \geq \mathbf{0}, \quad (18b)$$

$$q_{h_{T-1}}(\theta) \geq \mathbf{e}_{v_{T-1}}^\top R_{h_{T-1}}^\theta \sigma_{h_{T-1}} + \sum_{u_{T-1}, s_T} \mathcal{P}_{s_{T-1}, u_{T-1}}^{s_T} q_{h_T}(\theta), \quad (18c)$$

$$\forall h_T \in \mathcal{H}_T, \forall \theta \in \Theta, \forall v_T \in \mathcal{V},$$

$$\mathbf{1}^\top \boldsymbol{\eta}_{h_T} = \sigma_{h_{T-1}}(u_{T-1}), \quad \boldsymbol{\eta}_{h_T} \geq \mathbf{0}, \quad (18d)$$

$$q_{h_T}(\theta) \geq \mathbf{e}_{v_T}^\top R_{h_T}^{\theta^\top} \boldsymbol{\eta}_{h_T}, \quad (18e)$$

where $\mathbf{q}_{h_{T-1}}, \mathbf{q}_{h_T} \in \mathbb{R}^{|\Theta|}$, and $\boldsymbol{\eta}_{h_T} \in \mathbb{R}^{|\mathcal{U}(s_T)|}$ are the dual variables, and $\mathbf{e}_{v_{T-1}}, \mathbf{e}_{v_T} \in \mathbb{R}^{|\mathcal{V}|}$ are unit vectors with 1 on the entries corresponding to v_{T-1} and v_T . As shown later in the Attacker's LP, the vectors $\mathbf{q}$ correspond to the Attacker's performance under each Defender allocation action, while $\boldsymbol{\eta}_{h_T}$ characterizes the Attacker's policy at history h_T . By strong duality, the dual of (18) is feasible and admits the same finite value $B(\boldsymbol{\sigma}_{h_{T-1}})$ under fixed $\boldsymbol{\sigma}_{h_{T-1}}$. Since the additional Attacker policy simplex (18b) is non-empty and the finite reward function lower-bounds the objective (18a), strong duality extends to the combined minimization LP (18).

Dualizing (18) yields the two-stage Defender's maximization LP

$$V^*(h_{T-1}, \mathbf{b}_{T-1}) = \max_{\substack{\mathbf{z}_{h_{T-1}}, \ell_{h_{T-1}} \\ \mathbf{z}_{h_T}, \ell_{h_T}, \forall h_T}} \ell_{h_{T-1}} \quad (19a)$$

$$\text{s.t.} \quad \mathbf{1}^\top \mathbf{z}_{h_{T-1}}^\theta = b_{T-1}(\theta), \quad \mathbf{z}_{h_{T-1}}^\theta \geq \mathbf{0}, \quad \forall \theta \in \Theta, \quad (19b)$$

$$\ell_{h_{T-1}} \mathbf{1} \leq \sum_{\theta} R_{h_{T-1}}^\theta \mathbf{z}_{h_{T-1}}^\theta + \sum_{h_T} (\ell_{h_T} \mathbf{e}_{u_{T-1}}), \quad (19c)$$

$$\forall h_T \in \mathcal{H}_T,$$

$$\mathbf{1}^\top \mathbf{z}_{h_T}^\theta = \mathcal{P}_{s_{T-1}, u_{T-1}}^{s_T} z_{h_{T-1}}^\theta(v_{T-1}), \quad \mathbf{z}_{h_T}^\theta \geq \mathbf{0}, \quad \forall \theta \in \Theta, \quad (19d)$$

$$\ell_{h_T} \mathbf{1} \leq \sum_{\theta} R_{h_T}^\theta \mathbf{z}_{h_T}^\theta. \quad (19e)$$

The scalar variable ℓ_{h_T} in (19e) captures the performance at stage T . Note that ℓ_{h_T} is scaled by the probability of realizing the history $h_T = (h_{T-1}, u_{T-1}, v_{T-1}, s_T)$, since this probability is encoded in $\mathbf{z}_{h_T}^\theta$ as in (19d). The stage- $(T-1)$ value at $\ell_{h_{T-1}}$ is linked to stage- T values through the dynamic-program-like constraint (19c). On the right-hand side of (19c), the first summation captures the current performance, while the second represents the expected future performance. Since the transition probabilities are already incorporated in ℓ_{h_T} , the expression sums directly over all possible next histories. Moreover, constraint (19c) is an element-wise vector constraint, where each row corresponds to a possible Attacker action u_{T-1} . Combined with the maximization objective (19a), (19c) ensures optimal Defender strategy against all Attacker actions for this two-stage problem.

