Les Cahiers du GERAD ISSN: 0711–2440

The value of randomized strategies in distributionally robust risk averse network interdiction games

U. Sadana, E. Delage

G–2020–20

March 2020

La publication de ces rapports de recherche est rendue possible grâce au soutien de HEC Montréal, Polytechnique Montréal, Université McGill, Université du Québec à Montréal, ainsi que du Fonds de recherche du Québec – Nature et technologies.

Dépôt légal – Bibliothèque et Archives nationales du Québec, 2020
– Bibliothèque et Archives Canada, 2020

The publication of these research reports is made possible thanks to the support of HEC Montréal, Polytechnique Montréal, McGill University, Université du Québec à Montréal, as well as the Fonds de recherche du Québec – Nature et technologies.

Legal deposit – Bibliothèque et Archives nationales du Québec, 2020
– Library and Archives Canada, 2020

GERAD HEC Montréal
3000, chemin de la Côte-Sainte-Catherine
Montréal (Québec) Canada H3T 2A7

Tél. : 514 340-6053
Téléc. : 514 340-5665
info@gerad.ca
www.gerad.ca

The value of randomized strategies in distributionally robust risk averse network interdiction games

Utsav Sadana
Erick Delage

GERAD & Department of Decision Sciences, HEC Montréal, Montréal (Québec), Canada, H3T 2A7

utsav.sadana@gerad.ca
erick.delage@hec.ca

March 2020
Les Cahiers du GERAD
G–2020–20

Abstract: Conditional Value at Risk (CVaR) is widely used to account for the preferences of a risk-averse agent in the extreme loss scenarios. To study the effectiveness of randomization in interdiction games with an interdictor that is both risk and ambiguity averse, we introduce a *distributionally robust network interdiction game* where the interdictor randomizes over the feasible interdiction plans in order to minimize the worst-case CVaR of the flow with respect to both the unknown distribution of the capacity of the arcs and his mixed strategy over interdicted arcs. The flow player, on the contrary, maximizes the total flow in the network. By using the budgeted uncertainty set, we control the degree of conservatism in the model and reformulate the interdictor's non-linear problem as a bi-convex optimization problem. For solving this problem to any given optimality level, we devise a spatial branch and bound algorithm that uses the McCormick inequalities and reduced reformulation linearization technique (RRLT) to obtain convex relaxation of the problem. We also develop a column generation algorithm to identify the optimal support of the convex relaxation which is then used in the coordinate descent algorithm to determine the upper bounds. The efficiency and convergence of the spatial branch and bound algorithm is established in the numerical experiments. Further, our numerical experiments show that randomized strategies can have significantly better in-sample and out-of-sample performance than optimal deterministic ones.

Keywords: Conditional Value at Risk, distributionally robust optimization, interdiction game, column generation

1 Introduction

Game-theoretic models are increasingly being used to determine the best practice in security policy, e.g., ARMOR program in Los Angeles International Airport, see Pita et al. (2008), Pita et al. (2009), Kar et al. (2017). Yet, in many real-world applications, the parameters of an agent's decision model, e.g., the capacity of the arcs in a network interdiction game, can be undetermined. It was shown in Ben-Tal and Nemirovski (2000) that perturbations in the parameters of a linear programming problem can render the solution to become infeasible or significantly suboptimal. A popular technique to tackle this issue is Stochastic Programming (SP) which assumes that the distribution of the parameters is known. However, in many instances, the distribution of the random parameters is not exactly known. This can lead to post-decision disappointment referred to as the *optimizer's curse* (see Smith and Winkler (2006)). Alternatively, classical robust optimization models do not use any distributional information on the random parameters. However, imposing that only the worst-case outcome is to be taken into account can often lead to over-conservative solutions.

Distributionally Robust Optimization (DRO) acts as a compromise between SP and RO by requiring the probability distribution to lie in a distributional ambiguity set and seeking a solution that performs best according to the worst-case distribution. Hence, when the ambiguous distribution set is properly tuned, it can prevent both the post-decision disappointment of SP models and the over-conservatism of RO models. Furthermore, the set can be calibrated in ways that will provide statistical guarantees on the out-of-sample performance of the DRO solution. Recently, it was shown in Delage et al. (2019) that it can be beneficial to employ randomization in non-convex DRO problems. For an ambiguity-averse risk-neutral decision maker, i.e. one that minimizes worst-case expected value, Delage and Saif (2018) proposed an algorithm to identify such strategies in mixed-integer two-stage DRO problems. Bertsimas et al. (2016) also studied the value of randomization specifically in the context of a network interdiction game with known parameters and risk-neutral agents. Nevertheless, none of these works provide either theoretical or computational means of identifying optimal randomized solutions for agents that employ more general risk measures than expected value, such as the Conditional Value at Risk (CVaR) measure introduced in Rockafellar and Uryasev (2000). Such risk measures are especially relevant in security policy model's such as network interdiction problems where a decision maker, e.g., a law-enforcement agency, might be concerned by the possibility of incurring huge losses under certain scenarios, e.g. caused by a large flow of illegal drugs, weapons or money.

In this paper, we study a *distributionally robust network interdiction game* where the interdictor employs a CVaR risk measure to model his risk aversion. Our contributions can be summarized as follows:

- On the methodological side, we introduce for the first time ambiguity and risk aversion in network interdiction games where the interdictor minimizes the worst-case CVaR over both the unknown distribution of the capacities of the arcs and the distribution of interdicted arcs. This is in sharp contrast with the work of Loizou (2015) who considers an interdictor that employs CVaR to handle parameter uncertainty but an expected value to handle the uncertainty caused by his own randomized strategy. We show that the approach in Loizou (2015) can in fact produce a solution that is stochastically strictly dominated by the solution of our proposed model.
- On the algorithm side, we complement the work of Delage and Saif (2018) by designing the first algorithm that can identify optimal randomized strategies for an ambiguity averse risk averse agent, i.e. an agent that minimizes a worst-case convex risk measure other than the expected value. Our algorithm is based on a spatial branch and bound scheme (see Al-Khayyal and Falk (1983)) embedded with the column generation (CG) method. It will successfully identify high quality randomized solution for networks containing hundreds of nodes in a few minutes. We expect that these results should pave the way to developing algorithms for solving non-zero sum Stackelberg games where players use a CVaR risk measure, and are ambiguity-averse.
- On the empirical side, we provide evidence which indicates that a network interdictor can significantly benefit from randomization. We find that, while it is true that deterministic plans are

often worst-case CVaR optimal, in instances where randomization strictly improves this objective, the improvement is significant, i.e. a 4% improvement on average. Furthermore, for those case, the out-of-sample performance of the randomized strategy also always strictly improves the out-of-sample performance and by as much as 19% on average.

The rest of the paper is organized as follows. Section 2 gives an overview of the literature covering network interdiction games and DRO problems which are closely related to this work. We also briefly discuss the algorithms which have been proposed previously to solve non-convex optimization problems with bilinear constraints. Section 3 defines our *distributionally robust network interdiction game*. The robust counterpart of the DRO model is given in Section 4 where we also describe the spatial branch and bound algorithm embedded with the CG algorithm that is used to solve the interdictor's problem to global optimality. Numerical experiments are provided in Section 5 to demonstrate the convergence and efficiency of our algorithm and illustrate that randomization can significantly improve both in-sample and out-of-sample performance. Concluding remarks are given in Section 6.

Notations

Vectors are expressed in bold and matrices are represented by capital letters. $\mathbf{1}$ and $\mathbf{0}$ denote column vectors of 1's and 0's respectively. The identity matrix is denoted by I, and $\boldsymbol{e}_i$ captures its i-th column. The set of all probability measures on a finite discrete measurable space $(\mathcal{X}, F_{\mathcal{X}})$ is denoted by $\Delta\mathcal{X} \subseteq \mathbb{R}_+^{|\mathcal{X}|}$, where $F_{\mathcal{X}}$ denotes all subsets of $\mathcal{X}$.

2 Related literature

It is well-known that the illegal flow of drugs, weapons or other hazardous substances poses a threat to the security of a nation, see Magliocca et al. (2019) and the references in it. The law-enforcement agencies aim to reduce their flow while the adversaries, which may be smugglers or terrorist organizations, try to increase it. Network flow problems arise in diverse areas like transportation (Israeli and Wood (2002)), military manpower planning (Gass (1991)), medicine (Assimakopoulos (1987)). For an elaborate description of theory, algorithms and applications of network flow problems, refer to Ahuja et al. (1993). One of their useful applications is in determining the optimal interdiction policies for a network operator. Numerous network flow models have been developed to study the network interdiction game where the defender interdicts a set of arcs on the network to minimize the flow while the adversary maximizes it, see Wood (1993), Cormican et al. (1998), Smith et al. (2013), Smith and Song (2020)).

In Jain et al. (2010), it is shown that randomization can be useful in security applications like patrolling of airports. Recently, Bertsimas et al. (2016) introduced a randomized network interdiction game where only the interdictor can randomize. The authors assumed interdictor as well as the flow player to be risk-neutral, and the model parameters are known with certainty. In contrast to their model, Lei et al. (2018) assumed deterministic strategies for the players, and the effect of interdiction on the capacity of each arc is random. Also, the players use CVaR risk measure. In Atamturk et al. (2017), the capacities are assumed to be stochastic, and the interdictor minimizes the maximum flow-at-risk over a discrete set of actions.

Clearly, the literature on interdiction has been limited to SP models. Also, the benefit of using randomized strategies for a risk-averse agent in DRO problems has not been explored. Aghassi and Bertsimas (2006) have developed robust game theory models where players are expected utility maximizers, and have incomplete information on the true probability distribution of their payoffs. In order to account for ambiguity aversion in non-cooperative games, Loizou (2015) has proposed a distributionally robust game theory model where players use worst-case expected CVaR measure to evaluate the performance of their strategies. We show, using an example, that a single player counterpart of his model results in an optimal policy that is strictly stochastically dominated by the strategy produced by our model.

Alternatively, we propose the *distributionally robust network interdiction game* where the interdictor minimizes the worst-case CVaR of the flow with respect to both the unknown distribution of the capacity of the arcs and his mixed strategy over the feasible interdiction plans. In our model, both the interdictor and flow player can randomize over their respective action sets. Similar to Janjarassuk and Linderoth (2008), we assume that the flow player can observe the capacity of each arc as well as the interdicted arcs before selecting her optimal strategy. However, the success of interdiction is a Bernoulli random variable in Janjarassuk and Linderoth (2008) while it is deterministic in our model.

Hajinezhad and Shi (2018) and Hajinezhad and Hong (2019)) made separability assumptions on the objective functions to solve non-convex non-smooth optimization problems with bilinear constraints for machine learning and signal processing applications. We do not make those assumptions on the interdictor's bi-convex optimization problem and we devise the spatial branch and bound algorithm (see Al-Khayyal and Falk (1983)) that iteratively searches the feasible space of the problem to obtain global optimal solutions within a given tolerance. The rate of convergence of the algorithm relies on generating tighter lower and upper bounds. In the literature, McCormick inequalities (see McCormick (1976)) are commonly used to obtain convex relaxations of the non-convex non-linear programming problems with bilinear constraints. In order to obtain a tighter bound, Reformulation Linearization Technique (RLT) was proposed in Sherali and Alameddine (1992) wherein valid inequalities obtained by multiplying the pairs of feasible constraints are added to the relaxed problem. This increases the size of LP relaxation. Liberti and Pantelides (2006) proposed the reduced RLT, which in contrast to RLT, gives an exact reformulation of original non-convex problem with an additional number of linear equality constraints. The authors also show that RRLT combined with McCormick inequalities can result in tighter relaxations than applying McCormick directly to the bilinear terms.

3 Distributionally robust network interdiction game

Consider a directed graph $G = (V, E)$, where V and E denote the nodes and arcs, respectively. Let $e = (i, j)$ represent an arc on G, $\delta^-(i) = \{(i, j)|j \in V\}$ and $\delta^+(i) = \{(j, i)|j \in V\}$ denote, respectively, the set of arcs leaving and entering node $i \in V$. A flow in the graph G is denoted by a non-negative vector $\boldsymbol{x} \in \mathbb{R}_+^{|E|}$ so that for each $e \in E$, we have $x_e \leq c_e$, where the capacity of all arcs is denoted by $\boldsymbol{c} \in \mathbb{R}_+^{|E|}$. The conservation of flow at each node is ensured by

$$\sum_{e \in \delta^-(i)} x_e - \sum_{e \in \delta^+(i)} x_e = 0 \qquad \forall i \notin \{s, t\}, \tag{1}$$

where s and t denote the source and sink node, respectively. The flow player aims at maximizing the flow in the network. The interdictor, on the other hand, aims at minimizing the worst-case CVaR of the flow with respect to both the unknown distribution of the capacity of the arcs and his mixed strategy over the interdicted arcs. We assume that the interdictor has a budget to remove B arcs in the network. Let $\mathcal{L}$ denote the finite set of feasible plans for the interdictor where $\mathcal{L} = \{\boldsymbol{\ell} \in \{0, 1\}^{|E|} | \sum_e \ell_e \leq B\}$. Here, $\ell_e = 1$ if the interdictor removes arc e and $\ell_e = 0$ if arc e is not interdicted. The distribution of the capacities of all the arcs is only known to lie in set $\mathcal{Q}$. We assume that the distribution is discrete with a set of scenarios $\mathcal{K}$ supported on $\{\boldsymbol{c}^k\}_{k \in \mathcal{K}}$.

