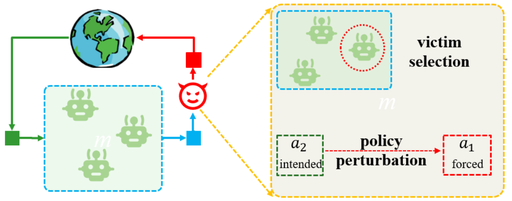
Robust Multi-agent Coordination via Evolutionary Generation of Auxiliary Adversarial Attackers

Lei Yuan*, Ziqian Zhang*, Ke Xue, Hao Yin, Feng Chen,
Cong Guan, Lihe Li, Chao Qian, Yang Yu
Nanjing University, Polixir Technologies

Introduction

Despite of promising potential of MARL, lack of robustness makes it difficult to be applied in the real world.

To solve a robust policy, we formalize LPA-Dec-POMDP and train a robust policy via evolutionary generation of attackers.



Method

Attacker Optimization Objective:

- minimize the reward of the ego-system
- sparsity prior regularization
- JSD diversity regularization

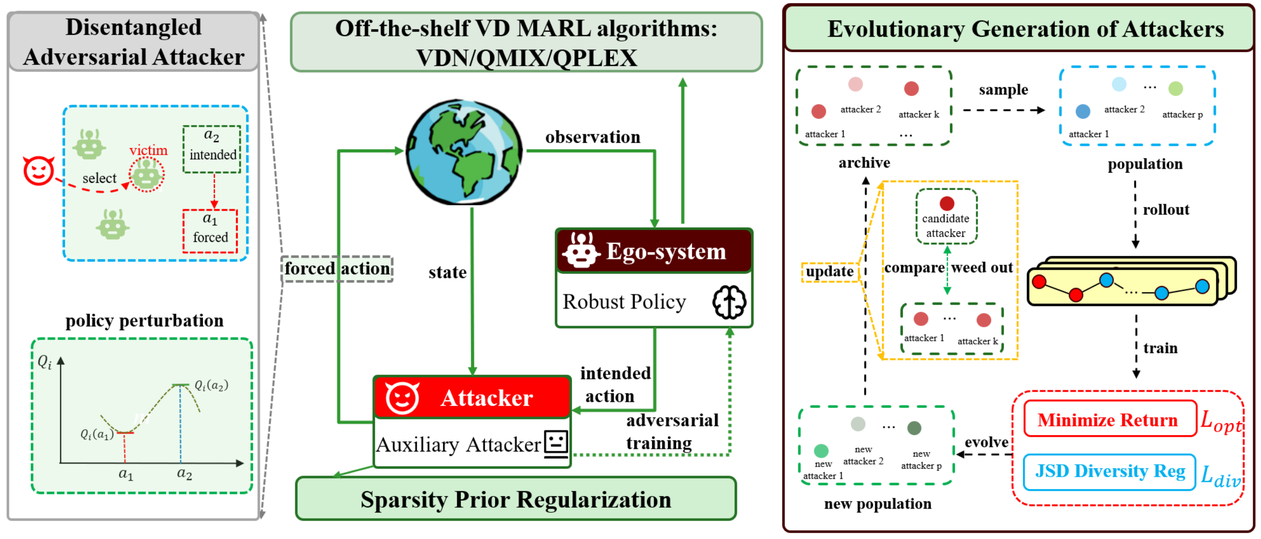
Attacker Population Generation:

- Quality-Diversity algorithm
- customized selection and update mechanism

Robustness Training Paradigm:

