Deep Reinforcement Learning for Green Security Games with Real-time Information

Yufei Wang[1], Ryan Shi[2], Lantao Yu[3], Yi Wu[4],
Rohit Singh[5], Lucas Joppa[6], Fei Fang[2]

Peking University[1], Carnegie Mellon University[2]
Stanford University[3], University of California, Berkeley[4]
World Wide Fund for Nature[5], Microsoft Research[6]

AAAI 19

Green Security Challenges

Environmental Resources

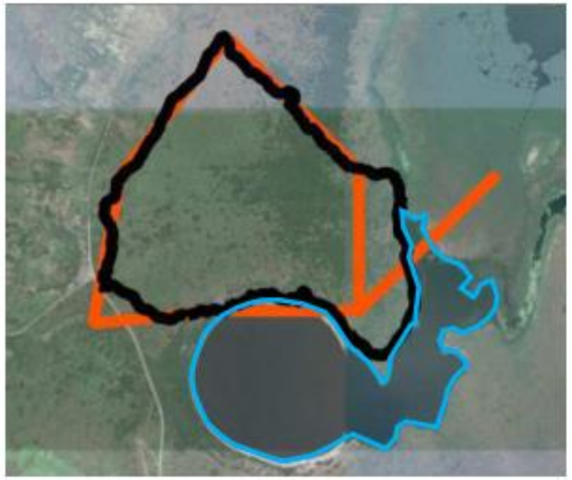
Endangered Wildlife

Fisheries

122 rhinos in 2009 **1215** rhinos in 2014

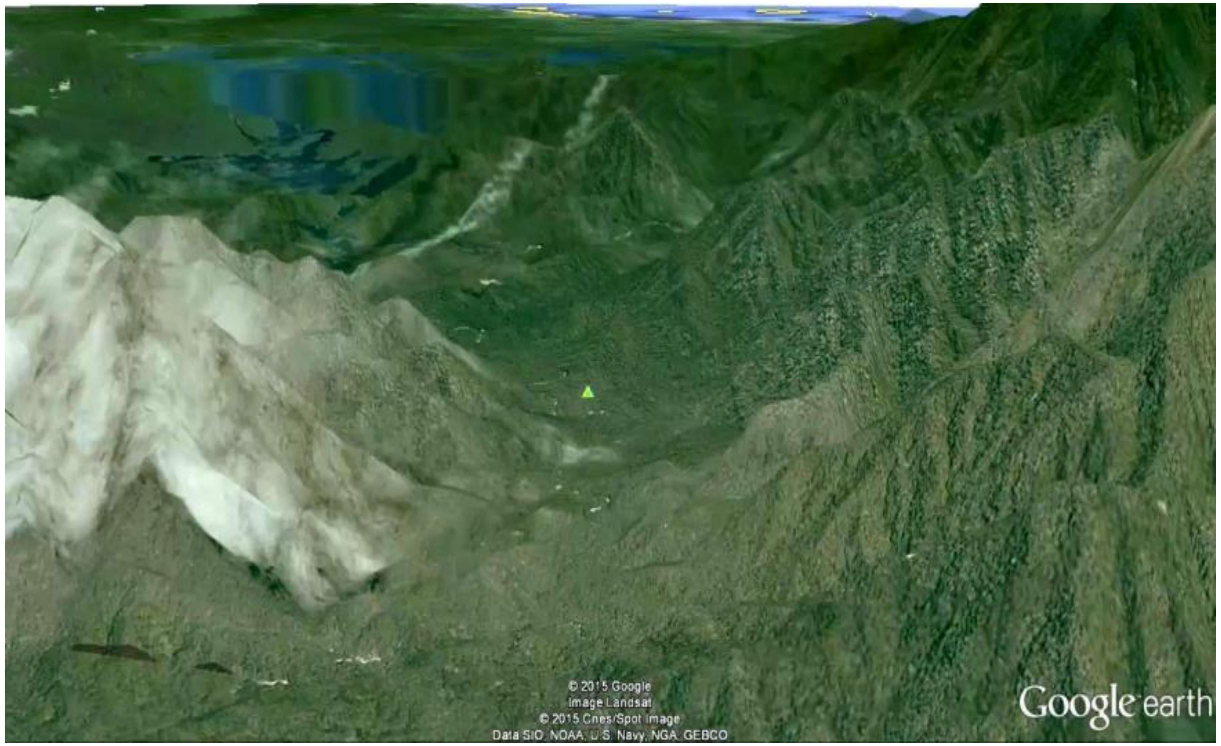
Green Security Games

- Green Security Games model the strategic interaction between law enforcement agencies (defenders) and their opponents (attackers). [Fang, Stone, and Tambe 2015; Fang et al. 2016; Xu et al. 2017]
- Design patrol routes with a limited number of patrol resources.