Similar to the model in Bertsimas et al. (2016), we assume that the randomized strategy of the interdictor is a probability distribution $\boldsymbol{u}$ over the set $\mathcal{L}$ where $\boldsymbol{u} \in \Delta\mathcal{L}$. For any interdiction plan $\boldsymbol{\ell}$ and scenario k, the flow player solves the following problem

$$f_{\boldsymbol{\ell},k} := \underset{\boldsymbol{x}}{\text{maximize}} \quad \boldsymbol{d}^T \boldsymbol{x} \tag{2a}$$

$$\text{subject to} \quad N\boldsymbol{x} = 0 \tag{2b}$$

$$0 \leq \boldsymbol{x} \leq C^k(\mathbf{1} - \boldsymbol{\ell}), \tag{2c}$$

where $\boldsymbol{d}^T \boldsymbol{x} = \sum_{e \in \delta^+(t)} x_e$, $C^k = \text{diag}(\boldsymbol{c}^k)$, and (2b) is a shorthand notation for constraint (1).

The interdictor solves the following distributionally robust network interdiction (DRNI) problem:

$$\text{(DRNI)} \quad \underset{\boldsymbol{u}\in\Delta\mathcal{L}}{\text{minimize}} \max_{\boldsymbol{q}\in\mathcal{Q}} \text{CVaR}^{\alpha}_{\boldsymbol{\ell}\sim\boldsymbol{u},k\sim\boldsymbol{q}}[f_{\boldsymbol{\ell},k}], \tag{3}$$

where CVaR is defined over the joint distribution of capacities and interdicted arcs. Namely, with risk aversion parameter $\alpha = [0, 1)$, CVaR is defined as

$$\text{CVaR}^{\alpha}_{\boldsymbol{\ell}\sim\boldsymbol{u},k\sim\boldsymbol{q}}[f_{\boldsymbol{\ell},k}] := \inf_{\zeta} \ \zeta + \frac{1}{1-\alpha}\sum_{\boldsymbol{\ell}}\sum_{k} q_k u_{\boldsymbol{\ell}}[f_{\boldsymbol{\ell},k} - \zeta]^+,$$

where $[f_{\boldsymbol{\ell},k} - \zeta]^+ := \max(f_{\boldsymbol{\ell},k} - \zeta, 0)$.

Remark 1 *Since the set $\mathcal{X} := \{\boldsymbol{x} \,|\, N\boldsymbol{x} = 0,\ 0 \le \boldsymbol{x} \le C^k(\mathbf{1} - \boldsymbol{\ell})\}$ of all possible s-t flows is convex, randomization is not beneficial for the flow player if he considers minimizing a convex risk measure of $\boldsymbol{d}^T\boldsymbol{x}$, such as, $\min_{\boldsymbol{u}_x\in\Delta\mathcal{X}} CVaR^{\alpha}_{\boldsymbol{x}\sim\boldsymbol{u}_x}[-\boldsymbol{d}^T\boldsymbol{x}]$ (see Delage et al. (2019)).*

Under this setting, one can show that the interdictor's DRNI problem can be cast as a bi-convex DRO problem.

Proposition 1 *When $\mathcal{Q}$ is a convex set, the interdictor's DRNI problem (3) is equivalent to the following bi-convex DRO problem*

$$\begin{aligned}
\underset{\boldsymbol{u},\zeta,\boldsymbol{\Delta},t,\boldsymbol{\eta}}{\text{minimize}} \quad & t && && \text{(4a)}\\
\textit{subject to} \quad & \zeta + \frac{1}{1-\alpha}\sum_{\boldsymbol{\ell}}\sum_{k} q_k \Delta_{\boldsymbol{\ell},k} \le t && \forall \boldsymbol{q}\in\mathcal{Q} && \text{(4b)}\\
& \Delta_{\boldsymbol{\ell},k} \ge u_{\boldsymbol{\ell}} f_{\boldsymbol{\ell},k} - \eta_{\boldsymbol{\ell}} && \forall \boldsymbol{\ell}\in\mathcal{L},\ k\in\mathcal{K} && \text{(4c)}\\
& \eta_{\boldsymbol{\ell}} = u_{\boldsymbol{\ell}}\zeta && \forall \boldsymbol{\ell}\in\mathcal{L} && \text{(4d)}\\
& \Delta_{\boldsymbol{\ell},k} \ge 0 && \forall \boldsymbol{\ell}\in\mathcal{L},\ k\in\mathcal{K} && \text{(4e)}\\
& \boldsymbol{u} \ge 0 && && \text{(4f)}\\
& \mathbf{1}^T\boldsymbol{u} = 1 && && \text{(4g)}\\
& 0 \le \zeta \le \bar{\zeta}, && && \text{(4h)}
\end{aligned}$$

where $\bar{\zeta} := \max_{k\in\mathcal{K}} f_{\mathbf{0},k}$.

Proof. See Appendix A.1. □

The above problem is non-convex due to the bilinear terms $u_{\boldsymbol{\ell}}\zeta$. A second difficulty resides in having to compute $f_{\boldsymbol{\ell},k}$ for each scenario $k \in \mathcal{K}$ and interdiction plan $\boldsymbol{\ell} \in \mathcal{L}$ in order to solve (4). This will be addressed algorithmically in Section 4.

In the following example, we show that the distributionally robust model proposed in Loizou (2015) identifies interdiction strategies which are strictly stochastically dominated by other feasible strategies. This indicates that Loizou's approach is not well motivated for this class of problems.

Example 1 *Consider an agent trying to reduce the maximum flow from point s to point t which are located on two respective sides of a river. The agent has a budget to interdict two routes. There are three routes available, $e \in \{T1,\ T2,\ B\}$, where routes $T1$ and $T2$ use two different tunnels to pass the river, while route B uses a bridge to do so. In normal traffic conditions, the capacities are, τ, τ, and ϵ, with $\frac{2\tau}{3} < \epsilon < \tau$, for routes $T1$, $T2$, and B respectively. Unfortunately, all three routes are susceptible to congestion on the day of interest. In the case of $T1$ and $T2$, it is known that the city is planning to do some repairs, which would decrease the flow by δ, using one repair team but no information is available regarding which tunnel it will be; the identity of the selected tunnel is denoted by $r \in \{1,\ 2\}$. On the other hand, there is also a weather forecast that predicts 50% chance of snow fall which would*

create the same decrease in flow, δ, on the bridge used by route B. We let δ satisfy $2(\tau-\epsilon)<\delta\leq\epsilon$. The flow player wants to maximize the flow from point s to point t.

Let the set of feasible interdiction plans be defined as follows:

$$\mathcal{L}=\{\{T1,B\},\{T2,B\},\{T1,T2\},\{T1\},\{T2\},\{B\},\{\varnothing\}\}.$$

The possible scenarios $k=\{1,2,3,4\}$ for the capacity of the three routes $T1$, $T2$ and B are given by:

$$\boldsymbol{c}^1=\begin{bmatrix}\tau-\delta\\ \tau\\ \epsilon-\delta\end{bmatrix},\ \boldsymbol{c}^2=\begin{bmatrix}\tau\\ \tau-\delta\\ \epsilon-\delta\end{bmatrix},\ \boldsymbol{c}^3=\begin{bmatrix}\tau-\delta\\ \tau\\ \epsilon\end{bmatrix},\ \boldsymbol{c}^4=\begin{bmatrix}\tau\\ \tau-\delta\\ \epsilon\end{bmatrix},$$

with respective probabilities q_1,q_2,q_3,q_4, such that $q_1+q_2=0.5$ and $q_3+q_4=0.5$.

Here are the numerical details regarding $f(\boldsymbol{\ell},k)$ which denotes the total flow when interdiction plan $\boldsymbol{\ell}$ is chosen, and scenario k is realized:

$$f(\boldsymbol{\ell},k):=\begin{cases}\tau & \text{if } \boldsymbol{\ell}=\{T1,B\} \text{ and } k\in\{1,3\},\\ \tau-\delta & \text{if } \boldsymbol{\ell}=\{T1,B\} \text{ and } k\in\{2,4\},\\ \tau-\delta & \text{if } \boldsymbol{\ell}=\{T2,B\} \text{ and } k\in\{1,3\},\\ \tau & \text{if } \boldsymbol{\ell}=\{T2,B\} \text{ and } k\in\{2,4\},\\ \epsilon-\delta & \text{if } \boldsymbol{\ell}=\{T1,T2\} \text{ and } k\in\{1,2\},\\ \epsilon & \text{if } \boldsymbol{\ell}=\{T1,T2\} \text{ and } k\in\{3,4\},\\ \tau+\epsilon-\delta & \text{if } \boldsymbol{\ell}=\{T1\} \text{ and } k\in\{1,4\},\\ \tau+\epsilon-2\delta & \text{if } \boldsymbol{\ell}=\{T1\} \text{ and } k=2,\\ \tau+\epsilon & \text{if } \boldsymbol{\ell}=\{T1\} \text{ and } k=3,\\ \tau+\epsilon-2\delta & \text{if } \boldsymbol{\ell}=\{T2\} \text{ and } k=1,\\ \tau+\epsilon-\delta & \text{if } \boldsymbol{\ell}=\{T2\} \text{ and } k\in\{2,3\},\\ \tau+\epsilon & \text{if } \boldsymbol{\ell}=\{T2\} \text{ and } k=4,\\ 2\tau-\delta & \text{if } \boldsymbol{\ell}=\{B\} \text{ and } k\in\{1,2,3,4\},\\ 2(\tau-\delta)+\epsilon & \text{if } \boldsymbol{\ell}=\{\varnothing\} \text{ and } k\in\{1,2\},\\ 2\tau-\delta+\epsilon & \text{if } \boldsymbol{\ell}=\{\varnothing\} \text{ and } k\in\{3,4\}.\end{cases}.$$

Consider two potential strategies:

$$\boldsymbol{u}^L=(0.5,0.5,0,0,0,0,0),\qquad \boldsymbol{u}^{SD}=(0,0,1,0,0,0,0),$$

where $\boldsymbol{u}^L$ denotes the strategy to interdict routes $(T1,B)$ with 50% probability and routes $(T2,B)$ with 50% probability; $\boldsymbol{u}^{SD}$ denotes the strategy to interdict routes $(T1,T2)$ with probability 1.

Our robust risk-averse approach $g^{SD}(\boldsymbol{u}):=\sup_{\boldsymbol{q}\in\mathcal{Q}} CVaR^{\alpha}_{\boldsymbol{\ell}\sim\boldsymbol{u},k\sim\boldsymbol{q}}[f(\boldsymbol{\ell},k)]$, with $\alpha\geq 50\%$ leads to the following evaluation:

$$g^{SD}(\boldsymbol{u}^L)=\tau \qquad \textit{v.s.} \qquad g^{SD}(\boldsymbol{u}^{SD})=\epsilon.$$

Since $\epsilon<\tau$, we get that $\boldsymbol{u}^{SD}$, interdicting both routes $T1$ and $T2$, is optimal.

Alternatively, the approach in Loizou (2015) can be summarized as minimizing $g^L(\boldsymbol{u}):=\sup_{\boldsymbol{q}\in\mathcal{Q}} CVaR^{\alpha}_{k\sim\boldsymbol{q}}[\mathbb{E}_{\boldsymbol{\ell}\sim\boldsymbol{u}}[f(\boldsymbol{\ell},k)]]$ and leads to the following evaluation

$$g^L(\boldsymbol{u}^L)=\tau-\delta/2 \qquad \textit{v.s.} \qquad g^L(\boldsymbol{u}^{SD})=\epsilon.$$

Since $\delta>2(\tau-\epsilon)$, the optimal decision in this case is to implement $\boldsymbol{u}^L$, i.e. interdicting $\{T1,B\}$ with 50% probability and $\{T2,B\}$ with 50% probability, is optimal. Yet, it is clear from a purely statistical point of view that $\boldsymbol{u}^{SD}$ strictly stochastically dominates $\boldsymbol{u}^L$ for all $\boldsymbol{q}\in\mathcal{Q}$. Specifically, we have that

$$\begin{aligned}\forall t\in\mathbb{R},\ \mathbb{P}_{\boldsymbol{\ell}\sim\boldsymbol{u}^L,k\sim\boldsymbol{q}}(f(\boldsymbol{\ell},k)\geq t)&=0.5\mathbb{1}_{t\leq\tau}+0.5\mathbb{1}_{t\leq\tau-\delta}\\ &\geq 0.5\mathbb{1}_{t\leq\epsilon}+0.5\mathbb{1}_{t\leq\epsilon-\delta}=\mathbb{P}_{\boldsymbol{\ell}\sim\boldsymbol{u}^{SD},k\sim\boldsymbol{q}}(f(\boldsymbol{\ell},k)\geq t)\,,\end{aligned}$$

where $\mathbb{1}_x$ denotes the indicator function of set x, and where inequality is strict when $\epsilon<t\leq\tau$ or $\epsilon-\delta<t\leq\tau-\delta$.

