Using Graph Convolutional Networks to Learn Interdiction Games

Kai Wang, **Aditya Mate**, **Bryan Wilder**, **Andrew Perrault** and **Milind Tambe**
University of Southern California
{wang319, aditya.mate, bwilder, aperraul, tambe}@usc.edu

Abstract

Illegal smuggling is one of the most important issues across countries, where more than $10 billion a year of illegal wildlife trafficking is conducted within transnational criminal networks. Governments have tried to deploy inspections at checkpoints to stop illegal smuggling, though the effect is quite limited due to vast protection areas but limited human resources. We study these problems from the perspective of network interdiction games with a boundedly rational attacker. In this paper, we aim to improve the efficiency of the limited number of checkpoints. The problem involves two main stages: i) a predictive stage to predict the attacker's behavior based on the historical interdiction; ii) a prescriptive stage to optimally allocate limited checkpoints to interdict the attacker. In this paper, we propose a novel boundedly rational model which resolves the issue of exponentially many attacker strategies by making memoryless assumption about the attacker's behavior. We show that the attacker's behavior can be reduced to an absorbing Markov chain, where the success probability of reaching any target can be computed analytically, thus optimized via any gradient-based optimization technique. We incorporate graph convolutional neural networks with K-hops look-ahead to model the attacker's reasoning. Our proposed model provides a new perspective to study the boundedly rationality in traditional interdiction games with graph structure. This novel model possesses nice analytical properties and scales up very well by avoiding enumerating all paths in the graph.

1 Introduction

The illegal wildlife trade is estimated to be the second-largest illegal trade worldwide [Warchol *et al.*, 2003]. Wildlife trafficking, the illegal poaching, transit, trade and sale of wildlife, generates more than $10 billion a year for transnational organized criminal networks[1]. Thousands of elephants die each year so that their tusks can be carved into religious objects [Christy, 2012]. Although checkpoints with inspections can efficiently stop smuggling, due to the limited human resource, we can hardly protect the entire vast area especially when the smuggler is also intelligent. In the previous literature, Stackelberg security games have been commonly used in the security related problems, including wildlife conservation [Fang *et al.*, 2015] and flight protection [Pita *et al.*, 2009; Wang *et al.*, 2018]. In this paper, we are interested in Stackelberg security games whose domains come with a graph structure, e.g., smuggling network, which are generally known as interdiction games [Washburn and Wood, 1995]. The smuggler, as an evading attacker, aims to find a path from one of the origins to one of the destinations; the interdicting player, as a defender, inspects one or more edges in the graph in order to prevent the attacker from reaching the target.

Previous studies about interdiction games mostly focus on the attacker's rationalily. When the attacker is perfectly rational, double oracles [Jain *et al.*, 2011; Jain *et al.*, 2013] and branch-and-cut [Fischetti *et al.*, 2019] algorithms were proposed to find the optimal defender strategy. When the attacker is boundedly rational, many boundedly rational models, including Quantal Response (QR) [McKelvey and Palfrey, 1995] and Subjective Utility Quantal Response (SUQR) [Nguyen *et al.*, 2013], were proposed to fit the attacker's rationality. These boundedly rational models can learn the attacker's behavior from the historical plays. Here, we mainly focus on the attacker's rationality in problems with graph structure, which is less studied in the previous literature.

The interdiction problem with boundedly rational attacker can be generally decomposed into two stages: predictive and prescriptive stages. The predictive stage is to predict the path that the attacker is most likely to be using, while the prescriptive stage is to optimally allocate the resource based on the previous prediction. However, the standard QR and SUQR models are not designed for graph-based problem and decision. People tried to reduce a graph-based interdiction game to a standard Stackelberg security game via enumerating all the possible attacker strategies [Yang *et al.*, 2011; Guo *et al.*, 2016; Ford *et al.*, 2015], where each potential path reduces to a new target that the attacker might choose. This can resolve the issue of graph-related decision but create another problem of exponentially or even infinitely many paths from any source to any destination. The huge amount of paths may also result in a sparse prediction in the predictive stage.