4.3 General T-stage LP

With the $\mathbf{z}$ -variables being defined analogously to (15) as

$$\mathbf{z}_{h_0}^\theta = b_0(\theta) \boldsymbol{\tau}_{h_0}^\theta, \quad \mathbf{z}_{h_t}^\theta = \mathcal{P}_{s_{t-1}, u_{t-1}}^{s_t} z_{h_{t-1}}^\theta(v_{t-1}) \boldsymbol{\tau}_{h_t}^\theta,$$

we can generalize the two-stage LP in (19) to the following generic T -stage Defender LP for $V^*(h_t, \mathbf{b}_0)$ using backward induction

$$V^*(h_0, \mathbf{b}_0) = \max_{\substack{\mathbf{z}_{h_t}, \ell_{h_t} \\ \forall t \in [0:T], h_t \in \mathcal{H}_t}} \ell_{h_0} \quad (20a)$$

$$\text{s.t. } \forall \theta \in \Theta,$$

$$\mathbf{1}^\top \mathbf{z}_{h_0}^\theta = b_0(\theta), \quad \mathbf{z}_{h_0}^\theta \geq \mathbf{0}, \quad (20b)$$

$$\ell_{h_0} \mathbf{1} \leq \sum_{\theta} R_{h_0}^\theta \mathbf{z}_{h_0}^\theta + \sum_{h_1} (\ell_{h_1} \mathbf{e}_{u_0}), \quad (20c)$$

$$\forall t \in [1:T-1], \forall h_t \in \mathcal{H}_t, \forall \theta \in \Theta,$$

$$\mathbf{1}^\top \mathbf{z}_{h_t}^\theta = \mathcal{P}_{s_{t-1}, u_{t-1}}^{s_t} z_{h_{t-1}}^\theta(v_{t-1}), \quad \mathbf{z}_{h_t}^\theta \geq \mathbf{0}, \quad (20d)$$

$$\ell_{h_t} \mathbf{1} \leq \sum_{\theta} R_{h_t}^\theta \mathbf{z}_{h_t}^\theta + \sum_{h_{t+1}} (\ell_{h_{t+1}} \mathbf{e}_{u_t}), \quad (20e)$$

$$\forall h_T \in \mathcal{H}_T, \forall \theta \in \Theta,$$

$$\mathbf{1}^\top \mathbf{z}_{h_T}^\theta = \mathcal{P}_{s_{T-1}, u_{T-1}}^{s_T} z_{h_{T-1}}^\theta(v_{T-1}), \quad \mathbf{z}_{h_T}^\theta \geq \mathbf{0}, \quad (20f)$$

$$\ell_{h_T} \mathbf{1} \leq \sum_{\theta} R_{h_T}^\theta \mathbf{z}_{h_T}^\theta, \quad (20g)$$

The Defender policy is reconstructed from the optimal solution $\mathbf{z}^*$ of (20) via the following normalization

$$\tau_{h_t}^{\theta*} = \begin{cases} \frac{\mathbf{z}_{h_t}^{\theta*}}{\mathbf{1}^\top \mathbf{z}_{h_t}^{\theta*}}, & \text{if } \mathbf{1}^\top \mathbf{z}_{h_t}^{\theta*} \neq 0, \\ \frac{1}{|\mathcal{V}|}, & \text{otherwise.} \end{cases}$$

4.4 Interpretation of the Defender's LP

The LP formulation (20) consists of two types of constraints. Constraints (20c), (20e), and (20g) encode optimality. Constraints (20b), (20d), and (20f) propagate the joint distribution over types and history realizations.

The LP formulation (20) offers insight into the asymmetric information structure in DRAG problems. The $\mathbf{z}$ variables encode joint distributions over types and histories, showing that the informed Defender plans over distributions of game trajectories. By shaping these trajectories, the Defender influences the uninformed Attacker's belief updates and decisions.