4 Solving the DRNI problem

In this section, we propose a numerical scheme for solving the DRNI problem. It is well-known that the tractability of a robust optimization problem depends on the structure of the uncertainty set (see Ben-Tal and Nemirovski (2008) and Ben-Tal et al. (2015)). In particular, for simplicity of exposition, we will focus on the case where the distribution ambiguity set contains perturbed versions of a reference distribution $\hat{\boldsymbol{q}} \in \Delta\mathcal{K}$.

Assumption 1 *Let $\mathcal{Q}$ be defined as follows:*

$$\mathcal{Q} := \{\boldsymbol{q} \in \mathbb{R}^{|\mathcal{K}|} \ | \exists \boldsymbol{z} \in \mathcal{Z}(\Gamma), \ \ \boldsymbol{q} \geq 0, \sum_{k=1}^{|\mathcal{K}|} q_k = 1, \boldsymbol{q} = \hat{\boldsymbol{q}} + \operatorname{diag}(\bar{\boldsymbol{q}})\boldsymbol{z}\},$$

where $\hat{\boldsymbol{q}} \in \Delta\mathcal{K}$ is a reference distribution, e.g., $\hat{\boldsymbol{q}} = \frac{1}{|\mathcal{K}|}\mathbf{1}$, while $\bar{\boldsymbol{q}}$ models the magnitude of potential perturbations. Moreover, the set of perturbations $\mathcal{Z}$ refers to the following "budgeted uncertainty set":

$$\mathcal{Z}(\Gamma) = \{\boldsymbol{z} \in \mathbb{R}^{|\mathcal{K}|} | -1 \leq \boldsymbol{z} \leq 1, \sum_{k=1}^{|\mathcal{K}|} |z_k| \leq \Gamma\},$$

where Γ denotes the maximum number of terms of $\hat{\boldsymbol{q}}$ that can be perturbed.

The budgeted uncertainty set introduced in Bertsimas and Sim (2004) provides the flexibility to study the trade-off between robustness and performance. In Bertsimas and Sim (2003) and Atamtürk and Zhang (2007), a budgeted uncertainty set was used to study network flow problems with data uncertainty. For a given budget, the maximum number of parameters which can deviate from their nominal values is controlled with the budget Γ. A valuable property of the budgeted uncertainty set is that the robust counterpart of a linear constraint is representable in a linear program. This implies that the non-linear bi-convex DRO problem (4) can be reformulated as a finite dimensional bi-convex optimization problem.

Proposition 2 *The robust counterpart of the interdictor's DRNI problem presented in (4) is given by*

$$\begin{aligned}
\underset{\substack{\boldsymbol{u}, t, \boldsymbol{w}, \boldsymbol{w}^-, \\ \boldsymbol{\eta}, \chi, \boldsymbol{\beta}, \Delta, \zeta}}{\text{minimize}} \quad & t && && \text{(5a)}\\
\text{subject to} \quad & \zeta + \sum_{k' \in \mathcal{K}} w_{k'} + \sum_{k' \in \mathcal{K}} w^-_{k'} + \Gamma\chi \\
& + \sum_{k' \in \mathcal{K}} \hat{q}_{k'}\beta_{k'} + \frac{1}{1-\alpha}\sum_{\boldsymbol{\ell} \in \mathcal{L}} \Delta_{\boldsymbol{\ell},k} - \beta_k \leq t && \forall k \in \mathcal{K}, && \text{(5b)}\\
& \chi \geq \bar{q}_k\beta_k - w_k && \forall k \in \mathcal{K} && \text{(5c)}\\
& \chi \geq -\bar{q}_k\beta_k - w^-_k && \forall k \in \mathcal{K} && \text{(5d)}\\
& \boldsymbol{w} \geq 0, \boldsymbol{w}^- \geq 0, \chi \geq 0 && && \text{(5e)}\\
& \text{(4c)} - \text{(4h)}.
\end{aligned}$$

Proof. See Appendix A.2. □

The procedure that we propose for solving problem (5) is motivated by the observation that ζ and $\boldsymbol{u}$ are complicating variables. Indeed, when either ζ or $\boldsymbol{u}$ is fixed, the problem reduces to a linear program. Since ζ is also constrained to lie in a bounded interval $\bar{\mathcal{I}} := [0, \bar{\zeta}]$, a spatial branch and bound scheme on ζ seems appropriate (see Chandraker and Kriegman (2008)).

Our implementation of the spatial branch and bound algorithm will rely on the existence of two operators. Namely, after defining the problem that we are interested in solving as

$$\begin{aligned} g(\bar{\mathcal{I}}) := \quad & \min_{\substack{\boldsymbol{u}, t, \boldsymbol{w}, \boldsymbol{w}^-, \\ \boldsymbol{\eta}, \chi, \boldsymbol{\beta}, \Delta, \zeta}} \quad t \\ & \text{subject to} \quad (4c) - (4g), (5b) - (5e), \\ & \qquad\qquad \zeta \in \bar{\mathcal{I}}\,, \end{aligned}$$

the algorithm will assume the existence of the following two bounding operators $g_{\text{lb}}(\mathcal{I})$ and $g_{\text{ub}}(\mathcal{I})$ which satisfy:

$$g_{\text{lb}}(\mathcal{I}) \leq g(\mathcal{I}) \leq g_{\text{ub}}(\mathcal{I})\,, \forall \mathcal{I} \subseteq \bar{\mathcal{I}}\,,$$

and such that for all sequence of intervals $\mathcal{I}_1, \mathcal{I}_2, \ldots$ converging to some ζ, the associated sequence of bounds $(g_{\text{lb}}(\mathcal{I}_j), g_{\text{ub}}(\mathcal{I}_j))$ converge to $g(\zeta)$. Finally, the operator $g_{\text{ub}}(\mathcal{I})$ will be such that one can always efficiently produce a $\zeta^*_{\text{ub}}(\mathcal{I}) \in \mathcal{I}$ such that $g(\{\zeta^*_{\text{ub}}\}) = g_{\text{ub}}(\mathcal{I})$.

With this in hand, we can describe the algorithm. First in words, the spatial branch and bound algorithm starts at a root node capturing $\bar{\mathcal{I}}$ and branches on this node by subdividing it into a number of sub-intervals, considered sub-nodes of the branch and bound tree. [1] Nodes are progressively selected and branched upon until either at node j, we have that $g_{\text{lb}}(\mathcal{I}_j)$ is close enough to $g_{\text{ub}}(\mathcal{I}_j)$, or $g_{\text{lb}}(\mathcal{I}_j)$ is larger than $g(\{\hat{\zeta}^*\})$ where $\hat{\zeta}^*$ is the best solution found so far. When no more nodes need to be branched upon, the algorithm can terminate and conclude that $\{\hat{\zeta}^*\}$ is close enough to being globally optimal. Based on $\hat{\zeta}^*$ it is then possible to get a nearly optimal solution $\hat{\boldsymbol{u}}^*$ from problem (5) where ζ is fixed to $\hat{\zeta}^*$. Finally, the exact worst-case CVaR of $\hat{\boldsymbol{u}}^*$ can be obtained by solving problem (5) where $\boldsymbol{u}$ is fixed to $\hat{\boldsymbol{u}}^*$. For clarity, Algorithm 1 presents the pseudocode for the procedure that was described.

Algorithm 1 Spatial branch and bound algorithm for solving problem (5)

procedure SpatialBranch&Bound(ϵ,n)
  $UB^* \leftarrow \infty$, $\mathcal{N} \leftarrow \{\bar{\mathcal{I}}\}$
  while $\mathcal{N} \neq \emptyset$ **do**
    Sort $\mathcal{N}$ according to $g_{\text{lb}}(\cdot)$
    Take first $\mathcal{I}$ out of $\mathcal{N}$
    if $g_{\text{ub}}(\mathcal{I}) < UB^*$ **then**
      $UB^* \leftarrow g_{\text{ub}}(\mathcal{I})$
      $\hat{\zeta}^* \leftarrow \zeta^*_{\text{ub}}(\mathcal{I})$
    end if
    Remove from $\mathcal{N}$, all $\mathcal{I}$ such that $g_{\text{lb}}(\mathcal{I}) \geq UB^*$
    if $g_{\text{ub}}(\mathcal{I}) > (1+\epsilon) g_{\text{lb}}(\mathcal{I}))$ **then**
      Divide $\mathcal{I}$ into n sub-intervals $\{\mathcal{I}_1, \ldots, \mathcal{I}_n\}$
      $\mathcal{N} \leftarrow \mathcal{N} \cup \{\mathcal{I}_1, \ldots, \mathcal{I}_n\}$
    end if
  end while
  Solve problem (5) with constraint $\zeta = \hat{\zeta}^*$ to get $\hat{\boldsymbol{u}}^*$
  Solve problem (5) with constraint $\boldsymbol{u} = \hat{\boldsymbol{u}}^*$ to get $\hat{t}^*$
  return $\hat{\boldsymbol{u}}^*, \hat{t}^*$
end procedure

We are left with describing how the two operators can be efficiently implemented.

4.1 Using RRLT with C&CG for $g_{\text{lb}}(\mathcal{I})$

In this section, we describe an efficient procedure that can be used to establish a lower bound for the optimal value of problem (5). This procedure will need to overcome the two underlying difficulties of problem (5), namely that constraint (4d) is bilinear in $\boldsymbol{u}$ and ζ, and that the size of this problem is exponential with respect to $|\mathcal{K}|$ due to the set $\mathcal{L}$. To tackle the first obstacle, we will employ a popular reduced reformulation linearization technique (see Liberti and Pantelides (2006)) that will relax the

[1] Implementation details: for any interval $\tilde{\mathcal{I}} := [\zeta_{\text{lb}},\ \zeta_{\text{ub}}]$, we create sub-intervals $[\zeta_{\text{lb}},\ \zeta_{\text{lb}} + p(\zeta_{\text{ub}} - \zeta_{\text{lb}})]$ and $[\zeta_{\text{lb}} + p(\zeta_{\text{ub}} - \zeta_{\text{lb}}),\ \zeta_{\text{ub}}]$, where $p = 0.2$ if $(\hat{\zeta}^* - \zeta_{\text{lb}}) < (\zeta_{\text{ub}} - \hat{\zeta}^*)$; $p = 0.8$ otherwise

problem to a linear program. The second obstacle will be dealt with by employing a column generation scheme (Desrosiers and Lübbecke (2005)) that only considers a subset $\hat{\mathcal{L}} \subset \mathcal{L}$ and progressively adds relevant support points to it until optimality conditions are satisfied.

Starting with the idea of relaxing the problem to a linear program, we follow similar steps as used in Liberti and Pantelides (2006). Namely, we start by introducing a set of redundant constraints in problem (5). This gives rise to the following equivalent optimization model

$$
\begin{aligned}
\underset{\substack{\boldsymbol{u},t,\boldsymbol{w},\boldsymbol{w}^-,\\ \boldsymbol{\eta},\chi,\boldsymbol{\beta},\Delta,\zeta}}{\text{minimize}} \quad & t && && \text{(6a)}\\
\text{subject to} \quad & (4c)-(4g),(5b)-(5e), && && \\
& \sum_{\boldsymbol{\ell}\in\mathcal{L}} \eta_{\boldsymbol{\ell}} = \zeta && && \text{(6b)}\\
& \eta_{\boldsymbol{\ell}} \geq u_{\boldsymbol{\ell}}\zeta_{\text{lb}} && \forall \boldsymbol{\ell}\in\mathcal{L} && \text{(6c)}\\
& \eta_{\boldsymbol{\ell}} \leq u_{\boldsymbol{\ell}}\zeta_{\text{ub}} && \forall \boldsymbol{\ell}\in\mathcal{L} && \text{(6d)}\\
& \eta_{\boldsymbol{\ell}} \geq \zeta + \zeta_{\text{ub}}(u_{\boldsymbol{\ell}}-1) && \forall \boldsymbol{\ell}\in\mathcal{L} && \text{(6e)}\\
& \eta_{\boldsymbol{\ell}} \leq \zeta + \zeta_{\text{lb}}(u_{\boldsymbol{\ell}}-1) && \forall \boldsymbol{\ell}\in\mathcal{L}, && \text{(6f)}\\
& \zeta_{\text{lb}} \leq \zeta \leq \zeta_{\text{ub}}, && && \text{(6g)}
\end{aligned}
$$

where ζ_{lb} and ζ_{ub} are the respective boundaries of $\mathcal{I}$. In problem (6), constraint (6b) is redundant since $\sum_{\boldsymbol{\ell}\in\mathcal{L}} u_\ell = 1$ implies that $\sum_{\boldsymbol{\ell}\in\mathcal{L}} u_\ell\zeta = \zeta$, and we have that $u_\ell\zeta = \eta_\ell$. On the other hand, constraints (6c)–(6f) are so-called McCormick inequalities (see McCormick (1976)) which are known to be redundant given that $\zeta \in \mathcal{I}$ and $0 \leq \boldsymbol{u} \leq 1$.