[1] According to https://www.state.gov/e/oes/ecw/wlt/

In the prescriptive stage, there is also scalability issue, where the defender can hardly solve the optimization problem due to exponential size of constraints and variables. In the previous literature, they usually require the graph to be directed acyclic graph (DAG) to reduce the number of paths, which largely restricts their applicability.

In this paper, we present a novel graph-based boundedly rational model by introducing an additional assumption on the attacker's memoryless property. We show that after adding the memoryless property, the problem reduces to an absorbing Markov chain where the success probability of reaching any destination from any source can be computed analytically. With the help of this novel model, we only need to learn the transition probability of each pair of neighbors and can avoid the issue of exponentially many paths. We provide analysis of the training process for both predictive and prescriptive stages. The predictive model also incorporates a graph convolutional network [Kipf and Welling, 2016] and K-hops lookahead method to remedy the myopia of the memoryless assumption, where these tricks allow the attacker to make decision based on non-local information. We demonstrate the applicability of our model analytically, where both the predictive and prescriptive stages are theoretically more scalable than the previous literature.

2 Preliminary and Model

2.1 Stackelberg Security Games (SSGs)

A Stackelberg security game is a two-player game composed of one defender and one attacker. The defender aims to protect a set of targets T with limited budget b where the defender can only fully protect up to b targets. Each target $t \in T$ is associated with a defender payoff $U^d(t) \leq 0$ and an attacker payoff $U^a(t) \geq 0$ when the target is successfully attacked. Once the defender commits her strategy, the attacker can conduct surveillance and choose one target to attack. We denote the defender's mixed strategy by $\mathbf{p} \in \mathbb{R}^{|T|}$, where $\mathbf{p}_t$ refers to the marginal probability that target t is protected. The probability that the attacker chooses target t is denoted by $\mathbf{q}_t(\mathbf{p}, \xi)$ (or abbreviated as $\mathbf{q}_t$), which is affected by the defender's strategy $\mathbf{p}$ and the feature ξ, e.g., the payoff value $U^a(t)$ of target t. We denote the payoff of catching the attacker (being caught) to be a constant $U^{d/a}_{\text{caught}}$ for defender (attacker). For simplicity and without loss of generality, we assume zero catching reward to the defender $U^d_{\text{caught}} = 0$. Then the defender's utility function can be written as:

$$\text{DefU}(\mathbf{p}; \mathbf{q}) = \sum\nolimits_{t \in T} (1 - \mathbf{p}_t)\mathbf{q}_t(\mathbf{p}, \xi)U^d(t) \qquad (1)$$

The function $\mathbf{q}$ can represent the attacker's rationality, e.g., $\mathbf{q}_t(\mathbf{p}, \xi) = 1$ if $t = \arg\max_{t' \in T}(1 - \mathbf{p}_{t'})U^a(t')$ else 0 refers to a fully rational attacker.

2.2 Bounded Rationality in SSGs

In the previous studies on SSGs, there is a rich literature about modeling adversary modeling. QR [McKelvey and Palfrey, 1995] was proposed to model the attacker's boundedly rational behavior, where a QR attacker attacks each target with probability proportional to the exponential of its payoff scaled by a constant λ: $\mathbf{q}_t(\mathbf{p}) \propto \exp(\lambda(1 - \mathbf{p}_t)U_a(t))$. SUQR [Nguyen *et al.*, 2013] is a variant of the QR model, where the probability is proportional to the exponential of a subjective utility or an attractiveness function:

$$\mathbf{q}_t(\mathbf{p}, \xi) \propto \exp(-w\mathbf{p}_t + \Phi(t, \xi)) \qquad (2)$$

where $w > 0$ is a constant representing the attacker's risk aversion and $\Phi(t, \xi)$ denotes the subjective utility of target t under feature ξ.

2.3 Interdiction Games

Interdiction games are an extension of the standard Stackelberg game where the attack consists of a path on a graph. Given a directed graph $G = (\mathcal{V}, \mathcal{E})$, the defender is trying to allocate limited number of checkpoints along edges in $\mathcal{E}$ while the attacker is trying to find a path from a source to a target without being caught. The defender's pure strategy is a set of edges to assign checkpoints while the attacker's pure strategy is a path. We divide the set of all vertices into $\mathcal{V} = S \cup T$, where $S = \{s_1, s_2, ..., s_{|S|}\}$ is the set of all possible sources and $T = \{t_1, t_2, ..., t_{|T|}\}$ is the set of all targets. Without loss of generality, we assume $S \cap T = \emptyset$. At each time, the attacker appears in one source $s \in S$ drawn from a given prior distribution $\pi \in \mathbb{R}^{|S|}$.