4.5 Computation Complexity

We next characterize the size of the LP in (20). For each history h_t at time $t \in [0:T]$, the LP contains one scalar variable ℓ_{h_t} and $|\Theta|$ vectors $\mathbf{z}_{h_t}^\theta \in \mathbb{R}^{|\mathcal{V}|}$. Since $|\mathcal{V}| = |\Theta|$, the total number of scalar variables is

$$N_{\text{var}} = \sum_{t=0}^T \sum_{h_t \in \mathcal{H}_t} (1 + |\Theta|^2) = \sum_{t=0}^T [|\mathcal{H}_t| \cdot (1 + |\Theta|^2)] \quad (21)$$

variables. For each h_t , the probability-flow constraints contribute $|\Theta|$ equalities, while the optimality constraint contributes $|\mathcal{U}(s_t)|$ inequalities. Thus, excluding the non-negativity bounds on $\mathbf{z}$, the number of constraints is

$$N_{\text{constr}} = \sum_{t=0}^T \sum_{h_t \in \mathcal{H}_t} (|\Theta| + |\mathcal{U}(s_t)|). \quad (22)$$

The $\mathbf{z}$ -variables additionally impose $\sum_{t=0}^T (|\mathcal{H}_t| \cdot |\Theta|^2)$ non-negativity bounds.

Let $|\mathcal{U}|_{\max} = \max_{s \in \mathcal{S}} |\mathcal{U}(s)|$, then, each stage has at most $|\mathcal{U}|_{\max} |\Theta| |\mathcal{S}|$ history branches. Hence $|\mathcal{H}_t|$ is upper bounded by $|\mathcal{U}|_{\max} |\Theta| |\mathcal{S}|$. The numbers of variables and constraints grow exponentially with the horizon T in the worst case, but polynomially with $|\mathcal{U}|_{\max}$, $|\Theta|$, and $|\mathcal{S}|$ for fixed T .

5 Attacker's Linear Program

By strong duality, the Attacker's equilibrium strategy σ^* can be solved from the dual of the Defender LP in (20), following a max-min formulation of $V^*(h_t, \mathbf{b}_t)$. Similar to the Defender LP, we use the change of variables $\boldsymbol{\eta}_{h_0} = \boldsymbol{\sigma}_{h_0}$ and $\boldsymbol{\eta}_{h_t} = \eta_{h_{t-1}}(u_{t-1}) \boldsymbol{\sigma}_{h_t}$, yielding the following T -stage Attacker LP

$$V^*(h_0, \mathbf{b}_0) = \min_{\substack{\boldsymbol{\eta}_{h_t}, \mathbf{q}_{h_t} \\ \forall t \in [0:T], h_t \in \mathcal{H}_t}} \sum_{\theta} b_0(\theta) q_{h_0}(\theta) \quad (23a)$$

$$\text{s.t. } \forall \theta \in \Theta, \forall v_0 \in \mathcal{V},$$

$$\mathbf{1}^\top \boldsymbol{\eta}_{h_0} = 1, \quad \boldsymbol{\eta}_{h_0} \geq \mathbf{0}, \quad (23b)$$

$$q_{h_0}(\theta) \geq \mathbf{e}_{v_0}^\top R_{h_0}^{\theta^\top} \boldsymbol{\eta}_{h_0} + \sum_{s_1, u_0} (\mathcal{P}_{s_0, u_0}^{s_1} q_{h_1}(\theta)), \quad (23c)$$

$$\forall t \in [1:T-1], \forall h_t \in \mathcal{H}_t, \forall \theta \in \Theta, \forall v_t \in \mathcal{V},$$

$$\mathbf{1}^\top \boldsymbol{\eta}_{h_t} = \eta_{h_{t-1}}(u_{t-1}), \quad \boldsymbol{\eta}_{h_t} \geq \mathbf{0}, \quad (23d)$$

$$q_{h_t}(\theta) \geq \mathbf{e}_{v_t}^\top R_{h_t}^{\theta^\top} \boldsymbol{\eta}_{h_t} + \sum_{s_{t+1}, u_t} (\mathcal{P}_{s_t, u_t}^{s_{t+1}} q_{h_{t+1}}(\theta)), \quad (23e)$$

$$\forall h_T \in \mathcal{H}_T, \forall \theta \in \Theta, \forall v_T \in \mathcal{V},$$

$$\mathbf{1}^\top \boldsymbol{\eta}_{h_T} = \eta_{h_{T-1}}(u_{T-1}), \quad \boldsymbol{\eta}_{h_T} \geq \mathbf{0}, \quad (23f)$$

$$q_{h_T}(\theta) \geq \mathbf{e}_{v_T}^\top R_{h_T}^{\theta^\top} \boldsymbol{\eta}_{h_T}. \quad (23g)$$

Both (20) and (23) compute $V^*(h_0, \mathbf{b}_0)$, but in two different ways. The Defender's and Attacker's equilibrium strategies are recovered by normalizing the optimal solutions of (20) and (23), respectively. Since (23) is the dual of (20), it has N_{constr} variables and N_{var} constraints, excluding variable-sign restrictions. Unlike the $\mathbf{z}$ variables, $\boldsymbol{\eta}$ encodes only the Attacker's action distribution along each game trajectory, reflecting the Attacker's informational disadvantage. Since

belief updates (3) are determined solely by the Defender's strategy, the Attacker neither controls nor plans how its belief evolves, but instead optimizes performance aggregated across all game types. Due to strong duality, the LPs in (20) and (23) constructed from the dynamic programs ensure the existence of PBNE in DRAG, as specified in Theorem 1.