- However, previous work mainly focuses on computing patrol strategy offline.

The effect of Real-time Information



Our contribution

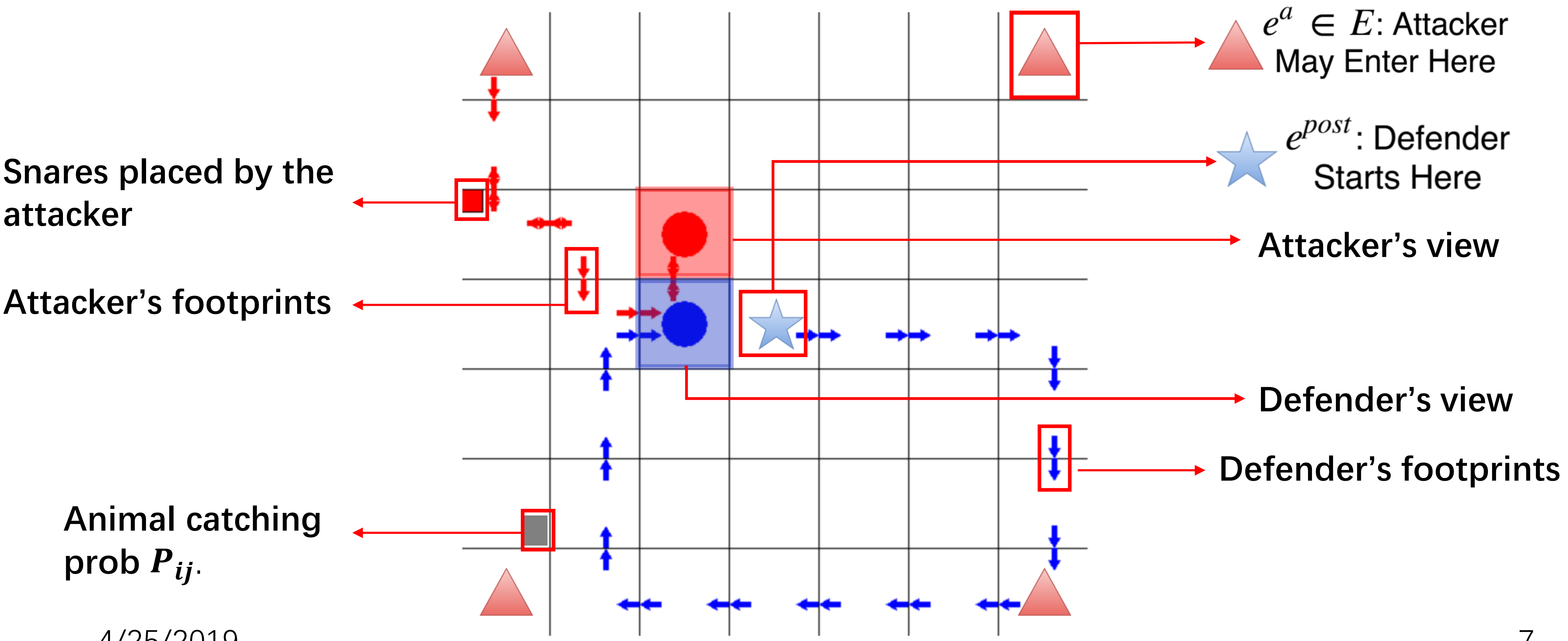
- We propose a game model GSG-I for green security games, that incorporates the vital real-time information.

- We design an efficient algorithm DeDOL that combines deep reinforcement learning and the classic Double Oracle framework in security games to solve zero-sum GSG-I.

- DeDOL is built upon the Policy Space Response Oracle framework, with two important domain-specific enhancements that improves the training efficiency.