We use $\alpha = \{v_1, v_2, ..., v_{|\alpha|}\}$ to denote a path with one starting vertex $v_1 \in S$ and a target vertex $v_{|\alpha|} \in T$. Each target $t \in T$ corresponds to payoffs $U^{d/a}(t)$ to the defender and the attacker. When the attacker gets caught by the defender, the attacker receives a penalty U^a_{caught} and the defender receives a reward U^d_{caught}. For simplicity and without loss of generality, we assume zero catching reward $U^d_{\text{caught}} = 0$. Thus we can write the attacker payoffs by $U^a = \{U^a(t_1), ..., U^a(t_{|T|}), U^a_{\text{caught}}\} \in \mathbb{R}^{|T|+1}$ and the defender payoffs by $U^d = \{U^d(t_1), ..., U^d(t_{|T|}), U^d_{\text{caught}}\} \in \mathbb{R}^{|T|+1}$.

With respect to the graph feature, each node $v \in \mathcal{V}$ comes with a node feature $\xi_v \in \mathbb{R}^D$. The features here could be the different characteristics of node v, e.g., the shortest distance to any landmark or the payoff of the current node $U^a(v)$ if $v \in T$. The entire graph feature is denoted by $\xi \in \mathbb{R}^{|\mathcal{V}| \times D}$.

We denote the set of all possible paths from any source to any target by $\mathcal{A}$, where this set could be infinitely many when the graph contains any cycle. Similarly, $\mathbf{p}$ denotes the defender strategy with $\mathbf{p}_e$ the marginal probability of covering edge $e \in \mathcal{E}$. The defender has limited number of resources b, i.e., the defender's mixed strategy needs to satisfy $\mathbf{1}^\top \mathbf{p} \leq b$.

2.4 Bounded Rationality in Interdiction Games

In this paper, we consider the attacker to be boundedly rational, where the attacker's mixed strategy is given by a function $\mathbf{q}(\mathbf{p}, \xi)$ representing the probability of choosing path α under coverage p) and features ξ. Given the coverage p, we can compute the defender expected utility:

$$\text{DefU}(\mathbf{p}; \mathbf{q}) = \sum_{\alpha \in \mathcal{A}} \prod_{e=(v_i, v_{i+1})} (1 - \mathbf{p}_e)\mathbf{q}_\alpha(\mathbf{p}, \xi)U^d(\alpha) \qquad (3)$$

where $U^d(\alpha) = U^d(t)$ is the defender utility when attacker successfully pass through path α to attack its target t.

The only difference between Equation (1) and (3) is that there are multiple layers of protection along the path α. Therefore the probability of successfully attacking a target is a product of all the success probability of crossing edge e in the path. However, this incurs the non-linearity of coverage p, which makes the above defender optimization problem generally a hard problem regardless the attacker's rationality. Furthermore, the set of all possible paths $\mathcal{A}$ could be exponentially large or infinitely many when there is any cycle. This results in a computational challenge to any boundedly rational assumption and also the optimization problem itself. Even if the function $\mathbf{q}_\alpha(\mathbf{p}, \xi), \alpha \in \mathcal{A}$ is given and there are only polynomial many paths $\mathcal{A}$, the defender optimization problem is still proven to be NP-hard [Jain *et al.*, 2011]. In the following sections, we will show that our proposed model can resolve these issues by imposing the memoryless property to the attacker behavior.

3 Problem Statement

At each instance, a directed graph $G = (\mathcal{V}, \mathcal{E})$ with node features ξ is presented to both the defender and the attacker. The attacker comes with a hidden underlying rationality function $\mathbf{q}$. The defender has access to the historical plays between the defender and the attacker, where each training instance contains a graph G' with node features ξ', the deployed coverage $\mathbf{p}'$, and the path α' the attacker chose to use.