Theorem 1 (Existence of PBNE). *The DRAG can be solved as a pair of primal-dual linear programs, which admits a unique and finite game value $V^*(h_0, \mathbf{b}_0)$ due to strong duality. Through normalization, the LPs' optimal solutions constitute a PBNE.*

6 Numerical Experiments

We evaluated the proposed LP approach on the 4×4 grid DRAG in Fig. 2, where the two candidate assets are marked in green and blue, and the prior is set to $\mathbf{b}_0 = [0.2, 0.8]$. The running reward for a type- θ Defender is defined as

$$r(h_t, \theta, u_t, v_t) = \mathbb{1}(v_t = \theta) \mathbb{1}(\Theta \cap s_{0:t} = \emptyset),$$

i.e., the Defender receives reward only when allocating to the true asset before the Attacker reaches any candidate asset.

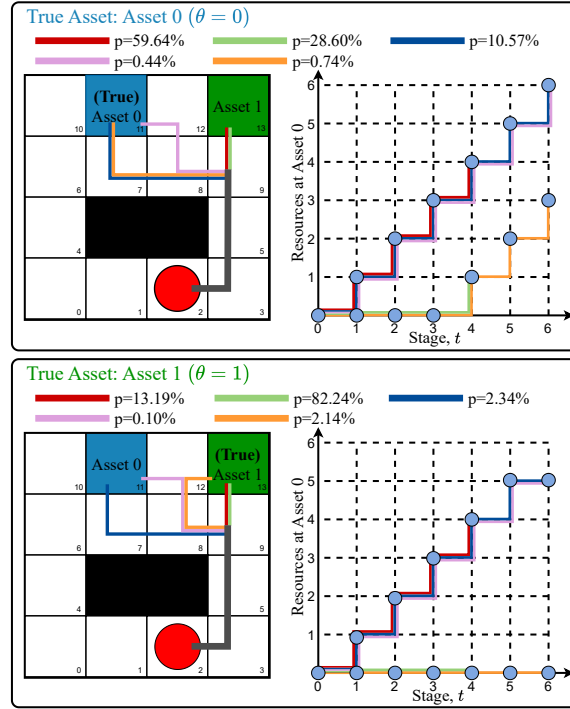


Fig. 2. Histories of a 4×4 DRAG. The label p denotes the probability of the game history h_t depicted in the corresponding color. *Top & Bottom:* Attacker trajectories and Defender allocation when Asset 0 & 1 is selected.

The game terminates when the Attacker either (i) fails to reach any asset by the horizon $T = 10$, or (ii) reaches one asset at the terminal time $\hat{t} \leq T$. At time $\hat{t}$, we assign the additional terminal reward

$$r_{\hat{t}} = -m \mathbb{1}(s_{\hat{t}} = \theta).$$

That is, if the Attacker reaches the true asset, the Defender is penalized with the Attacker threat level $m = 25$. Figure 2 presents the possible game histories h_t under the strategy profile (σ^*, τ^*) obtained from the LPs in (20) and (23). Since h_t contains both the Attacker trajectory and the Defender allocation history, we visualize them separately for clarity. With only two assets, the Defender's allocation is characterized by resources at Asset 0. For an arbitrary h_t , the corresponding Attacker trajectory and the Defender allocation are plotted using the same color.

6.1 Deceptive Behaviors

We first illustrate how (σ^*, τ^*) induces belief manipulation and leads to emergent deceptive behavior. In Fig. 3, we consider all trajectories following the parent history $\hat{h}_3$, under which the Attacker follows the state sequence (2, 3, 5, 9) and the Defender allocates to Asset 0 throughout. Along $\hat{h}_3$, the evolution of the Defender's policy and the Attacker's belief is characterized by $\{(\tau_{h_t}^\theta(0), b_t(0))\}_{t=0}^3$ in Fig. 3. At time $t = 0$, the Defender randomizes its allocation with policy $\tau_{h_0}^0 = [0.71, 0.29]$, $\tau_{h_0}^1 = [0.16, 0.84]$. This randomized strategy induces the posterior Attacker belief $\mathbf{b}_1 = [0.53, 0.47]$, which is later preserved by the type-invariant policy $\tau_{h_t}^\theta = [1, 0]$ until reaching state 9, strategically positioning the Attacker at an indifference point.