We obtain the linear relaxation of the above problem by removing constraint (4d) from problem (6). For completeness, we present this linear programming relaxation in full details below:

$$
\begin{aligned}
\underset{\substack{\boldsymbol{u},t,\boldsymbol{w},\boldsymbol{w}^-,\\ \boldsymbol{\eta},\chi,\boldsymbol{\beta},\Delta,\zeta}}{\text{minimize}} \quad & t && && \text{(7a)}\\
\text{subject to} \quad & \zeta + \sum_{k'\in\mathcal{K}} w_{k'} + \sum_{k'\in\mathcal{K}} w^-_{k'} + \Gamma\chi && && \\
& \quad + \sum_{k'\in\mathcal{K}} \hat{q}_{k'}\beta_{k'} + \frac{1}{1-\alpha}\sum_{\boldsymbol{\ell}\in\mathcal{L}} \Delta_{\boldsymbol{\ell},k} - \beta_k \leq t && \forall k\in\mathcal{K}, && \text{(7b)}\\
& \chi \geq \bar{q}_k\beta_k - w_k && \forall k\in\mathcal{K} && \text{(7c)}\\
& \chi \geq -\bar{q}_k\beta_k - w^-_k && \forall k\in\mathcal{K} && \text{(7d)}\\
& \mathbf{1}^T\boldsymbol{u} = 1 && && \text{(7e)}\\
& \sum_{\boldsymbol{\ell}\in\mathcal{L}} \eta_{\boldsymbol{\ell}} = \zeta && && \text{(7f)}\\
& \eta_{\boldsymbol{\ell}} \geq \zeta + \zeta_{\text{ub}}(u_{\boldsymbol{\ell}}-1) && \forall \boldsymbol{\ell}\in\mathcal{L} && \text{(7g)}\\
& \eta_{\boldsymbol{\ell}} \leq \zeta + \zeta_{\text{lb}}(u_{\boldsymbol{\ell}}-1) && \forall \boldsymbol{\ell}\in\mathcal{L}, && \text{(7h)}\\
& \zeta_{\text{lb}} \leq \zeta \leq \zeta_{\text{ub}} && && \text{(7i)}\\
& \boldsymbol{w} \geq 0, \boldsymbol{w}^- \geq 0, \chi \geq 0 && && \text{(7j)}\\
& \Delta_{\boldsymbol{\ell},k} \geq u_{\boldsymbol{\ell}} f_{\boldsymbol{\ell},k} - \eta_{\boldsymbol{\ell}} && \forall \boldsymbol{\ell}\in\mathcal{L},\, k\in\mathcal{K} && \text{(7k)}\\
& \Delta_{\boldsymbol{\ell},k} \geq 0, && \forall \boldsymbol{\ell}\in\mathcal{L}\, k\in\mathcal{K} && \text{(7l)}\\
& \boldsymbol{u} \geq 0 && && \text{(7m)}\\
& \eta_{\boldsymbol{\ell}} \geq u_{\boldsymbol{\ell}}\zeta_{\text{lb}} && \forall \boldsymbol{\ell}\in\mathcal{L} && \text{(7n)}\\
& \eta_{\boldsymbol{\ell}} \leq u_{\boldsymbol{\ell}}\zeta_{\text{ub}} && \forall \boldsymbol{\ell}\in\mathcal{L}. && \text{(7o)}
\end{aligned}
$$

In what follows, it will be useful to compactly represent (7) as follows:

$$\underset{\boldsymbol{x},\{\boldsymbol{y_\ell}\}_{\ell\in\mathcal{L}}}{\text{minimize}} \quad \boldsymbol{h}^T\boldsymbol{x} \tag{8a}$$

$$\text{subject to} \quad A\boldsymbol{x} + \sum_{\boldsymbol{\ell}\in\mathcal{L}} B_{\boldsymbol{\ell}}\boldsymbol{y_\ell} \leq \boldsymbol{s} \tag{8b}$$

$$W_{\boldsymbol{\ell}}\boldsymbol{y_\ell} \leq \boldsymbol{0} \qquad \forall \boldsymbol{\ell} \in \mathcal{L}, \tag{8c}$$

where (8b) summarizes constraints (7b), (7c), (7d), (7e), (7f), (7g), (7h), (7i), (7j) while (8c) summarizes constraints (7k), (7l), (7m), (7n), (7o). The decision variables are given by

$$\boldsymbol{x} = [\boldsymbol{\beta}^T \ \boldsymbol{w}^T \ \boldsymbol{w}^{-T} \ \zeta \ \chi \ t]^T,$$
$$\boldsymbol{y_\ell} = [\boldsymbol{\Delta}_{\boldsymbol{\ell}}^T \ u_{\boldsymbol{\ell}} \ \eta_{\boldsymbol{\ell}}]^T, \qquad \forall \boldsymbol{\ell} \in \mathcal{L}$$

where $\boldsymbol{x} \in \mathbb{R}^{3(|\mathcal{K}|+1)}$, $\boldsymbol{y_\ell} \in \mathbb{R}^{|\mathcal{K}|+2}$ and $\boldsymbol{\Delta_\ell} \in \mathbb{R}^{|\mathcal{K}|}$ for each $\boldsymbol{\ell} \in \mathcal{L}$, $A \in \mathbb{R}^{(5|\mathcal{K}|+2|\mathcal{L}|+7)\times(3(|\mathcal{K}|+1))}$, $B_{\boldsymbol{\ell}} \in \mathbb{R}^{(5|\mathcal{K}|+2|\mathcal{L}|+7)\times(|\mathcal{K}|+2)}$, $W_{\boldsymbol{\ell}} \in \mathbb{R}^{(2|\mathcal{K}|+3)\times(|\mathcal{K}|+2)}$, $\boldsymbol{s} \in \mathbb{R}^{5|\mathcal{K}|+2|\mathcal{L}|+7}$, $\boldsymbol{h} \in \mathbb{R}^{3(|\mathcal{K}|+1)}$. For the full description of A, B, W, s, h, refer to Appendix B.

The idea behind column generation methods is to consider that at optimality $\boldsymbol{y_\ell} \neq 0$ only for a small set of index $\ell \in \mathcal{L}$. This is a legitimate assumption for our DRNI problem where we expect that there should be an optimal strategy that only randomizes among a relatively small (non-exponential) number of interdiction plans. This was observed for instance in the distributionally robust risk neutral facility location problem studied in Delage and Saif (2018).

Given a set $\hat{\mathcal{L}} \subseteq \mathcal{L}$, by linear programming duality, we have that the solution of

$$\underset{\boldsymbol{x},\{\boldsymbol{y_\ell}\}_{\ell\in\mathcal{L}}}{\text{minimize}} \quad \boldsymbol{h}^T\boldsymbol{x} \tag{9a}$$

$$\text{subject to} \quad A\boldsymbol{x} + \sum_{\boldsymbol{\ell}\in\mathcal{L}} B_{\boldsymbol{\ell}}\boldsymbol{y_\ell} \leq \boldsymbol{s} \tag{9b}$$

$$W_{\boldsymbol{\ell}}\boldsymbol{y_\ell} \leq \boldsymbol{0} \qquad \forall \boldsymbol{\ell} \in \hat{\mathcal{L}} \tag{9c}$$

$$\boldsymbol{y}_\ell = 0 \qquad \forall \ell \in \mathcal{L}/\hat{\mathcal{L}}, \tag{9d}$$

is optimal with respect to problem (8) if and only if a solution of its dual problem

$$\underset{\boldsymbol{\psi},\{\boldsymbol{\sigma_\ell}\}_{\boldsymbol{\ell}\in\hat{\mathcal{L}}}}{\text{maximize}} \quad -\boldsymbol{\psi}^T\boldsymbol{s} \tag{10a}$$

$$\text{subject to} \quad \boldsymbol{h} + A^T\boldsymbol{\psi} = \boldsymbol{0} \tag{10b}$$

$$B_{\boldsymbol{\ell}}^T\boldsymbol{\psi} + W_{\boldsymbol{\ell}}^T\boldsymbol{\sigma_\ell} = \boldsymbol{0} \qquad \forall \boldsymbol{\ell} \in \hat{\mathcal{L}} \tag{10c}$$

$$\boldsymbol{\psi} \geq 0, \ \boldsymbol{\sigma_\ell} \geq 0 \qquad \forall \boldsymbol{\ell} \in \hat{\mathcal{L}}, \tag{10d}$$

where $\boldsymbol{\psi} \in \mathbb{R}^{5|\mathcal{K}|+2|\mathcal{L}|+7}$ and $\boldsymbol{\sigma_\ell} \in \mathbb{R}^{2|\mathcal{K}|+3}$ are the dual variables associated to constraints (9b) and (9c) respectively, can be completed with some $\boldsymbol{\sigma_\ell} \in \mathbb{R}^{2|\mathcal{K}|+3}$ for all $\boldsymbol{\ell} \in \mathcal{L}/\hat{\mathcal{L}}$ in a way that makes it feasible in the dual of problem (8), i.e. problem (10) where $\hat{\mathcal{L}}$ is replaced with $\mathcal{L}$.

In particular, this can be verified after solving problem (9) for some $\hat{\mathcal{L}} \subseteq \mathcal{L}$ by obtaining a set $(\hat{\boldsymbol{\psi}}, \{\hat{\boldsymbol{\sigma}}_{\boldsymbol{\ell}}\}_{\ell\in\hat{\mathcal{L}}})$ of optimal dual variables for constraints (9b) and (9c) and verifying if they satisfy the following condition:

$$\inf_{\boldsymbol{\ell}\in\mathcal{L}} \sup_{\boldsymbol{\sigma_\ell}\geq 0} \inf_{\boldsymbol{y_\ell}} \boldsymbol{\psi}^{*T} B_{\boldsymbol{\ell}}\boldsymbol{y_\ell} + \boldsymbol{\sigma}_{\boldsymbol{\ell}}^T W_{\boldsymbol{\ell}}\boldsymbol{y_\ell} \geq 0. \tag{11}$$

Furthermore, when condition (11) is not satisfied, a violating $\boldsymbol{\ell} \in \mathcal{L}$, which is necessarily not in $\hat{\mathcal{L}}$, can be identified and added to $\hat{\mathcal{L}}$ in order to improve the solution obtained by problem (9).

Two important observations need to be made at this point. First, fortunately enough problem (9) can be shown to reduce to a linear program which size is linear in $|\hat{\mathcal{L}}|$ given that only the decision

variables $(\boldsymbol{x}, \{\boldsymbol{y}_{\boldsymbol{\ell}}\}_{\boldsymbol{\ell}\in\hat{\mathcal{L}}})$ need to be optimized, while the only constraints indexed by some $\boldsymbol{\ell} \in \mathcal{L}/\hat{\mathcal{L}}$ are constraint (9d) and a subset of constraint (9b) which capture constraints (7g) and (7h). The latter two become redundant for any $\boldsymbol{\ell} \notin \hat{\mathcal{L}}$ since in those cases the constraints reduce to:

$$\zeta_{\text{lb}} \leq \zeta \leq \zeta_{\text{ub}} \qquad\qquad \forall \boldsymbol{\ell} \in \mathcal{L}/\hat{\mathcal{L}}.$$

Secondly, one can also show that condition (11) can be verified efficiently by solving a mixed-integer linear program of reasonable size as described in the following proposition.

Proposition 3 *Given some $\mathcal{I} \subseteq \bar{\mathcal{I}}$, verifying condition (11) is equivalent to verifying whether the optimal value of the following mixed-integer linear program is non-negative:*

$$\begin{aligned}
\underset{\substack{\boldsymbol{\ell}\in\mathcal{L},\boldsymbol{\Delta},\eta \\ \{\boldsymbol{\lambda}_k,\boldsymbol{\upsilon}_k,\boldsymbol{\Upsilon}_k\}_{k\in\mathcal{K}}}}{\text{minimize}} \quad & \frac{\boldsymbol{\varphi}^{*T}}{1-\alpha}\boldsymbol{\Delta} + p^* + \pi^*\eta && && \text{(12a)}\\
\text{subject to} \quad & \Delta_k \geq \boldsymbol{c}_k^T(\boldsymbol{\lambda}_k - \boldsymbol{\Upsilon}_k) - \eta && \forall k \in \mathcal{K} && \text{(12b)}\\
& \boldsymbol{\Upsilon}_k \leq \boldsymbol{\ell} && k \in \mathcal{K} && \text{(12c)}\\
& \boldsymbol{\Upsilon}_k \leq \boldsymbol{\lambda}_k && \forall k \in \mathcal{K} && \text{(12d)}\\
& \boldsymbol{\Upsilon}_k \geq \boldsymbol{\lambda}_k + \boldsymbol{\ell} - 1 && \forall k \in \mathcal{K} && \text{(12e)}\\
& \boldsymbol{\Upsilon}_k \geq \boldsymbol{0} && \forall k \in \mathcal{K} && \text{(12f)}\\
& \boldsymbol{\lambda}_k + N^T\boldsymbol{\upsilon}_k - \boldsymbol{d} \geq 0 && \forall k \in \mathcal{K} && \text{(12g)}\\
& 0 \leq \boldsymbol{\lambda}_k \leq 1 && \forall k \in \mathcal{K} && \text{(12h)}\\
& \Delta_k \geq 0 && \forall k \in \mathcal{K} && \text{(12i)}\\
& \eta \geq \zeta_{lb} && && \text{(12j)}\\
& \eta \leq \zeta_{ub} && && \text{(12k)}\\
& \boldsymbol{1}^T\boldsymbol{\ell} \leq B && && \text{(12l)}\\
& \boldsymbol{\ell} \in \{0,1\}^{|E|}, && && \text{(12m)}
\end{aligned}$$

where $\boldsymbol{\lambda}_k \in \mathbb{R}^{|E|}$, $\boldsymbol{\Delta} \in \mathbb{R}^{|\mathcal{K}|}, \eta \in \mathbb{R}$, $\boldsymbol{\upsilon}_k \in \mathbb{R}^{|V|}$ and $\boldsymbol{\Upsilon} \in \mathbb{R}^{|E|\times|\mathcal{K}|}$ while $\boldsymbol{\varphi}^*$, p^*, π^* are the elements of $\boldsymbol{\psi}^*$ associated with (7b), (7e), (7f) respectively.