Once the defender chooses a coverage $\{\mathbf{p}_e\}_{e\in\mathcal{E}}$ satisfying the budget constraint $\mathbf{1}^\top \mathbf{p} \leq b$, the attacker observes $\mathbf{p}$ and decides which path to use based on his own rationality function $\mathbf{q}$. The defender aims to learn the attacker rationality function from the training instances and thus chooses the optimal coverage to maximize the corresponding defender utility. In the standard machine learning literature, this problem is usually solved by a two-stage method, which is composed of a predictive stage and a prescriptive stage.

3.1 Two-stage Approach

Predictive stage is aiming to predict the attacker's action by approximating the attacker's rationality function $\mathbf{q}$. This can be thought of as a classification model, where given an input graph, node features, and edge coverage, the attacker will choose a path based on an underlying distribution, which is a function of all the input features. The goal here is to minimize the cross-entropy between our predicted distribution and the underlying distribution.

Prescriptive stage is trying to maximize the defender utility based on the learned attacker's rationality function $\tilde{\mathbf{q}}$. Given the learned function $\tilde{\mathbf{q}}$, we can express the defender optimization problem as Equation (5). The objective function is a non-convex function of coverage $\mathbf{p}$, where $\mathbf{p}$ needs to satisfy some budget constraints. This can be solved by standard constrained optimization techniques.

4 Predictive Model

4.1 Memoryless Attacker

In this paper, we assume the boundedly rational attacker to be memoryless. Given an attacker rationality function $\mathbf{q}$, defender coverage $\mathbf{p}$, and features ξ, the probability of using a path $\alpha = \{v_1, v_2, ..., v_{|\alpha|}\}$ can be written as:

$$\mathbf{q}_\alpha(\mathbf{p}, \xi) = \prod_{i=1}^{|\alpha|} \text{Prob}(v_i \to v_{i+1} | v_1, ..., v_i, \mathbf{p}, \xi) \quad (4)$$
$$= \prod_{i=1}^{|\alpha|} \text{Prob}(v_i \to v_{i+1} | \mathbf{p}, \xi) = \prod_{i=1}^{|\alpha|} \mathbf{q}_{v_i \to v_{i+1}}(\mathbf{p}, \xi)$$

We use $\mathbf{q}_{u\to v}(\mathbf{p}, \xi)$ to represent the probability of transiting from node u to v under coverage $\mathbf{p}$ and features ξ. Now the above Equation (3) can be further simplified as:

$$\text{DefU}(\mathbf{p}; \mathbf{q}) = \sum_{\alpha\in\mathcal{A}} \prod_{e=(v_i,v_{i+1})} (1 - \mathbf{p}_e)\mathbf{q}_e(\mathbf{p}, \xi)U^d(\alpha) \quad (5)$$

Under the memoryless assumption, we just need to learn the transition probability $\mathbf{q}_e(\mathbf{p}, \xi)$ for every edge $e \in \mathcal{E}$.

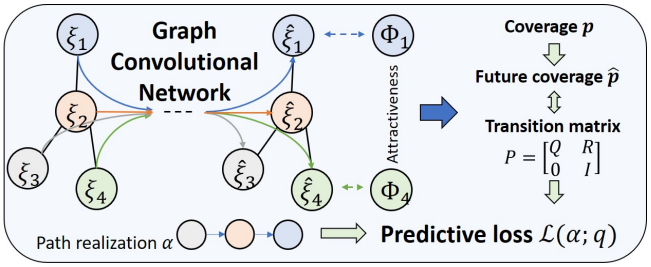

Figure 1: Predictive model with GCN and future coverage

4.2 Localized SUQR

Motivated by the SUQR model, we assume the probability that the attacker moves from u to v is proportional to $exp(-\omega\hat{\mathbf{p}}_{u\to v} + \Phi(v, \xi)) \; \forall v \in N_{\text{out}}(u)$, where $\hat{\mathbf{p}}_{u\to v}$ represents the amount of future coverage that the attacker will see after choosing edge $u \to v$, and $\Phi(v, \xi)$ represents the attractiveness of node v. In other words, the attacker tends to move toward the target with larger attractiveness $\Phi(v, \xi)$ but avoids using the edge $e = (u, v) \in \mathcal{E}$ been covered or with more coverage in the future. So the transition probability from u to any $v \in N_{\text{out}}(u)$ can be written as:

$$\mathbf{q}_{u\to v}(\mathbf{p}, \xi) = \frac{exp(-\omega\hat{\mathbf{p}}_{u\to v} + \Phi(v, \xi))}{\sum_{v'\in N_{\text{out}}(u)} exp(-\omega\hat{\mathbf{p}}_{u\to v'} + \Phi(v', \xi)} \quad (6)$$

which is a graph-based extension of the SUQR model.