At state 9, the Attacker faces two possible actions: moving **Left** toward Asset 0 or **Up** toward Asset 1, each yielding different payoffs depending on the type θ . Specifically, moving to Asset 0 yields payoffs $(-19, 1)$ under types $(0, 1)$, while moving to Asset 1 yields $(4, -25)$. Under the induced belief $[0.53, 0.47]$, the expected payoff of moving to either asset is identical, i.e.

$$0.53 \times (-19) + 0.47 \times 1 = 0.53 \times 4 + 0.47 \times (-25).$$

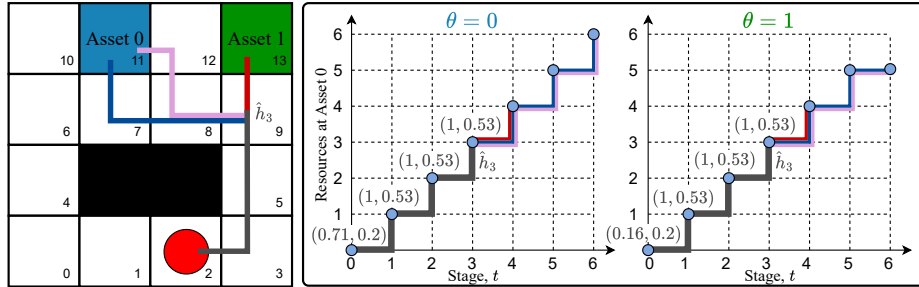


Fig. 3. Illustration of deception. *Left:* Attacker trajectories; *Middle & Right:* Type-0 & 1 Defender allocation histories.

Therefore, the Attacker is indifferent among all actions at state 9, which is the consequence of the Defender’s deliberate randomization, which strategically shapes the Attacker’s ambiguous belief $[0.53, 0.47]$. However, this randomization may appear “suboptimal”, since the type-0 Defender allocates to the decoy Asset 1 with a probability of 0.29, effectively “wasting” resources. As we will show later, inducing such ambiguity at the critical Attacker decision point in fact maximizes Defender performance, thus constituting *purposeful deception*. We use the term *purposeful deception* to describe deceptive behavior that emerges endogenously from the PBNE when influencing the Attacker’s belief and subsequent actions improves the Defender’s equilibrium performance. In contrast, optimizing a separate deception heuristic may encourage unnecessary deceptive behavior that degrades overall performance, resulting in “deception for the sake of deception”.

6.2 Effectiveness of Deception

The LP-generated strategy profile (σ^*, τ^*) in Fig. 2 leads to an expected performance $V_{LP}^*(s_0, \mathbf{b}_0) = -16.68$. To quantify the Defender’s informational advantage, we compared with the full-information game outcome $V_{FI}^*(s_0, \mathbf{b}_0)$, where Nature selects the true asset via $\mathbf{b}_0$ and publicly announces it. In the type- θ full-information game, the optimal Attacker strategy is the shortest path to the true asset θ , while the Defender best responds by allocating to θ at every step. Such a strategy pair leads to the full-information value

$$V_{FI}^*(s_0, \mathbf{b}_0) = 0.2 \times (-19) + 0.8 \times (-21) = -20.6.$$

With the *value of information* (VoI) [15] defined in (24), we can capture the Defender’s performance gain due to informational advantage:

$$\text{VoI}(s_0, \mathbf{b}_0) = \frac{V_{LP}^*(s_0, \mathbf{b}_0) - V_{FI}^*(s_0, \mathbf{b}_0)}{|V_{FI}^*(s_0, \mathbf{b}_0)|}. \quad (24)$$

For the 4×4 DRAG, the performance gap between the two values is 3.92, corresponding to a VoI of 19%. This indicates that, by purposefully allocating to the decoys, the Defender manipulates the Attacker’s belief, leading to a 19% improvement with respect to the full-information setting.

6.3 Equilibrium Validation

We next verify that (σ^*, τ^*) constitutes a PBNE by showing that neither player benefits from unilateral deviation.

Table 1. Performances under unilateral Attacker deviation.