Proof. See Appendix A.3. □

For completeness, we present the pseudocode of the column generation algorithm described above in Algorithm 2. Note that the algorithm is guaranteed to converge in a finite number of iterations given that at each iteration, either the algorithm terminates or a new element $\boldsymbol{\ell} \in \mathcal{L}$ is added to $\hat{\mathcal{L}}$, yet $|\mathcal{L}|$ is finite.

Algorithm 2 Column Generation Algorithm to solve problem (7)

procedure ColumnGeneration
  $\hat{\mathcal{L}} \leftarrow \{\boldsymbol{0}\}$
  while $\hat{\mathcal{L}} \neq \mathcal{L}$ **do**
    Solve problem (7) with $\eta_{\boldsymbol{\ell}} = 0$,$u_{\boldsymbol{\ell}} = 0$, $\boldsymbol{\Delta}_{\boldsymbol{\ell}} = \boldsymbol{0}$ for all $\boldsymbol{\ell} \in \mathcal{L}/\hat{\mathcal{L}}$
    Identify a set of optimal dual variables $\boldsymbol{\varphi}^*$, p^*, and π^* as defined in Proposition 3.
    Solve problem (12) to obtain optimal value v^* and optimal $\boldsymbol{\ell}^*$
    if $v^* \geq 0$ **then**
      return Optimal solution obtained in step 4
    else
      $\hat{\mathcal{L}} \leftarrow \hat{\mathcal{L}} \cup \{\boldsymbol{\ell}^*\}$
    end if
  end while
  return Solve (7) and return optimal solution
end procedure

4.2 Using coordinate descent for $g_{\text{ub}}(\mathcal{I})$

Given an interval $\mathcal{I} \subseteq \bar{\mathcal{I}}$ we are looking for an upper bound on $g(\mathcal{I})$ and a value of $\zeta \in \mathcal{I}$ such that $g(\mathcal{I})$ matches this upper bound. In order to accomplish this task, we will first look back at the solution of problem (7) to identify the optimal support $\mathcal{L}^*_{\text{lb}}$ and distribution $\boldsymbol{u}^*_{\text{lb}}$ of the lower bounding problem. We then perform coordinate descent on problem (5) where $\mathcal{L}$ is replaced with $\mathcal{L}^*_{\text{lb}}$ iterating between a step where $\boldsymbol{u}$ stays fixed at the best solution found so far, initially at $\boldsymbol{u}^*_{\text{lb}}$, and a step where it is rather ζ that stays fixed. In both cases, the problem reduces to a linear program whose size is linear in the size of $\mathcal{L}^*_{\text{lb}}$. This procedure is considered to have converged when the relative improvement on optimal value is considered small enough. For completeness, we provide the pseudocode for the coordinate descent algorithm in Algorithm 3.

Algorithm 3 Coordinate Descent Algorithm to obtain upper bound on problem (5)

procedure CoordinateDescent(ϵ, $\mathcal{L}^*_{\text{lb}}$, $\boldsymbol{u}^*_{\text{lb}}$)
  $\bar{\boldsymbol{u}}^* \leftarrow \boldsymbol{u}^*_{\text{lb}}$, $\mathcal{L} \leftarrow \mathcal{L}^*_{\text{lb}}$
  do
    Solve problem (5) with constraint $\boldsymbol{u} = \bar{\boldsymbol{u}}^*$ to get optimal value t^*_1 and optimal $\zeta^*_{\text{ub}}(\mathcal{I})$
    Solve problem (5) with constraint $\zeta = \zeta^*_{\text{ub}}(\mathcal{I})$ to get optimal value t^*_2 and optimal $\bar{\boldsymbol{u}}^*$
  while $t^*_2 < (1-\epsilon)t^*_1$
  $g_{\text{ub}}(\mathcal{I}) \leftarrow t^*_1$
  return $g_{\text{ub}}(\mathcal{I})$, $\zeta^*_{\text{ub}}(\mathcal{I})$
end procedure

5 Numerical experiments

We performed a series of numerical experiments to show the convergence and numerical efficiency of the spatial branch and bound algorithm (in Section 5.1), and to compare the in-sample and out-of-sample performance of optimal randomized and deterministic interdiction plan strategies in Section 5.2. All algorithms were implemented in Matlab 2019b using the YALMIP toolbox and CPLEX 12.9.0 was used to solve all continuous and mixed-integer linear programs. The DRNI problem instances were generated using the network structure given in Figure 1. Within the same column, the arcs can point upward or downward with equal probability whereas between different columns, the arcs always point in the direction of the sink. The number of rows is denoted by m and the number of columns is denoted by n. This class of network instances is the same as the one used in Cormican et al. (1998), Janjarassuk and Linderoth (2008) and Atamturk et al. (2017) except that we allow any arc to be interdicted whereas they have assumed that the arcs in the first and last column and those leaving the source node (s) or entering the sink (t) are not interdictable and have infinite capacity. For each DRNI problem instance, capacity vector scenarios are drawn from a factor model, $\boldsymbol{c} := F\boldsymbol{\xi}$, with each ξ_i independently distributed according to an exponential distribution with mean μ_i, for some fixed $F \in \mathbb{R}^{|E|\times 2}$ and $\boldsymbol{\mu} \in \mathbb{R}^2$ that were randomly generated for the given instance.

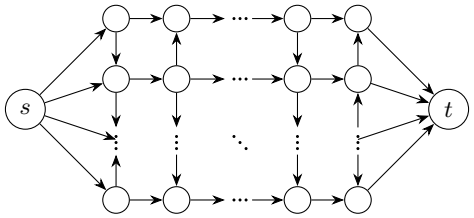

Figure 1: Network for numerical experiments

5.1 Convergence of spatial branch and bound algorithm

In this subsection, we show that the spatial branch and bound algorithm described in Section 4 converges quickly to approximately optimal solutions. We randomly generate 10 instances of sizes

10×10, 20×20 and 30×30 of the DRNI problem as described in the introduction of this section. For each instance, we consider $|K| = 20$ randomly generated scenarios from the constructed factor model. We then solve the instances for a convergence tolerance, ϵ, either equal to 0.01% or 0.0001%, terminating the algorithm if the higher precision is not achieved after 1 hour. The results of the numerical experiments for $\Gamma = 2$, $B = 5$, $\hat{\boldsymbol{q}} = (1/20)\mathbf{1}$, $\bar{\boldsymbol{q}} = \mathbf{1}$, $\alpha = 0.05$ are given in Table 1. It can be seen that for networks of sizes 10×10 and 20×20 , the spatial branch and bound algorithm converges within 1 hour to an accuracy of $\epsilon = 0.0001\%$ for 9 problem instances (out of 10) whereas we found that for networks of size 30×30 the algorithm did not reach this level of accuracy within 1 hr for 2 instances out of 10. On the other hand, for sizes of problem 10×10 and 20×20, all instances converged within less than 1 hour to a precision of $\epsilon = 0.01\%$, reaching on average a gap close to 0.007% and 0.001% respectively in the case of the problematic instances. However, for one instance of problem 30×30, the algorithm did not converge in 1 hr even for 0.01% precision, and the gap was found to be close to 0.25%.

Table 1: Convergence of spatial branch and bound algorithm for 10 randomly generated instances. * avg. cpu time is reported for those instances which converged with .0001% precision in 1 hour

Problem size ($m \times n$)	avg. cpu time (min)*		gap $> 0.0001\%$ after 1 hr	
	$\epsilon = 0.0001\%$	$\epsilon = 0.01\%$	# instances	optimality gap
10×10, $\lvert E\rvert = 200$	2.86	1.85	1	0.007%
20×20, $\lvert E\rvert = 800$	8.32	3.13	1	0.001%
30×30, $\lvert E\rvert = 1800$	7.80	6.44	2	0.14%

5.2 In-sample and out-of-sample performance of randomized and deterministic strategies

To illustrate the benefit of randomization, we focus our attention on the network given in Figure 2 with 4 rows and 2 columns while the factor model continues to be regenerated for each instance. We are interested in comparing the in-sample and out-of-sample performance of strategies that are obtained using only $|\mathcal{K}| = 20$ scenarios from the underlying distribution. To do so, we generated 100 set of samples for $\{\boldsymbol{c}^1, \ldots, \boldsymbol{c}^{20}\}$ for each level of $\Gamma \in \{0,\ 0.1,\ 0.5,\ 1,\ 10,\ 20\}$. For each set of samples, we compare the in-sample performance of both the randomized interdiction strategy $\hat{\boldsymbol{u}}^*_{20}$ and the deterministic strategy $\hat{\ell}^*_{20}$ that optimizes the DRNI problem constructed from this "observed" sample set, with $B = 1$, $\alpha = 0.05$, $\hat{\boldsymbol{q}} = (1/20)\mathbf{1}$, and $\bar{\boldsymbol{q}} = \mathbf{1}$. We then compare the performance of the in-sample optimal strategies $\hat{\boldsymbol{u}}^*_{20}$ and $\hat{\boldsymbol{\ell}}^*_{20}$ on the "unobserved" underlying distribution using a Monte-Carlo simulation of 100000 scenarios. The in-sample optimal randomized strategy $\hat{\boldsymbol{u}}^*_{20}$ is obtained by using the spatial branch and bound algorithm outlined in Section 4 with $\epsilon < 0.01\%$, while the optimal in-sample deterministic strategy $\hat{\boldsymbol{\ell}}^*_{20}$ is computed by solving the MILP given in Appendix C.

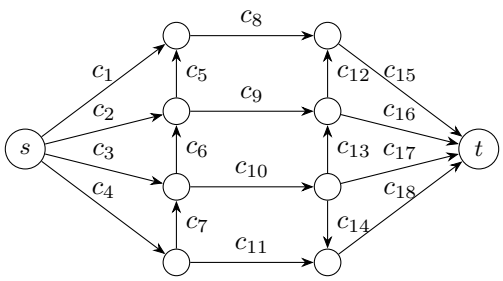

Figure 2: Example network

In terms of in-sample performance, we are interested in comparing the added value of the randomized strategy, which can be defined as the relative gap between the worst-case CVaR obtained by the

deterministic and randomized strategies:

$$\text{VRS} = \frac{\max_{\boldsymbol{q}\in\hat{\mathcal{Q}}_{20}} \text{CVaR}^{\alpha}_{k\sim\boldsymbol{q}}[f_{\hat{\boldsymbol{\ell}}^{*}_{20},k}] - \max_{\boldsymbol{q}\in\hat{\mathcal{Q}}_{20}} \text{CVaR}^{\alpha}_{\boldsymbol{\ell}\sim\hat{\boldsymbol{u}}^{*}_{20},k\sim\boldsymbol{q}}[f_{\boldsymbol{\ell},k}]}{\max_{\boldsymbol{q}\in\hat{\mathcal{Q}}_{20}} \text{CVaR}^{\alpha}_{\boldsymbol{\ell}\sim\hat{\boldsymbol{u}}^{*}_{20},k\sim\boldsymbol{q}}[f_{\boldsymbol{\ell},k}]} \times 100\%.$$

where $\hat{\mathcal{Q}}_{20}$ is centered at the empirical distribution over the sample set. Table 2 reports the number of instances, for each level of Γ, for which VRS was equal to zero, between zero and 1%, and above 1%. For the later case, the average VRS is also reported. We find that, for $\Gamma \in \{0,\ 0.1\}$, all 100 instances achieved a VRS below 1%. This is reasonable given that, as Γ goes to 0, the DRNI problem converges to CVaR minimization with known distribution for which it is known that deterministic strategies are optimal. We also observe that as distributional ambiguity is increased, the randomized strategies perform significantly better than deterministic ones with respect to the in-sample DRNI problem instance, reaching for $\Gamma = 20$ an average VRS of 4.67%.