4.3 Attractiveness with GCNs

In order to incorporate the graph structure with node features, we apply graph convolutional networks (GCNs) introduced by [Kipf and Welling, 2016]. We use a graph convolution network with input graph G and node features ξ. The target is the attractiveness of each node. The graph convolution network outputs an embedded feature for each node, which will be fed into another fully connected neural network to obtain the attractiveness. The convolution layer can capture the nearby features, which allows us to capture some non-local attacker decisions. The learned attractiveness value will be used to compute the transition probability in Equation (6), which is differentiable thus back-propagatable.

The GCNs described in [Kipf and Welling, 2016] do not rely on the graph structure. So we can train a single GCN with the smuggling data of different countries and apply the trained model to other new countries.

4.4 Future Coverage and K-hops Lookahead

Previous literature has shown evidence that humans make decision based on future risk even if they are boundedly rational [Wittmann and Paulus, 2008], though with a bounded awareness that only allows humans to track up to limited time steps [Chugh and Bazerman, 2007]. Therefore, we take the future coverage $\hat{\mathbf{p}}_{u\to v}$ conditional on choosing edge $u \to v$ into account, which represents the attacker's perceived future risk of this decision. If both the attractiveness and the future coverage value are given, we can compute the transition matrix A_{line} of the line graph, where $[A_{\text{line}}]_{(w,u),(u,v)} = \text{Prob}((w,u) \to (u,v)) = \text{Prob}(u \to v)$ denotes the transition probability between adjacent edges. Then the future coverage conditional on choosing edge $u \to v$ can be written as:

$$\hat{\mathbf{p}}_{u\to v} = \sum_{k=0}^{K} \gamma^k \sum_{e\in\mathcal{E}} \text{Prob}(\text{reach } e \text{ in } k \text{ hops}) \cdot \mathbf{p}_e$$
$$= \sum_{k=0}^{K} \gamma^k (\mathbf{1}_{u\to v}^\top A_{\text{line}}^k)\mathbf{p} = \mathbf{1}_{u\to v}^\top (\sum_{k=0}^{K} \gamma^k A_{\text{line}}^k)\mathbf{p} \quad (7)$$

where $0 \le \gamma \le 1$ is a future discount factor and K is the maximum number of hops lookahead, representing the limited awareness of the attacker.

Since the transition matrix A_{line} is also a function of future coverage $\hat{\mathbf{p}}$, we can close the loop by assuming the future coverage to be a fixed point of a given function This can be solved by fixed-point theorem and all the derivatives can be obtained by implicit function theorem.

4.5 Training Data and Loss Function

In the domain of smuggling networks, we do not have the data of neither the actual attractiveness nor future coverage. Instead, we have access to the realization of the attacker's chosen path. Therefore, in our training and testing data, each instance $(G, \xi, \mathbf{p}, \alpha)$ contains a graph G with node features ξ, edge coverage $\mathbf{p}$, and the path that the attacker chose $\alpha \in \mathcal{A}$. We use the cross-entropy between the actual ground truth path α and our predicted distribution generated from $\mathbf{q}$ as loss function, which can be written as:

$$\mathcal{L}(D_{\text{train}}; \mathbf{q}) = \sum_{(G,\xi,\mathbf{p},\alpha)\in D_{\text{train}}} -\log(\text{Prob}(\alpha \text{ was taken}))$$
$$= -\sum_{(G,\xi,\mathbf{p},\alpha)\in D_{\text{train}}} \sum_{e\in\alpha} \log \mathbf{q}_e(\mathbf{p},\xi) \quad (8)$$

where the $\mathbf{q}$ function is described in Equation (6) and the attractiveness and future coverage are as discussed in the previous sections. We can train the hyperparameters of GCN by any standard gradient descent method.