$V^*(s_0, \mathbf{b}_0, \sigma, \tau^*)$	LP-A, σ^*	RS-A	HPSP-A	LPSP-A
LP-D, τ^*	-16.68	-8.32	-16.68	-0.85

Attacker Deviation We consider the following alternative Attacker strategies: RS-A samples actions uniformly; HPSP-A and LPSP-A follow a shortest path to the highest-prior and the lowest-prior assets, respectively. Table 1 reports the performances of the alternative Attacker strategies against the Defender strategy τ^* . As expected, all deviations yield values no smaller than $V_{LP}^*(s_0, \mathbf{b}_0)$, confirming that the (minimizing) Attacker cannot improve by deviating from σ^* . Notably, HPSP-A attains the same performance as σ^* . However, the Defender can best respond against HPSP-A by consistently allocating to the true asset, thus exploiting HPSP-A. This yields a value of -16 , strictly improving the Defender’s payoff relative to the equilibrium.

Table 2. Performances under unilateral Defender deviation.

$V^*(s_0, \mathbf{b}_0, \sigma^*, \tau)$	LP-D, τ^*	RS-D	TC-D	T0-D	C0-D	C1-D
LP-A, σ^*	-16.68	-16.85	-16.68	-16.88	-16.80	-16.76

Defender Deviation We experimented with the Defender alternatives: RS-D uniformly samples actions; TC-D and T0-D respectively allocate to true and decoy assets. ; C0-D and C1-D respectively allocate to Asset 0 and Asset 1 regardless of the selected true asset. Table 2 shows that all deviations yield performances no greater than $V_{LP}^*(s_0, \mathbf{b}_0)$, validating the optimality of τ^* . Although TC-D achieves the same performance as τ^* , the Attacker can exploit it by moving toward the true asset identified immediately after the first observation, yielding a value of -20.6 that is strictly lower than $V_{LP}^*(s_0, \mathbf{b}_0)$.

7 Conclusion

We introduced the Deceptive Resource Allocation Game (DRAG) and studied purposeful deception under a Bayesian game framework. While the informed player does not fully control the observed signal, we showed that DRAG still admits a linear program for computing Perfect Bayesian Nash Equilibria. Numerical results show that purposeful deception arises by shaping beliefs to influence decisions at critical states, rather than simply maximizing inference bias or uncertainty. Future work includes extension to two-sided information asymmetry and larger-scale environments.

Acknowledgements: This work was supported by the U.S. Army Research Laboratory (ARL) through the Distributed and Collaborative Intelligent Systems and Technology (DCIST) Collaborative Research Alliance (CRA) under Grant W911NF-17-2-0181 and Office of Naval Research (ONR) award N00014-23-1-2308.

Disclosure of Interests: The authors have no competing interests to declare that are relevant to the content of this article.

A Appendix

A.1 Proof of Proposition 1

Proof. For $t \in [1 : T]$, given the observed history $h_t = (h_{t-1}, u_{t-1}, v_{t-1}, s_t)$, we will show (3) outputs $b_t(\theta)$ under Assumption 1.

By Definition 1, the numerator of (3) can be interpreted as

$$\begin{aligned} b_{t-1}(\theta) \tau_{h_{t-1}}^\theta(v_{t-1}) &= \mathcal{P}^\tau(\theta \mid h_{t-1}) \mathcal{P}^\tau(v_{t-1} \mid h_{t-1}, \theta) \\ &= \frac{\mathcal{P}(h_{t-1}, \theta)}{\mathcal{P}(h_{t-1})} \frac{\mathcal{P}(v_{t-1}, h_{t-1}, \theta)}{\mathcal{P}(h_{t-1}, \theta)} \\ &= \mathcal{P}^\tau(v_{t-1}, \theta \mid h_{t-1}). \end{aligned}$$

The denominator of (3) is a marginalization of the numerator, which can be interpreted as

$$\sum_{\theta'} b_{t-1}(\theta') \tau_{h_{t-1}}^{\theta'}(v_{t-1}) = \mathcal{P}^\tau(v_{t-1} \mid h_{t-1}).$$