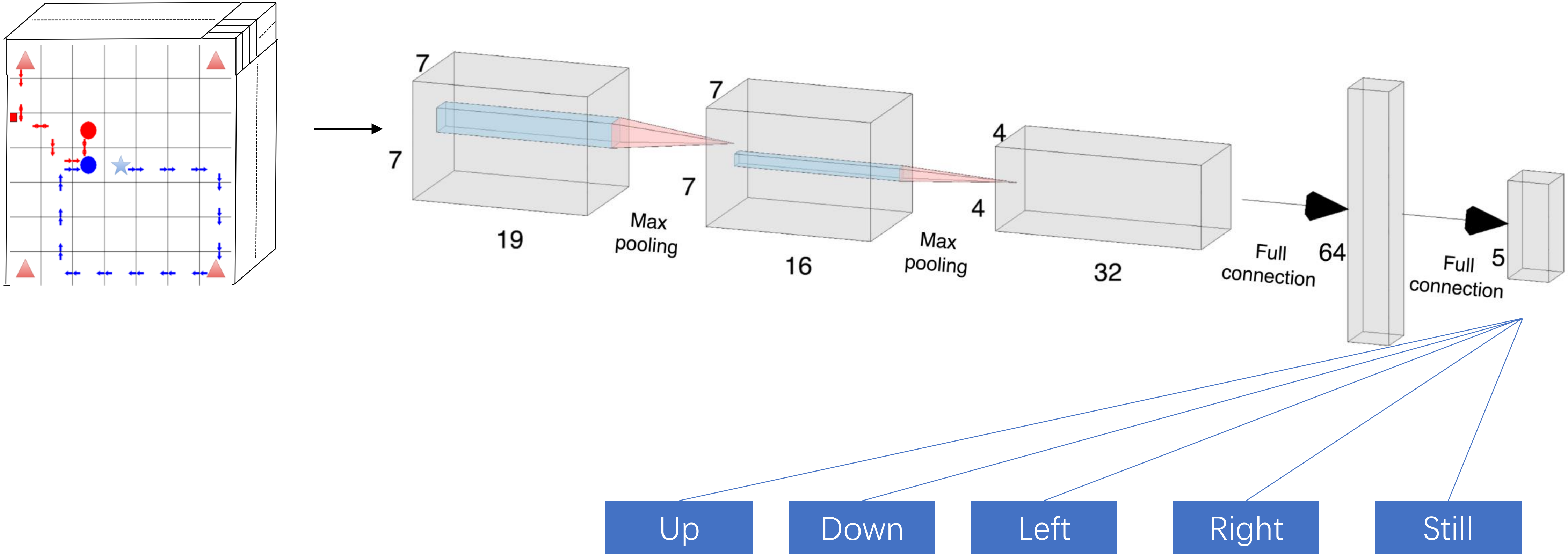
Our proposed GSG-I Model

- GSG-I: Green Security Games with Real-Time Information



Approximating Best Response against a fixed opponent with DQN

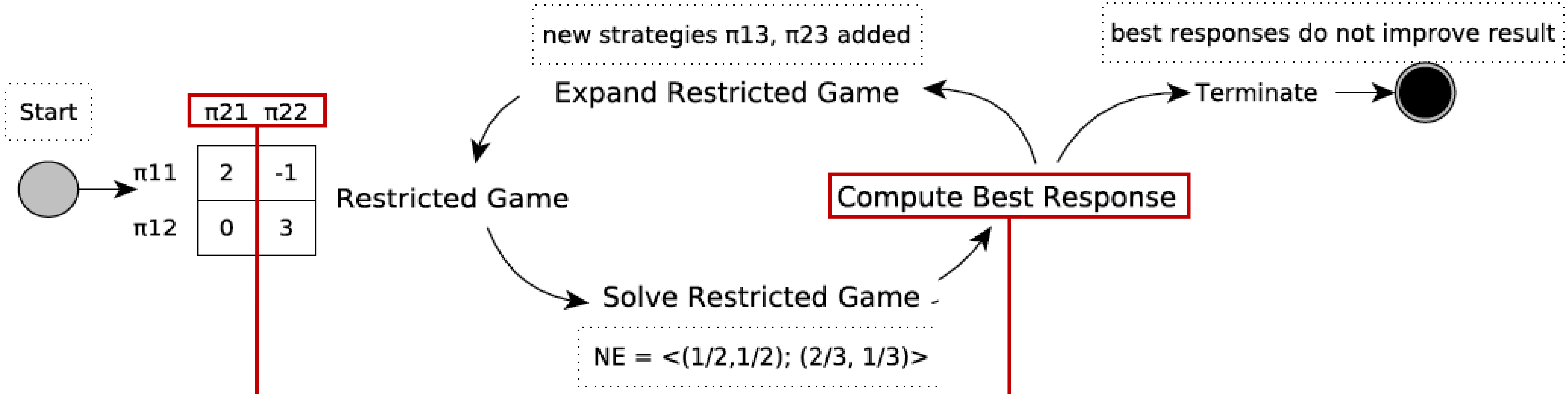
[Mnih et al. 2015]