5 Prescriptive Model

Given a graph G with node features ξ, the defender's goal is to choose an optimal coverage $\mathbf{p}^*$ to maximize her own objective value under the budget constraint.

5.1 Absorbing Markov Chain

Given a chosen coverage $\mathbf{p}$ and the attacker's current location u, then the probability that the attacker gets caught conditional on choosing edge $e = (u, v)$ is given by $\mathbf{p}_e$. Due to Equation (6), we know the attacker will choose edge e with probability $q_{u\to v}(\mathbf{p}, \xi)$. Therefore the probability that the attacker can actually reach node v in the next step is $q_u(v, \xi)(1 - p_e)$. We add a dummy node v_{caught} to represent the state of being caught. Therefore we can represent the transition probability from node u to v_{caught} by summing up the probability of getting caught. Once the attacker reaches either any terminal (e.g., black market) or v_{caught}, the attacker cannot go back to any other states. Therefore, we can model the attacker's behavior as an absorbing Markov chain with all these terminals and the dummy node v_{caught} to be absorbing states. Then we just need to compute the probability that the attacker ends up being in each absorbing state.

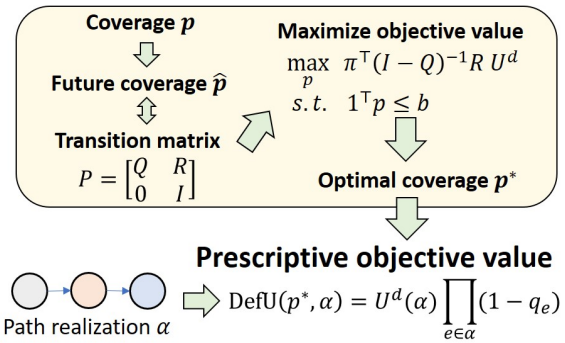

Figure 2: Prescriptive model with constrained optimization problem

Given coverage p, if we sort the vertices out by intermediate states then absorbing states, then the transition matrix can be written as: $P = \begin{bmatrix} Q & R \\ 0 & I \end{bmatrix}$, where I is an identity matrix representing once the attacker reaches the absorbing state, he would never transit to other states. The absorbing probability can be easily written as $B = (I - Q)^{-1} R \in \mathbb{R}^{|S|\times(|T|+1)}$, where the entry B_{ij} indicates the probability that the attacker initiates from state i and ends up being in absorbing state j.

Notice that the transition matrix P and its sub-components Q, R are all functions of coverage p. Given the initial source distribution $\pi \in \mathbb{R}^{|S|}$ and the defender payoff of absorbing states $R^d \in \mathbb{R}^{|T|+1}$, the defender's expected payoff $\pi^\top B U^d$ and optimization problem can be written as:

$$\max_p \quad \pi^\top B(\mathbf{p}) U^d \quad (9)$$
$$\text{s.t.} \quad \mathbf{1}^\top \mathbf{p} \le b, \qquad \mathbf{p}_e \ge 0 \quad \forall e \in \mathcal{E} \quad (10)$$

This optimization problem can be solved by any optimization technique, e.g., Sequential Least SQuares Programming (SLSQP) [Kraft, 1985; Bertsekas and Tsitsiklis, 1996].

6 Conclusions

We propose a novel model of boundedly rationality which resolves the problem of enumerating all possible paths by assuming memoryless property. The attacker's behavior can be reduced to an absorbing Markov chain with learned transition probability. The defender utility under this assumption can also be analytically computed, which enables any gradient-based optimization technique to be used here. These conclude our two-stage model which can be used to learn and deal with a boundedly rational attacker.

References

[Bertsekas and Tsitsiklis, 1996] Dimitri P Bertsekas and John N Tsitsiklis. *Neuro-dynamic programming*, volume 5. Athena Scientific Belmont, MA, 1996.

[Christy, 2012] Bryan Christy. Ivory worship. *National Geographic*, 222(4):28–61, 2012.

[Chugh and Bazerman, 2007] Dolly Chugh and Max H Bazerman. Bounded awareness: What you fail to see can hurt you. *Mind & Society*, 6(1):1–18, 2007.