Therefore, (3) can be expressed as follows

$$\begin{aligned} \frac{b_{t-1}(\theta) \tau_{h_{t-1}}^\theta(v_{t-1})}{\sum_{\theta'} b_{t-1}(\theta') \tau_{h_{t-1}}^{\theta'}(v_{t-1})} &= \frac{\mathcal{P}^\tau(v_{t-1}, \theta \mid h_{t-1})}{\mathcal{P}^\tau(v_{t-1} \mid h_{t-1})} \\ &= \frac{\mathcal{P}(v_{t-1}, \theta, h_{t-1})}{\mathcal{P}(h_{t-1})} \frac{\mathcal{P}(h_{t-1})}{\mathcal{P}(v_{t-1}, h_{t-1})} \\ &= \mathcal{P}^\tau(\theta \mid h_{t-1}, v_{t-1}) \\ &= b_t(\theta), \end{aligned}$$

where the last equality is a result of Assumption 1. ■

A.2 Objective Upper-Boundedness

Lemma 1. For a fixed history h_{T-1} , define

$$\bar{r}_{T-1} := \max_{\theta, u_{T-1}, v_{T-1}} r(h_{T-1}, \theta, u_{T-1}, v_{T-1}), \quad (25a)$$

$$\bar{r}_T := \max_{h_T, \theta, u_T, v_T} r(h_T, \theta, u_T, v_T). \quad (25b)$$

Then, every feasible solution of (17) satisfies

$$\sigma_{h_{T-1}}^\top \left\{ \sum_{\theta} R_{h_{T-1}}^\theta \mathbf{z}_{h_{T-1}}^\theta + \left[\sum_{v_{T-1}, s_T} \ell_{h_T} \right]_{u_{T-1}} \right\} \leq \bar{r}_{T-1} + \bar{r}_T. \quad (26)$$

Proof. Since the state and action spaces and the horizon are finite, the game tree of DRAG is finite. Hence, the maxima $\bar{r}_{T-1}$ and $\bar{r}_T$ in (25) are finite and attainable.

Under a fixed $\sigma_{h_{T-1}}$, we will show the objective (17a) is bounded above by $\bar{r}_{T-1} + \bar{r}_T$ for all feasible solutions. From the first part of (17a), we have

$$\begin{aligned} \sigma_{h_{T-1}}^\top \sum_{\theta} R_{h_{T-1}}^\theta \mathbf{z}_{h_{T-1}}^\theta &= \sum_{\theta, u_{T-1}, v_{T-1}} \sigma_{h_{T-1}}(u_{T-1}) r(h_{T-1}, \theta, u_{T-1}, v_{T-1}) z_{h_{T-1}}^\theta(v_{T-1}) \\ &\leq \bar{r}_{T-1} \sum_{u_{T-1}} \left(\sigma_{h_{T-1}}(u_{T-1}) \sum_{\theta} \mathbf{1}^\top \mathbf{z}_{h_{T-1}}^\theta \right) \end{aligned} \quad (27)$$

$$\begin{aligned}
&= \bar{r}_{T-1} \sum_{u_{T-1}} \left(\sigma_{h_{T-1}}(u_{T-1}) \sum_{\theta} b_{T-1}(\theta) \right) \\
&= \bar{r}_{T-1},
\end{aligned} \tag{28}$$

where (27) follows from the definition of $\bar{r}_{T-1}$ in (25a) and $\mathbf{z}_{h_{T-1}}^{\theta} \geq \mathbf{0}$; the last two equalities use (17b), $\sum_{\theta} b_{T-1}(\theta) = 1$, and $\mathbf{1}^{\top} \sigma_{h_{T-1}} = 1$.

For any successor history $h_T = (h_{T-1}, u_{T-1}, v_{T-1}, s_T)$, notice constraint (17d) implies, for every $u_T \in \mathcal{U}(s_T)$,

$$\ell_{h_T} \leq \mathbf{e}_{u_T}^{\top} \sum_{\theta} R_{h_T}^{\theta} \mathbf{z}_{h_T}^{\theta} \leq \bar{r}_T \sum_{\theta} \mathbf{1}^{\top} \mathbf{z}_{h_T}^{\theta} = \bar{r}_T \sum_{\theta} \mathcal{P}_{s_{T-1}, u_{T-1}}^{s_T} z_{h_{T-1}}^{\theta}(v_{T-1}). \tag{29}$$