$$Q(s_t, a_t) \leftarrow r_t + \gamma Q^{\text{target}}(s_{t+1}, \arg\max_{a'} Q(s_{t+1}, a'))$$

Double Oracle & Policy Space Response Oracle

[McMahan, Gordon, and Blum 2003; Lanctot et al. 2017]



Parameterized policy + Approximate Best Response using Deep RL → PSRO

Initial heuristic strategy

Attacker: parameterized heuristic

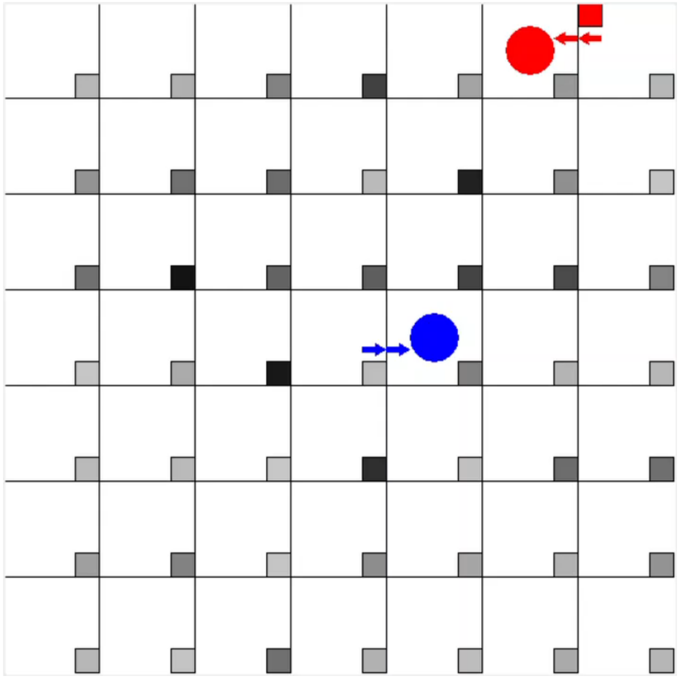
- Moving Direction

$$\pi_a(a_t^a = k | s_t^a) = \frac{\exp(w_p \cdot \bar{P}_k + w_i \cdot I_k + w_o \cdot O_k)}{\sum_z \exp(w_p \cdot \bar{P}_z + w_i \cdot I_z + w_o \cdot O_z)}$$