Table 2: In-sample value of randomized solution performances

Γ	VRS = 0	0 < VRS < 1%	VRS ≥ 1%	
	# instances	# instances	# instances	avg. VRS
0	100	0	0	0
0.1	98	2	0	0
0.5	94	4	2	2.09%
1	89	5	6	3.48%
10	78	12	10	4.45%
20	78	14	8	4.67%

Table 3: Comparison of out-of-sample CVaR performance for deterministic and randomized strategy for instances where randomized strategy improved on the in-sample performance of deterministic strategy by at least 1%

Γ	# of instances	avg. CVaR_r	avg. CVaR_d	# of instances $\text{CVaR}_r < \text{CVaR}_d$	$(\text{CVaR}_r - \text{CVaR}_d)/\text{CVaR}_d$
0.5	2	1.09	1.16	2	36%
1	6	0.84	1.07	6	19%
10	10	0.77	0.93	10	17%
20	8	0.64	0.78	8	19%

In terms of out-of-sample performance, we only consider problem instances for which the VRS was above 1%, given that otherwise the optimal in-sample randomized and deterministic strategies are too similar. We denote the optimal out-of-sample CVaR corresponding to randomized strategies and deterministic interdiction plans by CVaR_r and CVaR_d respectively. It can be clearly seen from Table 3 that randomized strategies have lower average CVaR than deterministic interdiction plans. In all the out-of-sample instances, we found that the optimal randomized strategies always have lower CVaR than optimal deterministic interdiction plans. We argue that this evidence confirms that there is a real observable benefit for the network interdictor for employing randomized interdiction plans in a risk averse network interdiction game, both in a distributionally robust setting (c.f. the in-sample performance comparison) and in a setting where the network capacity distribution information comes from a limited number of observed realizations.

6 Conclusions

In this paper, we introduced a distributionally robust risk averse network interdiction game to model the strategic interactions between a risk-averse interdictor and the flow player. We solved the interdictor's non-linear bi-convex DRO problem by first reformulating it as a bi-convex optimization problem using LP duality and then devising a spatial branch and bound algorithm. After observing that the optimal randomized strategy can be supported on a small number of interdiction plans in DRO problems, we developed a column generation algorithm that can be used to efficiently determine the convex relaxation of the interdictor's problem. Our numerical experiments showed that 1) our

proposed spatial branch and bound algorithm can efficiently solve distributionally robust interdiction games of reasonable sizes; 2) randomization can be quite effective in reducing the risk exposure obtained from the optimal deterministic interdiction strategy both when comparing in-sample worst-case CVaR and out-of-sample CVaR performances.

Given that Stackelberg games with single leader and multiple followers have been extensively applied in the literature, it would be interesting as future work to extend our algorithm in a way that can address Stackelberg games with a leader that is both risk and ambiguity averse while followers implement a Nash equilibrium that accounts for their respective risk aversion.

A Proofs

A.1 Proof of Proposition 1

We substitute the expression of CVaR in (3) to obtain

$$\underset{\boldsymbol{u}\in\Delta\mathcal{L}}{\text{minimize}} \max_{\boldsymbol{q}\in\mathcal{Q}} \inf_{\zeta} \quad \zeta + \frac{1}{1-\alpha}\sum_{\boldsymbol{\ell}}\sum_{k} q_k u_{\boldsymbol{\ell}}[f_{\boldsymbol{\ell},k} - \zeta]^+. \tag{13}$$

We prove by contradiction that an optimal ζ, denoted by ζ^*, lies between 0 and $\bar{\zeta}$. First, we assume that $\zeta^* = \zeta_L < 0$ is the largest optimal value for ζ. The maximum flow $f_{\boldsymbol{\ell},k}$ for any $\boldsymbol{\ell}$ and k is bounded below by 0, hence for any $\boldsymbol{q} \in \mathcal{Q}$ the CVaR equals

$$\zeta_L + \frac{1}{1-\alpha}\sum_{\boldsymbol{\ell}}\sum_{k} q_k u_{\boldsymbol{\ell}}(f_{\boldsymbol{\ell},k} - \zeta_L) = \frac{(1-\alpha)\zeta_L - \zeta_L}{1-\alpha} + \frac{1}{1-\alpha}\sum_{\boldsymbol{\ell}}\sum_{k} q_k u_{\boldsymbol{\ell}} f_{\boldsymbol{\ell},k}.$$

However, we arrive at a contradiction since $\zeta = 0$ is at least as good as ζ_L:

$$\frac{-\alpha\zeta_L}{1-\alpha} + \frac{1}{1-\alpha}\sum_{\boldsymbol{\ell}}\sum_{k} q_k u_{\boldsymbol{\ell}} f_{\boldsymbol{\ell},k} \geq \frac{1}{1-\alpha}\sum_{\boldsymbol{\ell}}\sum_{k} q_k u_{\boldsymbol{\ell}} f_{\boldsymbol{\ell},k} = 0 + \frac{1}{1-\alpha}\sum_{\boldsymbol{\ell}}\sum_{k} q_k u_{\boldsymbol{\ell}}[f_{\boldsymbol{\ell},k} - 0]^+,$$

for any $\alpha \in [0, 1)$. So, we can conclude that $\zeta^* := 0$ is also optimal.

Alternatively, we can assume that $\zeta^* = \zeta_H > \bar{\zeta}$ is the smallest optimal solution for ζ. In this case, for any $\boldsymbol{q} \in \mathcal{Q}$ we have that CVaR equals:

$$\zeta_H + \frac{1}{1-\alpha}\sum_{\boldsymbol{\ell}}\sum_{k} q_k u_{\boldsymbol{\ell}}[f_{\boldsymbol{\ell},k} - \zeta_H]^+ = \zeta_H,$$

since $f_{\boldsymbol{\ell},k} \leq f_{\mathbf{0},k} \leq \bar{\zeta} < \zeta_H$. Yet, for $\zeta = \bar{\zeta}$, the worst-case CVaR, given by $\bar{\zeta}$, is strictly less than ζ_H which contradicts our assumption that ζ_H is the smallest optimal solution for ζ.

We now turn ourselves to establishing the reformulation presented as problem (4). Let $v(\zeta, \boldsymbol{q}, \boldsymbol{u}) = \zeta + \frac{1}{1-\alpha}\sum_{\boldsymbol{\ell}}\sum_{k} q_k u_{\boldsymbol{\ell}}[f_{\boldsymbol{\ell},k} - \zeta]^+$. Since $v(\zeta, \boldsymbol{q}, \boldsymbol{u})$ is convex in ζ for all $\boldsymbol{q} \in \mathcal{Q}$ while being linear in $\boldsymbol{q}$ for all $\zeta \in [0, \bar{\zeta}]$, and since $[0, \bar{\zeta}]$ is bounded and $\mathcal{Q}$ is convex, it follows from Sion's minimax theorem (see Sion (1958)) that (13) is equivalent to

$$\underset{\boldsymbol{u}\in\Delta\mathcal{L}}{\text{minimize}} \min_{0\leq\zeta\leq\bar{\zeta}} \max_{\boldsymbol{q}\in\mathcal{Q}} \quad \zeta + \frac{1}{1-\alpha}\sum_{\boldsymbol{\ell}}\sum_{k} q_k u_{\boldsymbol{\ell}}[f_{\boldsymbol{\ell},k} - \zeta]^+. \tag{14}$$

Since we are minimizing over $\boldsymbol{u}$ and ζ, we have an equivalent reformulation of (14) given by

$$\begin{aligned} \underset{\substack{\boldsymbol{u}\in\Delta\mathcal{L},\boldsymbol{\Delta}\geq 0,\ \boldsymbol{q}\in\mathcal{Q} \\ 0\leq\zeta\leq\bar{\zeta}}}{\text{minimize}} \max \quad & \zeta + \frac{1}{1-\alpha}\sum_{\boldsymbol{\ell}}\sum_{k} q_k \Delta_{\boldsymbol{\ell},k} \\ \text{subject to} \quad & \Delta_{\boldsymbol{\ell},k} \geq u_{\boldsymbol{\ell}} f_{\boldsymbol{\ell},k} - u_{\boldsymbol{\ell}}\zeta \qquad \forall \boldsymbol{\ell} \in \mathcal{L},\ k \in \mathcal{K}. \end{aligned}$$

Employing an epigraph representation of the above objective function and introducing the decision variable $\eta_{\boldsymbol{\ell}} = u_{\boldsymbol{\ell}}\zeta$, we obtain problem (4).

A.2 Proof of Proposition 2

Substituting $\boldsymbol{q} = \hat{\boldsymbol{q}} + \mathrm{diag}(\bar{\boldsymbol{q}})\boldsymbol{z}$, we obtain that (4b) is equivalent to

$$\max_{\boldsymbol{z}:\boldsymbol{z}\in\mathcal{Z}(\Gamma),\sum_{k\in\mathcal{K}}(\hat{q}_k+\bar{q}_k z_k)=1,\hat{q}_k+\bar{q}_k z_k\geq 0,\forall k\in\mathcal{K}} \zeta + \frac{1}{1-\alpha}\sum_{\boldsymbol{\ell}\in\mathcal{L}}\sum_{k\in\mathcal{K}}(\hat{q}_k+\bar{q}_k z_k)\Delta_{\boldsymbol{\ell},k} - t \leq 0, \tag{15}$$

where we will exploit a well-known equivalent representation of $\mathcal{Z}(\Gamma)$:

$$\mathcal{Z}(\Gamma) = \left\{ \boldsymbol{z}\in\mathbb{R}^{|\mathcal{K}|} \;\middle|\; \exists \boldsymbol{\delta}^+,\boldsymbol{\delta}^-\in\mathbb{R}^{|\mathcal{K}|}, \begin{array}{l} \boldsymbol{z}=\boldsymbol{\delta}^+-\boldsymbol{\delta}^- \\ 0\leq\boldsymbol{\delta}^+\leq 1 \\ 0\leq\boldsymbol{\delta}^-\leq 1 \\ \sum_{k\in\mathcal{K}}\delta_k^+ + \delta_k^- \leq \Gamma \end{array} \right\}.$$

Let us define $\gamma_k := \frac{1}{1-\alpha}\sum_{\boldsymbol{\ell}\in\mathcal{L}}\Delta_{\boldsymbol{\ell},k}$, so that we can rewrite the maximization problem in (15) as follows

$$\begin{aligned}
\underset{\boldsymbol{\delta}^+,\boldsymbol{\delta}^-}{\text{maximize}} \quad & \zeta + \sum_{k\in\mathcal{K}}\left(\gamma_k(\hat{q}_k+\bar{q}_k(\delta_k^+-\delta_k^-))\right) - t && && \text{(16a)}\\
\text{subject to} \quad & 0\leq\boldsymbol{\delta}^+\leq 1 && && \text{(16b)}\\
& 0\leq\boldsymbol{\delta}^-\leq 1 && && \text{(16c)}\\
& \sum_{k\in\mathcal{K}}\delta_k^+ + \delta_k^- \leq \Gamma && && \text{(16d)}\\
& \sum_{k\in\mathcal{K}}\bar{q}_k\left(\delta_k^+-\delta_k^-\right) = 0 && && \text{(16e)}\\
& \hat{q}_k + \bar{q}_k\delta_k^+ - \bar{q}_k\delta_k^- \geq 0 && \forall k\in\mathcal{K}. && \text{(16f)}
\end{aligned}$$

Since $\boldsymbol{\delta}^+ = \boldsymbol{\delta}^- = 0$ is feasible, we can use strong LP duality to obtain an equivalent problem:

$$\begin{aligned}
\underset{\substack{\boldsymbol{w},\boldsymbol{w}^-,\\ \chi,\bar{\mu},\bar{\boldsymbol{\beta}}}}{\text{minimize}} \quad & \zeta + \sum_{k\in\mathcal{K}} w_k + \sum_{k\in\mathcal{K}} w_k^- + \Gamma\chi + \sum_{k\in\mathcal{K}}\left((\gamma_k+\bar{\beta}_k)\hat{q}_k\right) - t && \\
\text{subject to} \quad & \bar{q}_k\gamma_k - w_k - \chi + \bar{q}_k\bar{\mu} + \bar{q}_k\bar{\beta}_k \leq 0 && \forall k\in\mathcal{K}\\
& -\bar{q}_k\gamma_k - w_k^- - \chi - \bar{q}_k\bar{\mu} - \bar{q}_k\bar{\beta}_k \leq 0 && \forall k\in\mathcal{K}\\
& \boldsymbol{w}\geq 0, \boldsymbol{w}^-\geq 0, \chi\geq 0, \bar{\boldsymbol{\beta}}\geq 0, &&
\end{aligned}$$

where $\boldsymbol{w}\in\mathbb{R}^{|\mathcal{K}|}$, $\boldsymbol{w}^-\in\mathbb{R}^{|\mathcal{K}|}$, $\chi\in\mathbb{R}$, $\bar{\mu}\in\mathbb{R}$, and $\bar{\boldsymbol{\beta}}\in\mathbb{R}^{|\mathcal{K}|}$ are the dual variables associated to constraints (16b) to (16f) respectively. Substituting $\boldsymbol{\beta} := \boldsymbol{\gamma} + \bar{\mu} + \bar{\boldsymbol{\beta}}$ and then $\mu := -\bar{\mu}$, we obtain

$$\begin{aligned}
\underset{\substack{\boldsymbol{w},\boldsymbol{w}^-,\\ \chi,\mu,\boldsymbol{\beta}}}{\text{minimize}} \quad & \zeta + \sum_{k\in\mathcal{K}} w_k + \sum_{k\in\mathcal{K}} w_k^- + \Gamma\chi + \sum_{k\in\mathcal{K}}\hat{q}_k\beta_k + \mu - t && && \text{(17a)}\\
\text{subject to} \quad & \mu \geq \gamma_k - \beta_k && \forall k\in\mathcal{K} && \text{(17b)}\\
& \chi \geq \bar{q}_k\beta_k - w_k && \forall k\in\mathcal{K} && \text{(17c)}\\
& \chi \geq -\bar{q}_k\beta_k - w_k^- && \forall k\in\mathcal{K} && \text{(17d)}\\
& \boldsymbol{w}\geq 0, \boldsymbol{w}^-\geq 0, \chi\geq 0. && && \text{(17e)}
\end{aligned}$$

