Proceedings of the 2025 Winter Simulation Conference
E. Azar, A. Djanatliev, A. Harper, C. Kogler, V. Ramamohan, A. Anagnostou, and S. J. E. Taylor, eds.

SIMULATION-BASED MULTI-AGENT REINFORCEMENT LEARNING FOR NETWORK INTERDICTION GAMES

Xudong Wang[1]

[1]Dept. of Industrial and Systems Eng., University of Tennessee - Knoxville, Knoxville, TN, USA

ABSTRACT

Network interdiction problems capture adversarial interactions between a defender seeking to preserve flow in a network and an attacker aiming to disrupt it. Traditional approaches model this as a bilevel optimization problem, which quickly becomes intractable in large or dynamic networks. In this work, we investigate a simulation-based framework where both the defender and attacker are modeled as reinforcement learning (RL) agents. Using a fixed network topology, episodes of play simulate interdiction and defense actions, evaluate post-interdiction maximum flow, and provide rewards to each agent. The defender learns policies that maximize residual flow, while the attacker learns to minimize it, yielding a competitive zero-sum setting. The simulation demonstrates that both agents adaptively learn mixed strategies and converge toward a stable equilibrium distribution.

1 INTRODUCTION

Network interdiction arises when adversaries disrupt critical infrastructures while defenders seek to maintain performance (Brown et al. 2006). Applications include power grids, transportation, and military logistics (Morton et al. 2007). The maximum flow interdiction variant captures resilience by limiting flow between a source and sink through arc removal or protection. Traditionally, interdiction is formulated as a bilevel Stackelberg game solved with mixed-integer programming or heuristics (Smith and Song 2020), but these approaches become intractable on large or dynamic networks and assume perfect rationality. Recent advances in reinforcement learning (RL) and simulation provide alternatives: modeling attacker and defender as competing RL agents can yield equilibrium-like policies (Zhang et al. 2021), while simulation enables scalability and experimentation under uncertainty.

2 PROBLEM DESCRIPTION

We consider a directed network $G = (V, E)$ with source S, sink T, and arc capacities $\{C_{ij}\}_{(i,j)\in E}$. Two decision-making agents interact on this network: a defender, who may protect up to D arcs per round, and an attacker, who may interdict up to A arcs. The defender's objective is to maximize the remaining network flow after interdiction, while the attacker aims to minimize it.

At each round t, the state of the environment is represented as $s_t = \{C_{ij}, X_{ij}^{(t)}, Y_{ij}^{(t)}, Z_{ij}^{(t)}\}_{(i,j)\in E}$, where C_{ij} is the arc capacity, $X_{ij}^{(t)}$, $Y_{ij}^{(t)}$, $Z_{ij}^{(t)}$ indicate whether arc (i, j) is defended, attacked, available.

The defender chooses a protection action $a_{t+1}^{D} \sim \pi_D(\cdot|s_t)$, selecting up to D arcs, while the attacker chooses an interdiction action $a_{t+1}^{A} \sim \pi_A(\cdot|s_t)$, selecting up to A arcs. The environment resolves these actions with defense precedence, so defended arcs remain functional, while undefended attacked arcs are removed. After resolution, the maximum flow F_{t+1} is recomputed on the residual network.

Rewards are defined in a zero-sum form based on normalized flow $\hat{F}_{t+1} = F_{t+1}/F_0$, where F_0 is the baseline flow in the intact network: $r_{t+1}^{D} = \hat{F}_{t+1}$, $r_{t+1}^{A} = 1 - \hat{F}_{t+1}$. This reward structure aligns the defender's utility with network resilience and the attacker's utility with flow disruption.

3 SIMULATION FRAMEWORK

Figure 1 embeds the attacker and defender agents in a simulation loop, where states, actions, and rewards are exchanged at each round. The framework evaluates the effects of interdiction and protection, computes maximum flow using the residual network, and provides feedback to the agents. Policies are updated through reinforcement learning algorithms coded in Python and linked to the simulation platform via Pypeline to communicate between the learning module and the simulation.

It allows both agents to learn adaptive strategies through repeated interactions rather than relying on optimization models. Over many episodes, the defender and attacker policies are expected to approximate mixed strategies, offering insight into equilibrium behavior under adversarial conditions.

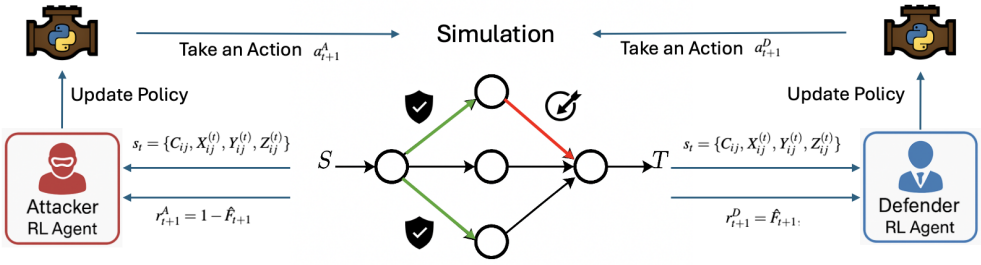
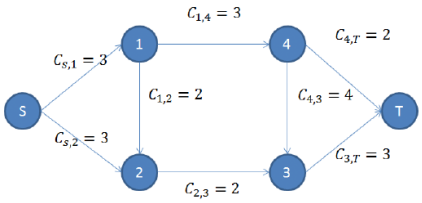
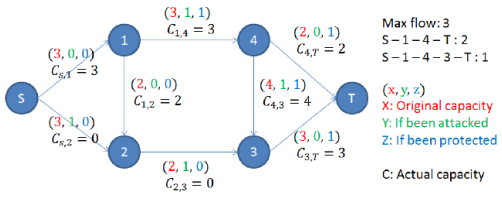


Figure 1: Simulation-based multi-agent reinforcement learning framework for network interdiction.

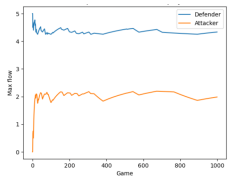
4 EXPERIMENT



(a) The example directed graph.



(b) Max-flow after a step.



(c) Best expected results for each agent.

Figure 2: Experimental results on a sample interdiction game.

We conduct an experiment on the network in Figure 2a with a defense budget of three and an attack budget of two. In the sequential simulation, both agents converge after 1,000 rounds to mixed strategies: $[0.338, 0.131, 0.531]$ for the defender and $[0.226, 0.552, 0.222]$ for the attacker. As shown in Figure 2c, the defender's expected reward stabilizes above 4, while the attacker's remains around 2, reflecting the defender's structural advantage when protecting more arcs than can be interdicted. Compared to this adaptive outcome, stochastic programming, in which both players assumes the opposite side will choose all strategies equally, yields 3.5 for the defender and 1.875 for the attacker, illustrating that equal-probability game favors the attacker, while sequential learning can help the defender to find a better strategy.

REFERENCES

Brown, G., M. Carlyle, J. Salmerón, and K. Wood. 2006. "Defending Critical Infrastructure". *Interfaces* 36(6):530–544.

Morton, D. P., F. Pan, and K. J. Saeger. 2007. "Models for Nuclear Smuggling Interdiction". *IIE Transactions* 39(1):3–14.

Smith, J. C., and Y. Song. 2020. "A Survey of Network Interdiction Models and Algorithms". *European Journal of Operational Research* 283(3):797–811.

Zhang, K., Z. Yang, and T. Başar. 2021. "Multi-Agent Reinforcement Learning: A Selective Overview of Theories and Algorithms". *Handbook of Reinforcement Learning and Control*:321–384.