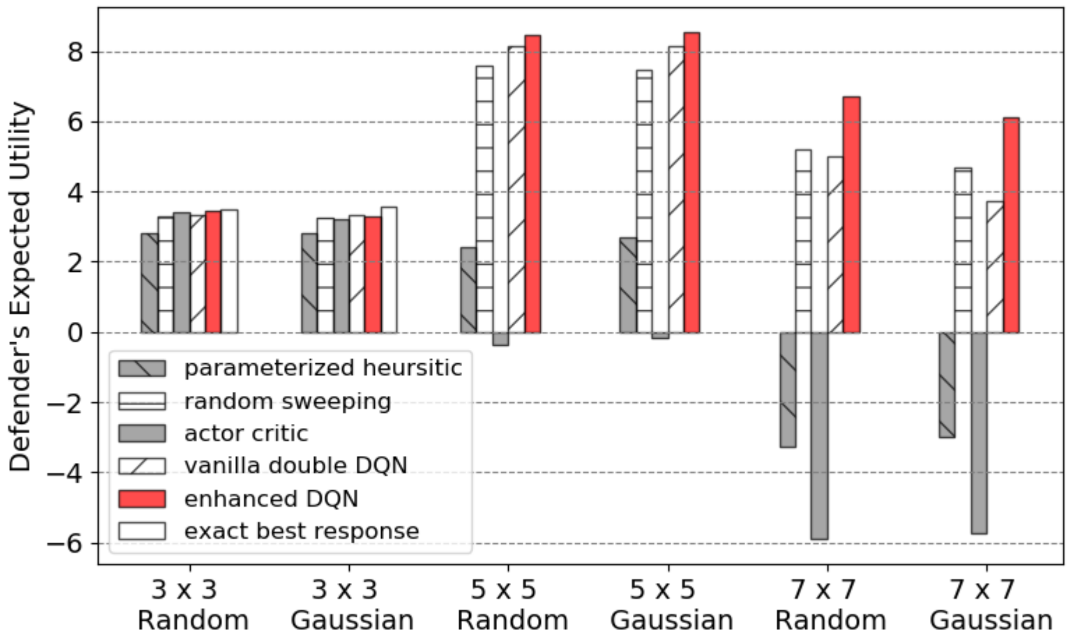
- Attacking Tool Placement

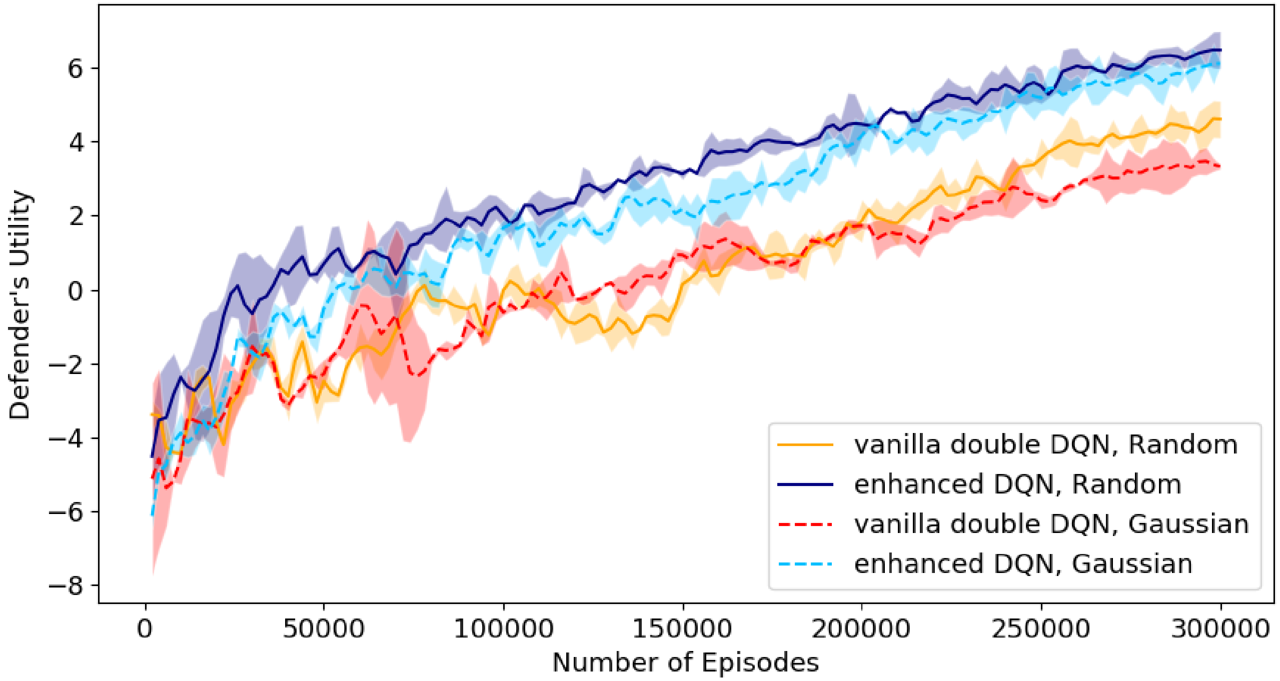
$$\eta_a(b_t^a = 1 | s_t^a) = \frac{\exp(P_{m,n} / \tau)}{\sum_i \sum_j \exp(P_{i,j} / \tau)}$$