Combining the above problem with (4), we obtain

$$\begin{aligned}
\underset{\substack{\boldsymbol{u},t,\boldsymbol{w},\boldsymbol{w}^-,\\ \boldsymbol{\eta},\chi,\boldsymbol{\beta},\Delta,\zeta}}{\text{minimize}} \quad & t \\
\text{subject to} \quad & \text{(4c)} - \text{(4h)}, \text{(17c)} - \text{(17e)} \\
& \zeta + \sum_{k'\in\mathcal{K}} w_{k'} + \sum_{k'\in\mathcal{K}} w_{k'}^- + \Gamma\chi + \sum_{k'\in\mathcal{K}}\hat{q}_{k'}\beta_{k'} + \frac{1}{1-\alpha}\sum_{\boldsymbol{\ell}\in\mathcal{L}}\Delta_{\boldsymbol{\ell},k} - \beta_k \leq t \qquad \forall k\in\mathcal{K},
\end{aligned}$$

where we were able to replace μ with $\max_{k\in\mathcal{K}}\frac{1}{1-\alpha}\sum_{\boldsymbol{\ell}\in\mathcal{L}}\Delta_{\boldsymbol{\ell},k} - \beta_k$.

A.3 Proof of Proposition 3

We start by repeating the definition of condition (11):

$$\inf_{\boldsymbol{\ell}\in\mathcal{L}} \sup_{\boldsymbol{\sigma_\ell}\geq 0} \inf_{\boldsymbol{y_\ell}} \boldsymbol{\psi}^{*T} B_{\boldsymbol{\ell}} \boldsymbol{y_\ell} + \boldsymbol{\sigma}_{\boldsymbol{\ell}}^T W_{\boldsymbol{\ell}} \boldsymbol{y_\ell} \geq 0.$$

We will first argue that the order of $\sup_{\boldsymbol{\sigma_\ell}\geq 0}$ and $\inf_{\boldsymbol{y_\ell}}$ can be changed without affecting the value that is obtained. In particular,

$$\sup_{\boldsymbol{\sigma_\ell}\geq 0} \inf_{\boldsymbol{y_\ell}} \boldsymbol{\psi}^{*T} B_{\boldsymbol{\ell}} \boldsymbol{y_\ell} + \boldsymbol{\sigma}_{\boldsymbol{\ell}}^T W_{\boldsymbol{\ell}} \boldsymbol{y_\ell} = \sup_{\boldsymbol{\sigma_\ell}\geq 0: B_{\boldsymbol{\ell}}^T \boldsymbol{\psi}^* + W_{\boldsymbol{\ell}}^T \boldsymbol{\sigma_\ell} = 0} 0,$$

while

$$\inf_{\boldsymbol{y_\ell}} \sup_{\boldsymbol{\sigma_\ell}\geq 0} \boldsymbol{\psi}^{*T} B_{\boldsymbol{\ell}} \boldsymbol{y_\ell} + \boldsymbol{\sigma}_{\boldsymbol{\ell}}^T W_{\boldsymbol{\ell}} \boldsymbol{y_\ell} = \inf_{\boldsymbol{y_\ell}: W_{\boldsymbol{\ell}} \boldsymbol{y_\ell} \leq 0} \boldsymbol{\psi}^{*T} B_{\boldsymbol{\ell}} \boldsymbol{y_\ell}.$$

Since the two problems are dual of each other and $\inf_{\boldsymbol{y_\ell}: W_{\boldsymbol{\ell}} \boldsymbol{y_\ell} \leq 0} \boldsymbol{\psi}^{*T} B_{\boldsymbol{\ell}} \boldsymbol{y_\ell}$ is feasible with $\boldsymbol{y_\ell} = 0$, we can conclude by strong LP duality that the two values are the same.

We thus obtain that the left-hand side of condition (11) can be obtained by solving:

$$\underset{\boldsymbol{\ell}\in\mathcal{L}, \boldsymbol{y_\ell}}{\text{minimize}} \sup_{\boldsymbol{\sigma_\ell}\geq 0} \boldsymbol{\psi}^{*T} B_{\boldsymbol{\ell}} \boldsymbol{y_\ell} + \boldsymbol{\sigma}_{\boldsymbol{\ell}}^T W_{\boldsymbol{\ell}} \boldsymbol{y_\ell}.$$

On taking the dual of the inner maximization problem, we have

$$\begin{aligned} \underset{\boldsymbol{\ell}\in\mathcal{L}, \boldsymbol{y_\ell}}{\text{minimize}} \quad & \boldsymbol{\psi}^{*T} B_{\boldsymbol{\ell}} \boldsymbol{y_\ell} \\ \text{subject to} \quad & W_{\boldsymbol{\ell}} \boldsymbol{y_\ell} \leq 0, \end{aligned}$$

which can be written more carefully as

$$\begin{aligned} \underset{\boldsymbol{\ell}\in\mathcal{L}, \bar{\boldsymbol{\Delta}}, \bar{u}, \bar{\eta}}{\text{minimize}} \quad & \frac{\boldsymbol{\varphi}^{*T}}{1-\alpha} \bar{\boldsymbol{\Delta}} + p^* \bar{u} + \pi^* \bar{\eta} && && \text{(18a)} \\ \text{subject to} \quad & \bar{\Delta}_k \geq \bar{u} f^*_{\boldsymbol{\ell} k} - \bar{\eta} && \forall k \in \mathcal{K} && \text{(18b)} \\ & \bar{\Delta}_k \geq 0 && \forall k \in \mathcal{K} && \text{(18c)} \\ & \bar{u} \geq 0 && && \text{(18d)} \\ & \bar{\eta} \geq \bar{u} \zeta_{\text{lb}} && && \text{(18e)} \\ & \bar{\eta} \leq \bar{u} \zeta_{\text{ub}} && && \text{(18f)} \\ & \boldsymbol{\ell} \in \{0,1\}^{|E|} && && \text{(18g)} \\ & \mathbf{1}^T \boldsymbol{\ell} = 1, && && \text{(18h)} \end{aligned}$$

where $\bar{\boldsymbol{\Delta}} \in \mathbb{R}^{|\mathcal{K}|}, \bar{u} \in \mathbb{R}, \bar{\eta} \in \mathbb{R}$, and where $\boldsymbol{\varphi}^*$, p^*, and π^* are the terms of the dual vector $\boldsymbol{\varphi}^*$ associated with constraints (7b), (7e), and (7f) respectively.

Next, we can observe that when $\bar{u} = 0$, problem (18) necessarily evaluates to zero. From this we conclude that $\bar{u} > 0$ can be added to problem (18) without affecting the conclusion when used to check condition (11). Moreover, $\bar{u}$ can be pulled out of (18) after replacing $\boldsymbol{\Delta} := (1/\bar{u})\bar{\boldsymbol{\Delta}}$ and $\eta := \bar{\eta}/\bar{u}$ to obtain:

$$\begin{aligned} \underset{\boldsymbol{\ell}\in\mathcal{L}, \boldsymbol{\Delta}, \eta}{\text{minimize}} \quad & \frac{\boldsymbol{\varphi}^{*T}}{1-\alpha} \boldsymbol{\Delta} + p^* + \pi^* \eta && && \text{(19a)} \\ \text{subject to} \quad & \Delta_k \geq f^*_{\boldsymbol{\ell} k} - \eta && \forall k \in \mathcal{K} && \text{(19b)} \\ & \Delta_k \geq 0 && \forall k \in \mathcal{K} && \text{(19c)} \\ & \eta \geq \zeta_{\text{lb}} && && \text{(19d)} \end{aligned}$$

$$\eta \leq \zeta_{\text{ub}} \tag{19e}$$
$$\boldsymbol{\ell} \in \{0,1\}^{|E|} \tag{19f}$$
$$\mathbf{1}^T \boldsymbol{\ell} = 1, \tag{19g}$$

Next, in order to solve the problem (19), we need to make explicit the relation between $\boldsymbol{\ell}$ and $f_{\boldsymbol{\ell},k}$ for each scenario k. One way is to exploit the dual problem associated with problem (2) which is given by

$$\begin{aligned} f_{\boldsymbol{\ell},k} = \quad & \min_{\boldsymbol{\upsilon},\boldsymbol{\lambda}} \quad (\mathbf{1}-\boldsymbol{\ell})^T C^k \boldsymbol{\lambda} \\ \text{subject to} \quad & \boldsymbol{\lambda} + N^T \boldsymbol{\upsilon} - \boldsymbol{d} \geq 0 \\ & 0 \leq \boldsymbol{\lambda} \leq 1, \end{aligned}$$

where $\boldsymbol{\upsilon} \in \mathbb{R}^{|V|}$ and $\boldsymbol{\lambda} \in \mathbb{R}^{|E|}$ are duals associated with the constraints (2b) and (2c) respectively. The dual vector $\boldsymbol{\lambda}$ is bounded above by 1 since a unit increase in capacity of an arc can increase the flow by at most one unit, see, (Cormican et al., 1998, Lemma 1).

We therefore have reduced the evaluation of the left-hand side of condition (11) to solving the following non-linear mixed integer programming (NL-MIP) problem

$$\underset{\substack{\boldsymbol{\ell}\in\mathcal{L},\boldsymbol{\Delta},\eta \\ \{\boldsymbol{\lambda}_k,\boldsymbol{\upsilon}_k\}_{k\in\mathcal{K}}}}{\text{minimize}} \quad \frac{\boldsymbol{\varphi}^{*T}}{1-\alpha}\boldsymbol{\Delta} + p^* + \pi^*\eta \tag{20a}$$

$$\begin{aligned} \text{subject to} \quad & \Delta_k \geq \mathbf{1}^T C^k \boldsymbol{\lambda}_k - \boldsymbol{\ell}^T C^k \boldsymbol{\lambda}_k - \eta && \forall k \in \mathcal{K} && \text{(20b)} \\ & \boldsymbol{\lambda}_k + N^T \boldsymbol{\upsilon}_k - \boldsymbol{d} \geq 0 && \forall k \in \mathcal{K} && \text{(20c)} \\ & 0 \leq \boldsymbol{\lambda}_k \leq 1 && \forall k \in \mathcal{K} && \text{(20d)} \\ & \Delta_k \geq 0 && \forall k \in \mathcal{K} && \text{(20e)} \\ & \eta \geq \zeta_{\text{lb}} && && \text{(20f)} \\ & \eta \leq \zeta_{\text{ub}} && && \text{(20g)} \\ & \boldsymbol{\ell} \in \{0,1\}^{|E|} && && \text{(20h)} \\ & \mathbf{1}^T \boldsymbol{\ell} \leq B. && && \text{(20i)} \end{aligned}$$

The non-linearity in the above problem is due to the bilinear terms $\boldsymbol{\ell}^T C_k \boldsymbol{\lambda}_k = \boldsymbol{c}_k^T \operatorname{diag}(\boldsymbol{\ell})\boldsymbol{\lambda}_k$. We can linearize them to obtain an equivalent MILP since $\boldsymbol{\ell}$ is binary:

$$\underset{\substack{\boldsymbol{\ell}\in\mathcal{L},\boldsymbol{\Delta},\eta \\ \{\boldsymbol{\lambda}_k,\boldsymbol{\upsilon}_k,\boldsymbol{\Upsilon}_k\}_{k\in\mathcal{K}}}}{\text{minimize}} \quad \frac{\boldsymbol{\varphi}^{*T}}{1-\alpha}\boldsymbol{\Delta} + p^* + \pi^*\eta$$

$$\begin{aligned} \text{subject to} \quad & \Delta_k \geq \boldsymbol{c}_k^T \boldsymbol{\lambda}_k - \boldsymbol{c}_k^T \boldsymbol{\Upsilon}_k - \eta && \forall k \in \mathcal{K} \\ & \boldsymbol{\Upsilon}_k \leq \boldsymbol{\ell} && k \in \mathcal{K} \\ & \boldsymbol{\Upsilon}_k \leq \boldsymbol{\lambda}_k && \forall k \in \mathcal{K} \\ & \boldsymbol{\Upsilon}_k \geq \boldsymbol{\lambda}_k + \boldsymbol{\ell} - 1 && \forall k \in \mathcal{K} \\ & \boldsymbol{\Upsilon}_k \geq \mathbf{0} && \forall k \in \mathcal{K} \\ & \text{(20c)} - \text{(20i)}, \end{aligned}$$

where each $\boldsymbol{\Upsilon}_k \in \mathbb{R}^{|E|}$ is a linearization of $\operatorname{diag}(\boldsymbol{\ell})\boldsymbol{\lambda}_k$.