Using (29), the definition of $\bar{r}_T$ in (25b), and the probability-flow constraints on $\mathbf{z}_{h_T}^{\theta}$ in (17b) and (17c), the second part of (17a) can be expressed as

$$\begin{aligned}
\sigma_{h_{T-1}}^{\top} \left[\sum_{v_{T-1}, s_T} \ell_{h_T} \right]_{u_{T-1}} &= \sum_{u_{T-1}, v_{T-1}, s_T} \sigma_{h_{T-1}}(u_{T-1}) \ell_{h_T} \\
&\leq \bar{r}_T \sum_{u_{T-1}} \left(\sigma_{h_{T-1}}(u_{T-1}) \sum_{v_{T-1}, s_T, \theta} \mathcal{P}_{s_{T-1}, u_{T-1}}^{s_T} z_{h_{T-1}}^{\theta}(v_{T-1}) \right) \\
&= \bar{r}_T \sum_{u_{T-1}} \left(\sigma_{h_{T-1}}(u_{T-1}) \sum_{s_T} \left(\mathcal{P}_{s_{T-1}, u_{T-1}}^{s_T} \sum_{\theta} \mathbf{1}^{\top} \mathbf{z}_{h_{T-1}}^{\theta} \right) \right) \\
&= \bar{r}_T.
\end{aligned}$$

Combining the inequalities for both parts of (17a) yields the result in (26). ■

References

1. Almeshekah, M.H., Spafford, E.H.: Cyber Security Deception. Cyber Deception: Building the Scientific Foundation pp. 23–50 (2016)
2. Bertsimas, D., Tsitsiklis, J.: Introduction to Linear Optimization. Athena Scientific (1997)
3. Chen, S., Savas, Y., Karabag, M.O., Sadler, B.M., Topcu, U.: Deceptive Planning for Resource Allocation. In: 2024 American Control Conference (ACC). pp. 4188–4195 (2024). <https://doi.org/10.23919/ACC60939.2024.10644373>
4. Flesch, J., Predtetchinski, A., Sudderth, W.: Superfair Stochastic Games. Applied Mathematics & Optimization **88**(3), 75 (2023). <https://doi.org/10.1007/s00245-023-10051-z>
5. Guan, Y., Maity, D., Tsiotras, P.: Strategic Concealment of Environment Representations in Competitive Games. IEEE Control Systems Letters **9**, 2609–2614 (2025)
6. Karabag, M.O., Ornik, M., Topcu, U.: Deception in Supervisory Control. IEEE Transactions on Automatic Control **67**(2), 738–753 (2022)
7. Koller, D., Megiddo, N., von Stengel, B.: Efficient Computation of Equilibria for Extensive Two-Person Games. Games and Economic Behavior **14**(2), 247–259 (1996). <https://doi.org/10.1006/game.1996.0051>
8. Li, L., Shamma, J.: LP Formulation of Asymmetric Zero-Sum Stochastic Games. 53rd IEEE Conference on Decision and Control pp. 1930–1935 (2014)
9. Lloyd, M.: The Art of Military Deception. Pen and Sword (2003)

10. Maschler, M., Solan, E., Zamir, S.: Game Theory. Cambridge University Press (2013)
11. Nikitas, A., Parkinson, S., Vallati, M.: The deceitful Connected and Autonomous Vehicle: Defining the concept, contextualising its dimensions and proposing mitigation policies. *Transport Policy* **122**, 1–10 (2022)
12. Ouyang, Y., Tavafoghi, H., Teneketzis, D.: Dynamic Games With Asymmetric Information: Common Information Based Perfect Bayesian Equilibria and Sequential Decomposition. *IEEE Transactions on Automatic Control* **62**(1), 222–237 (2017)
13. Pereira, R., Oren, N., Meneguzzi, F.: Landmark-Based Heuristics for Goal Recognition. In: *AAAI Conf. Artif. Intell.* vol. 31 (2017)
14. Ramírez, M., Geffner, H.: Goal Recognition over POMDPs: Inferring the Intention of a POMDP Agent. In: *Int. Joint Conf. on Artif. Intell.* pp. 2009–2014 (2011)
15. Rostobaya, V., Berneburg, J., Guan, Y., Dorothy, M., Shishika, D.: Deceptive Path Planning: A Bayesian Game Approach. *arXiv preprint arXiv:2506.13650* (2025)
16. Rostobaya, V., Guan, Y., Berneburg, J., Dorothy, M., Shishika, D.: Deception by Motion: The Eater and the Mover Game. *IEEE Control Systems Letters* **7**, 3157–3162 (2023)
17. Rostobaya, V., Guan, Y., Berneburg, J., Shishika, D.: Linear programming approach to deceptive path planning game with goal selection. *arXiv preprint arXiv:2605.16548* (2026)
18. Wayllace, C., Hou, P., Yeoh, W., Son, T.C.: Goal Recognition Design with Stochastic Agent Action Outcomes. In: *Int. Joint Conf. on Artif. Intell.* pp. 3279–3285 (2016)
19. Yildiz, M.: Chapter 16: Dynamic Games with Incomplete Information (2012), MIT OpenCourseWare