Defender: random sweeping

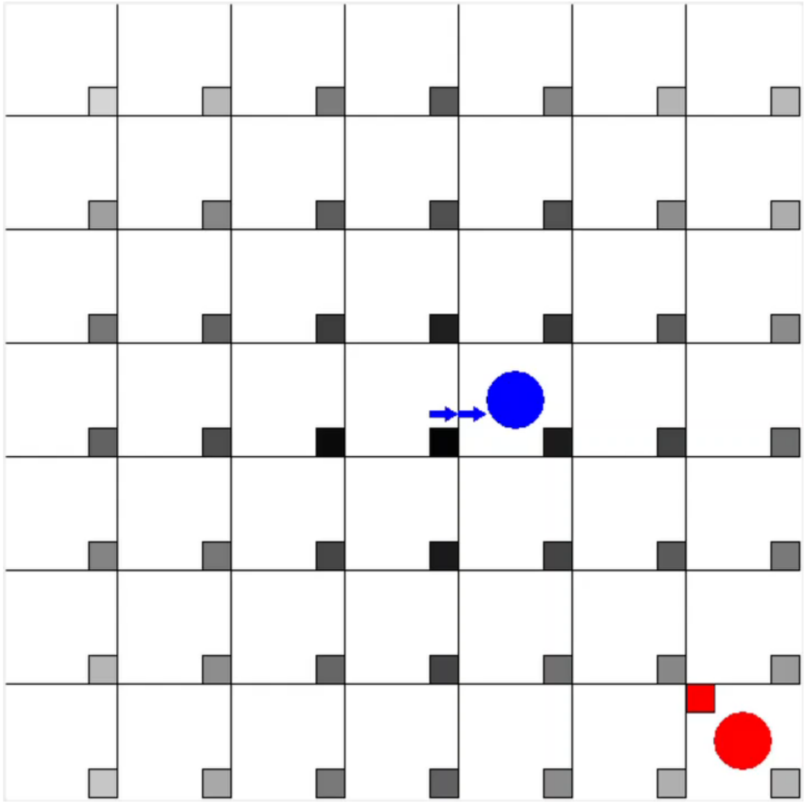


DQN strategy performance

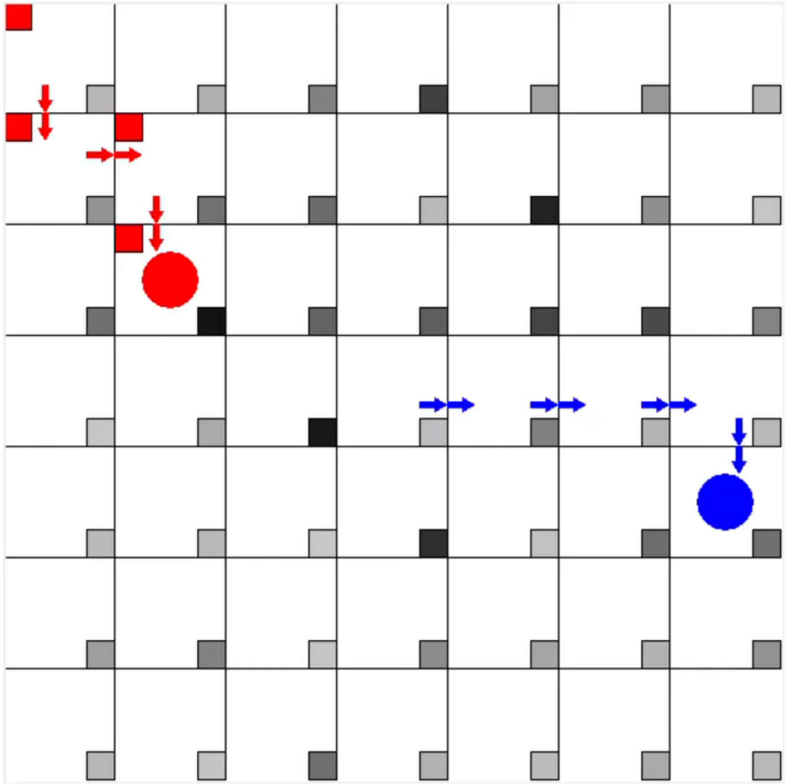




DQN strategy demo

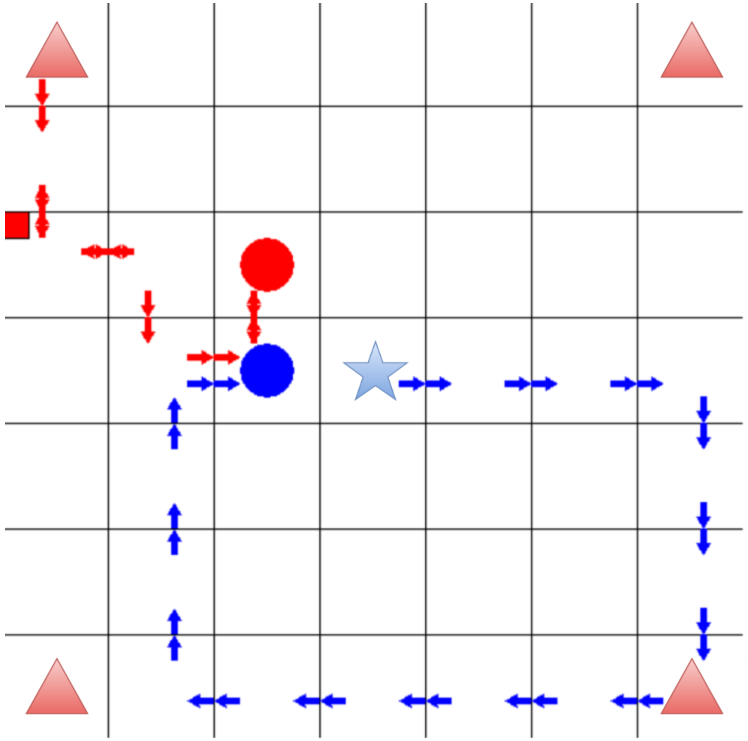