B Matrices

The coefficient matrices in (8) are given by:

$$\boldsymbol{h} = \begin{pmatrix} \mathbf{0} \\ \mathbf{0} \\ \mathbf{0} \\ \mathbf{0} \\ \mathbf{0} \\ 1 \end{pmatrix}, \quad W_{\boldsymbol{\ell}} = \begin{pmatrix} -I & \boldsymbol{f}_{\boldsymbol{\ell}} & -\mathbf{1} & \text{(7k)} \\ -I & \mathbf{0} & \mathbf{0} & \text{(7l)} \\ 0 & -1 & 0 & \text{(7m)} \\ 0 & \zeta_{\text{lb}} & -1 & \text{(7n)} \\ 0 & -\zeta_{\text{ub}} & 1 & \text{(7o)} \end{pmatrix},$$

$$A = \begin{pmatrix} \mathbf{1}\hat{\boldsymbol{q}}^T - I & \mathbf{1}\mathbf{1}^T & \mathbf{1}\mathbf{1}^T & \mathbf{1} & \Gamma\mathbf{1} & -\mathbf{1} & \text{(7b)} \\ \operatorname{diag}(\bar{\boldsymbol{q}}) & -I & 0 & 0 & -\mathbf{1} & 0 & \text{(7c)} \\ -\operatorname{diag}(\bar{\boldsymbol{q}}) & 0 & -I & 0 & -\mathbf{1} & 0 & \text{(7d)} \\ 0 & 0 & 0 & 0 & 0 & 0 & \text{(7e)} \\ 0 & 0 & 0 & 0 & 0 & 0 & \text{(7e)} \\ 0 & 0 & 0 & -1 & 0 & 0 & \text{(7f)} \\ 0 & 0 & 0 & 1 & 0 & 0 & \text{(7f)} \\ 0 & 0 & 0 & \mathbf{1} & 0 & 0 & \text{(7g)} \\ 0 & 0 & 0 & -\mathbf{1} & 0 & 0 & \text{(7h)} \\ 0 & 0 & 0 & 1 & 0 & 0 & \text{(7i)} \\ 0 & 0 & 0 & -1 & 0 & 0 & \text{(7i)} \\ 0 & -I & 0 & 0 & 0 & 0 & \text{(7j)} \\ 0 & 0 & -I & 0 & 0 & 0 & \text{(7j)} \\ 0 & 0 & 0 & 0 & -1 & 0 & \text{(7j)} \end{pmatrix}, \quad B_{\boldsymbol{\ell}} = \begin{pmatrix} \frac{1}{1-\alpha} I & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & -1 \\ 0 & \zeta_{\text{ub}}\boldsymbol{e}_{\boldsymbol{\ell}} & -\boldsymbol{e}_{\boldsymbol{\ell}} \\ 0 & -\zeta_{\text{lb}}\boldsymbol{e}_{\boldsymbol{\ell}} & \boldsymbol{e}_{\boldsymbol{\ell}} \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \quad \boldsymbol{s} = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 1 \\ -1 \\ 0 \\ 0 \\ \zeta_{\text{ub}}\mathbf{1} \\ -\zeta_{\text{lb}}\mathbf{1} \\ \zeta_{\text{ub}} \\ -\zeta_{\text{lb}} \\ 0 \\ 0 \\ 0 \end{pmatrix}.$$

where for each $\boldsymbol{\ell} \in \mathcal{L}$, $\boldsymbol{f}_{\boldsymbol{\ell}} \in \mathbb{R}^{|\mathcal{K}|}$ is the vector of maximum flows for each scenario in $|\mathcal{K}|$.

C Solving for deterministic strategy

Since ζ is bounded above by $\bar{\zeta}$, we can linearize the relation $\eta_{\boldsymbol{\ell}} = u_{\boldsymbol{\ell}}\zeta$ in (5) to obtain the following MILP:

$$\begin{aligned}
\underset{\substack{\boldsymbol{u}, t, \boldsymbol{\eta}, \boldsymbol{w}, \boldsymbol{w}^- \\ \chi, \boldsymbol{\beta}, \boldsymbol{\Delta}, \zeta}}{\text{minimize}} \quad & t \\
\text{subject to} \quad & \zeta + \sum_{k' \in \mathcal{K}} w_{k'} + \sum_{k' \in \mathcal{K}} w^-_{k'} + \Gamma\chi \\
& + \sum_{k' \in \mathcal{K}} \hat{q}_{k'}\beta_{k'} + \frac{1}{1-\alpha} \sum_{\boldsymbol{\ell} \in \mathcal{L}} \Delta_{\boldsymbol{\ell},k} - \beta_k \le t && \forall k \in \mathcal{K} \\
& \chi \ge \bar{q}_k\beta_k - w_k && \forall k \in \mathcal{K} \\
& \chi \ge -\bar{q}_k\beta_k - w^-_k && \forall k \in \mathcal{K} \\
& \boldsymbol{w} \ge 0, \boldsymbol{w}^- \ge 0, \chi \ge 0 \\
& \Delta_{\boldsymbol{\ell},k} \ge u_{\boldsymbol{\ell}} f_{\boldsymbol{\ell},k} - \eta_{\boldsymbol{\ell}} && \forall \boldsymbol{\ell} \in \mathcal{L}, k \in \mathcal{K} \\
& \Delta_{\boldsymbol{\ell},k} \ge 0 && \forall \boldsymbol{\ell} \in \mathcal{L}, k \in \mathcal{K} \\
& \eta_{\boldsymbol{\ell}} \ge 0 && \forall \boldsymbol{\ell} \in \mathcal{L} \\
& \eta_{\boldsymbol{\ell}} \ge \zeta - (1 - u_{\boldsymbol{\ell}})\bar{\zeta} && \forall \boldsymbol{\ell} \in \mathcal{L} \\
& \eta_{\boldsymbol{\ell}} \le \zeta && \forall \boldsymbol{\ell} \in \mathcal{L} \\
& \eta_{\boldsymbol{\ell}} \le \bar{\zeta} u_{\boldsymbol{\ell}} && \forall \boldsymbol{\ell} \in \mathcal{L} \\
& \boldsymbol{u} = \{0,1\}^{|\mathcal{L}|} \\
& \mathbf{1}^T\boldsymbol{u} = 1.
\end{aligned}$$

References

Aghassi, M. and Bertsimas, D. (2006). Robust game theory. Mathematical Programming, 107:231–273.

Ahuja, R. K., Magnanti, T. L., and Orlin, J. B. (1993). Network Flows: Theory, Algorithms, and Applications. Prentice-Hall, Inc., NJ, USA.

Al-Khayyal, F. A. and Falk, J. E. (1983). Jointly constrained biconvex programming. Mathematics of Operations Research, 8(2):273–286.

Assimakopoulos, N. (1987). A network interdiction model for hospital infection control. Computers in Biology and Medicine, 17(6):413–422.

Atamturk, A., Deck, C., and Jeon, H. (2017). Successive quadratic upper-bounding for discrete mean-risk minimization and network interdiction.

Atamtürk, A. and Zhang, M. (2007). Two-stage robust network flow and design under demand uncertainty. Operations Research, 55(4):662–673.

Ben-Tal, A., den Hertog, D., and Vial, J.-P. (2015). Deriving robust counterparts of nonlinear uncertain inequalities. Mathematical Programming, 149(1):265–299.

Ben-Tal, A. and Nemirovski, A. (2000). Robust solutions of linear programming problems contaminated with uncertain data. Mathematical Programming, 88(3):411–424.

Ben-Tal, A. and Nemirovski, A. (2008). Selected topics in robust convex optimization. Mathematical Programming, 112(1):125–158.

Bertsimas, D., Nasrabadi, E., and Orlin, J. B. (2016). On the power of randomization in network interdiction. Operations Research Letters, 44(1):114–120.

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

Bertsimas, D. and Sim, M. (2004). The price of robustness. Operations Research, 52(1):35–53.

Chandraker, M. and Kriegman, D. (2008). Globally optimal bilinear programming for computer vision applications. In 2008 IEEE Conference on Computer Vision and Pattern Recognition, pages 1–8. IEEE Computer Society, Anchorage, AK.

Cormican, K. J., Morton, D. P., and Wood, R. K. (1998). Stochastic network interdiction. Operations Research, 46(2):184–197.

Delage, E., Kuhn, D., and Wiesemann, W. (2019). "dice"-sion–making under uncertainty: When can a random decision reduce risk? Management Science, 65(7):3282–3301.

Delage, E. and Saif, A. (2018). The value of randomized solutions in mixed-integer distributionally robust optimization problems. Les Cahiers du GERAD G–2018–45.

Desrosiers, J. and Lübbecke, M. E. (2005). A primer in column generation. In Desaulniers, G., Desrosiers, J., and Solomon, M. M., editors, *Column Generation*, pages 1–32. Springer US, Boston, MA.

Gass, S. I. (1991). Military manpower planning models. Computers & Operations Research, 18(1):65–73.

Hajinezhad, D. and Hong, M. (2019). Perturbed proximal primal–dual algorithm for nonconvex nonsmooth optimization. Mathematical Programming, 176(1):207–245.

Hajinezhad, D. and Shi, Q. (2018). Alternating direction method of multipliers for a class of nonconvex bilinear optimization: convergence analysis and applications. Journal of Global Optimization, 70(1):261–288.

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

Jain, M., Tsai, J., Pita, J., Kiekintveld, C., Rathi, S., Tambe, M., and Ordóñez, F. (2010). Software assistants for randomized patrol planning for the lax airport police and the federal air marshal service. INFORMS Journal on Applied Analytics, 40(4):267–290.

Janjarassuk, U. and Linderoth, J. (2008). Reformulation and sampling to solve a stochastic network interdiction problem. Networks, 52(3):120–132.

Kar, D., Nguyen, T. H., Fang, F., Brown, M., Sinha, A., Tambe, M., and Jiang, A. X. (2017). Trends and applications in stackelberg security games. In Basar, T. and Zaccour, G., editors, Handbook of Dynamic Game Theory, pages 1–47. Springer International Publishing, Cham.

Lei, X., Shen, S., and Song, Y. (2018). Stochastic maximum flow interdiction problems under heterogeneous risk preferences. Computers & Operations Research, 90:97–109.

Liberti, L. and Pantelides, C. C. (2006). An exact reformulation algorithm for large nonconvex nlps involving bilinear terms. Journal of Global Optimization, 36(2):161–189.

Loizou, N. (2015). Distributionally robust game theory.

Magliocca, N. R., McSweeney, K., Sesnie, S. E., Tellman, E., Devine, J. A., Nielsen, E. A., Pearson, Z., and Wrathall, D. J. (2019). Modeling cocaine traffickers and counterdrug interdiction forces as a complex adaptive system. Proceedings of the National Academy of Sciences, 116(16):7784–7792.

McCormick, G. P. (1976). Computability of global solutions to factorable nonconvex programs: Part I–Convex underestimating problems. Mathematical Programming, 10(1):147–175.

Pita, J., Jain, M., Marecki, J., Ordóñez, F., Portway, C., Tambe, M., Western, C., Paruchuri, P., and Kraus, S. (2008). Deployed ARMOR Protection: The Application of a Game Theoretic Model for Security at the Los Angeles International Airport. In Proc. Seventh Internat. Joint Conf. Autonomous Agents Multiagent Systems, pages 125–132, Richland, SC. International Foundation for Autonomous Agents and Multiagent Systems.

Pita, J., Jain, M., Ordóñez, F., Portway, C., Tambe, M., Western, C., Paruchuri, P., and Kraus, S. (2009). Using Game Theory for Los Angeles Airport Security. AI Magazine, 30(1):43–57.

Rockafellar, R. and Uryasev, S. (2000). Optimization of conditional value-at-risk. Journal of Risk, 2:21–41.

Sherali, H. D. and Alameddine, A. (1992). A new reformulation-linearization technique for bilinear programming problems. Journal of Global Optimization, 2(4):379–410.

Sion, M. (1958). On general minimimax theorems. Pacific Journal of Mathematics, 8(1):171–176.

Smith, J. C., Prince, M., and Geunes, J. (2013). Modern network interdiction problems and algorithms. In Pardalos, P. M., Du, D.-Z., and Graham, R. L., editors, Handbook of Combinatorial Optimization, pages 1949–1987. Springer New York, New York, NY.

Smith, J. C. and Song, Y. (2020). A survey of network interdiction models and algorithms. European Journal of Operational Research, 283(3):797–811.

Smith, J. E. and Winkler, R. L. (2006). The optimizer's curse: Skepticism and postdecision surprise in decision analysis. Management Science, 52(3):311–322.

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