[Fang *et al.*, 2015] Fei Fang, Peter Stone, and Milind Tambe. When security games go green: Designing defender strategies to prevent poaching and illegal fishing. In *Twenty-Fourth International Joint Conference on Artificial Intelligence*, 2015.

[Fischetti *et al.*, 2019] Matteo Fischetti, Ivana Ljubić, Michele Monaci, and Markus Sinnl. Interdiction games and monotonicity, with application to knapsack problems. *INFORMS Journal on Computing*, 2019.

[Ford *et al.*, 2015] Benjamin Ford, Thanh Nguyen, Milind Tambe, Nicole Sintov, and Francesco Delle Fave. Beware the soothsayer: From attack prediction accuracy to predictive reliability in security games. In *International Conference on Decision and Game Theory for Security*, pages 35–56. Springer, 2015.

[Guo *et al.*, 2016] Qingyu Guo, Bo An, Yair Zick, and Chunyan Miao. Optimal interdiction of illegal network flow. 2016.

[Jain *et al.*, 2011] Manish Jain, Dmytro Korzhyk, Ondřej Vaněk, Vincent Conitzer, Michal Pěchouček, and Milind Tambe. A double oracle algorithm for zero-sum security games on graphs. In *The 10th International Conference on Autonomous Agents and Multiagent Systems-Volume 1*, pages 327–334. International Foundation for Autonomous Agents and Multiagent Systems, 2011.

[Jain *et al.*, 2013] Manish Jain, Vincent Conitzer, and Milind Tambe. Security scheduling for real-world networks. In *Proceedings of the 2013 international conference on Autonomous agents and multi-agent systems*, pages 215–222. International Foundation for Autonomous Agents and Multiagent Systems, 2013.

[Kipf and Welling, 2016] Thomas N Kipf and Max Welling. Semi-supervised classification with graph convolutional networks. *arXiv preprint arXiv:1609.02907*, 2016.

[Kraft, 1985] Dieter Kraft. On converting optimal control problems into nonlinear programming problems. In *Computational mathematical programming*, pages 261–280. Springer, 1985.

[McKelvey and Palfrey, 1995] Richard D McKelvey and Thomas R Palfrey. Quantal response equilibria for normal form games. *Games and economic behavior*, 10(1):6–38, 1995.

[Nguyen *et al.*, 2013] Thanh Hong Nguyen, Rong Yang, Amos Azaria, Sarit Kraus, and Milind Tambe. Analyzing the effectiveness of adversary modeling in security games. In *Twenty-Seventh AAAI Conference on Artificial Intelligence*, 2013.

[Pita *et al.*, 2009] James Pita, Manish Jain, Fernando Ordónez, Christopher Portway, Milind Tambe, Craig Western, Praveen Paruchuri, and Sarit Kraus. Using game theory for los angeles airport security. *AI magazine*, 30(1):43–43, 2009.

[Wang *et al.*, 2018] Kai Wang, Qingyu Guo, Phebe Vayanos, Milind Tambe, and Bo An. Equilibrium refinement in security games with arbitrary scheduling constraints. In *Proceedings of the 17th International Conference on Autonomous Agents and MultiAgent Systems*, pages 919–927. International Foundation for Autonomous Agents and Multiagent Systems, 2018.

[Warchol *et al.*, 2003] Greg L Warchol, Linda L Zupan, and Willie Clack. Transnational criminality: An analysis of the illegal wildlife market in southern africa. *International Criminal Justice Review*, 13(1):1–27, 2003.

[Washburn and Wood, 1995] Alan Washburn and Kevin Wood. Two-person zero-sum games for network interdiction. *Operations research*, 43(2):243–251, 1995.

[Wittmann and Paulus, 2008] Marc Wittmann and Martin P Paulus. Decision making, impulsivity and time perception. *Trends in cognitive sciences*, 12(1):7–12, 2008.

[Yang *et al.*, 2011] Rong Yang, Christopher Kiekintveld, Fernando Ordonez, Milind Tambe, and Richard John. Improving resource allocation strategy against human adversaries in security games. In *Twenty-Second International Joint Conference on Artificial Intelligence*, 2011.