DQN patroller against a heuristic poacher on 7x7 Gaussian grid.



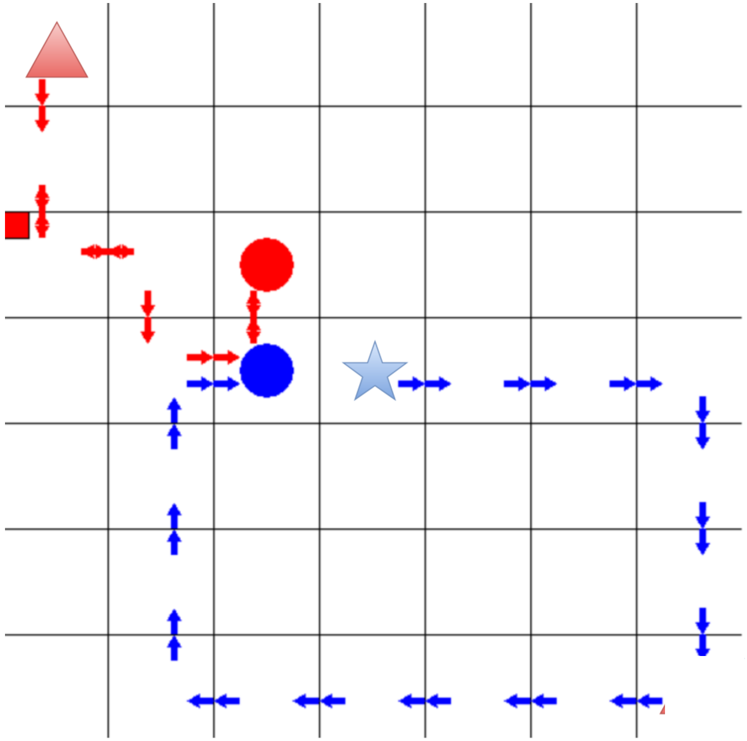
DQN poacher against a random sweeping patroller on 7x7 random grid.

Local modes: restrict attacker's entry point



Ease training

Provide good building block strategy

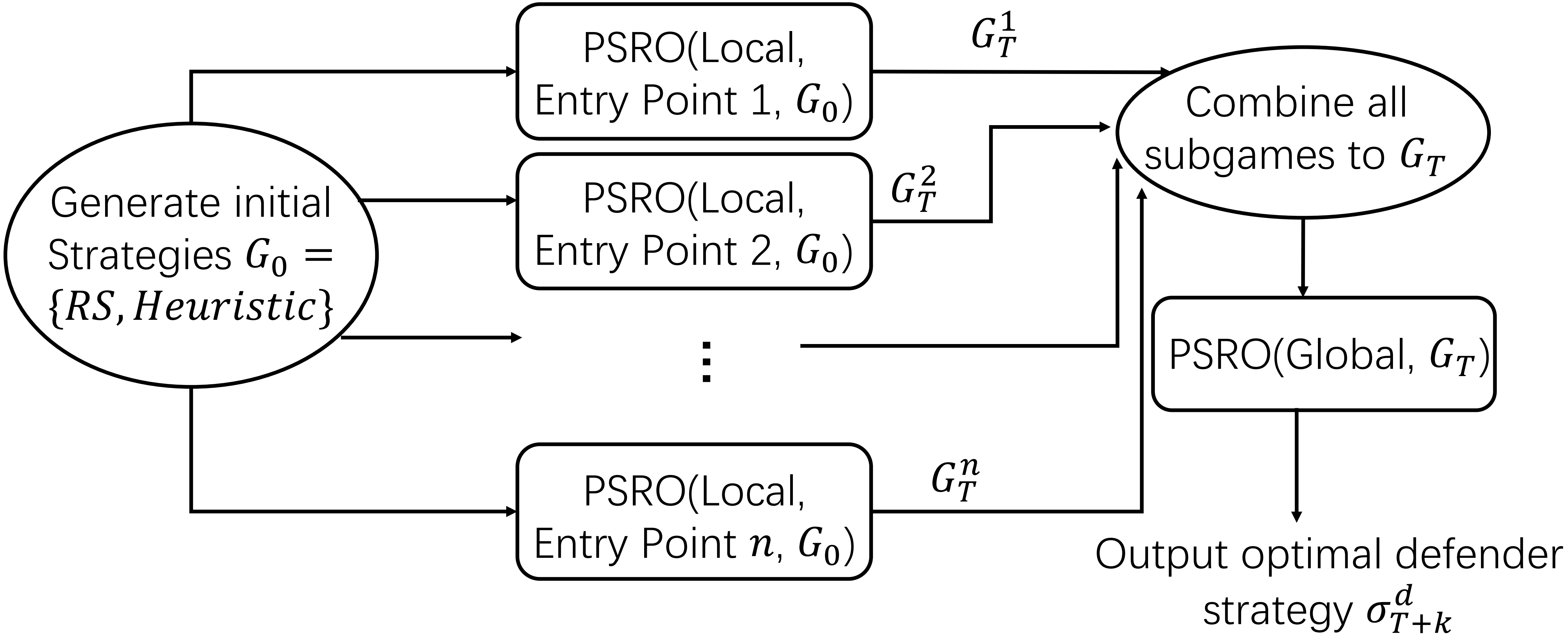


Global mode with 4 entry points

local mode with 1 top-left entry point

The DeDOL algorithm workflow

- Deep-Q Network based Double Oracle enhanced with Local modes



DeDOL Performance

	Random Sweeping	Vanilla PSRO	DeDOL Pure Global Mode	DeDOL Local + Global Mode	DeDOL Pure Local Mode	CFR
3×3 Random	-0.04	0.65 (16)	0.73 (16)	**0.85** (10 + 2)	0.71 (20)	**1.01** (3500)
3×3 Gaussian	-0.09	0.52 (16)	0.75 (16)	**0.86** (10 + 2)	0.75(20)	**1.05** (3500)
5×5 Random	-1.91	-8.98 (4)	-1.63 (4)	-0.42 (4 + 1)	**-0.25** (5)	-
5×5 Gaussian	-1.16	-9.09 (4)	-0.43 (4)	**0.60** (4 + 1)	-2.41 (5)	-
7×7 Random	-4.06	-10.65 (4)	-2.00 (4)	**-0.54** (3 + 1)	-1.72(5)	-
7×7 Gaussian	-4.25	-10.08 (4)	-4.15 (4)	**-2.35** (3 + 1)	-2.62(5)	-

- Vanilla PSRO: no heuristic initial strategy, no local mode
- CFR: counter-factual regret minimization
- Metric: highest expected utility against a best-response poacher across all iterations
- **DeDOL with local mode** performs the best!

Summary

- We incorporates the real-time information in Green Security Games, and proposed a new game model.
- We design an efficient algorithm DeDOL to compute the optimal patrol strategy in GSG-I, which is built upon the Double Oracle / Policy Space Response Oracle framework with domain-specific enhancements.

Thank you!

Yufei Wang
wang.yufei@pku.edu.cn

Zheyuan Ryan Shi
ryanshi@cmu.edu

Lantao Yu
lantaoyu@hotmail.com

Yi Wu
jxwuyi@gmail.com

Rohit Singh
rsingh@wwfnet.org

Lucas Joppa
lujoppa@microsoft.com

Fei Fang
feifang@cmu.edu

The Azure resources are provided by Microsoft for Research AI for Earth award program.