- application of any off-the-shelf MARL algorithm

Workflow

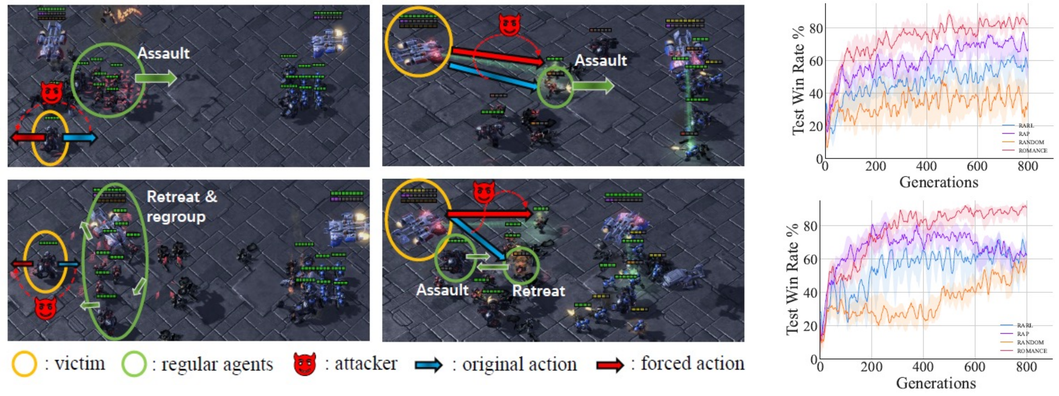


Experiments

Performance on different SMAC maps

Method	Map_Name	2s3z $K=8$	3m $K=4$	3s_vs_3z $K=8$	8m $K=5$	MMM $K=8$	1c3s5z $K=6$	+/−/≈
Natural	vanilla QMIX	92.8 ± 1.62	$\mathbf{97.9 \pm 1.02}$	98.3 ± 0.78	98.2 ± 0.45	95.8 ± 1.59	88.8 ± 2.13	1/1/4
	RARL	96.4 ± 1.19	86.0 ± 5.38	80.6 ± 27.5	95.3 ± 3.31	89.3 ± 7.01	76.9 ± 9.85	0/4/2
	RAP	$\mathbf{98.1 \pm 0.76}$	91.3 ± 4.93	$\mathbf{99.3 \pm 0.51}$	91.7 ± 7.96	95.3 ± 4.98	86.7 ± 10.5	0/1/5
	RANDOM	98.0 ± 0.60	95.3 ± 2.07	99.6 ± 0.35	$\mathbf{98.6 \pm 0.90}$	93.8 ± 7.56	93.1 ± 4.41	1/0/5
	ROMANCE	97.9 ± 1.34	96.0 ± 1.83	97.8 ± 1.78	94.3 ± 3.94	$\mathbf{97.1 \pm 1.49}$	$\mathbf{93.9 \pm 1.24}$	
Random Attack	vanilla QMIX	78.8 ± 1.28	$\mathbf{78.7 \pm 1.49}$	87.0 ± 0.36	66.2 ± 2.08	70.0 ± 3.97	66.6 ± 2.03	0/5/1
	RARL	84.3 ± 2.40	67.6 ± 5.01	70.1 ± 29.1	75.7 ± 7.00	62.2 ± 10.2	56.5 ± 10.8	0/5/1
	RAP	87.3 ± 1.87	73.5 ± 3.49	89.8 ± 4.81	$\mathbf{78.4 \pm 8.22}$	84.2 ± 9.05	66.8 ± 9.66	0/1/5
	RANDOM	83.9 ± 6.38	76.4 ± 2.27	91.9 ± 1.32	72.0 ± 3.46	72.9 ± 7.09	60.5 ± 21.3	0/2/4
	ROMANCE	$\mathbf{89.1 \pm 1.97}$	78.1 ± 5.13	$\mathbf{93.0 \pm 1.82}$	76.2 ± 5.36	$\mathbf{85.8 \pm 8.66}$	$\mathbf{77.9 \pm 1.96}$	
EGA	vanilla QMIX	26.7 ± 4.28	20.7 ± 2.13	30.9 ± 1.52	42.7 ± 9.79	37.9 ± 3.13	35.2 ± 8.66	0/6/0
	RARL	56.1 ± 11.8	86.1 ± 0.98	60.9 ± 14.2	66.3 ± 7.25	41.5 ± 11.6	35.3 ± 4.00	0/6/0
	RAP	64.1 ± 11.9	84.0 ± 4.27	65.1 ± 4.41	84.4 ± 8.88	74.9 ± 15.5	45.4 ± 6.83	0/4/2
	RANDOM	48.3 ± 17.3	66.2 ± 16.6	54.4 ± 7.83	55.6 ± 12.5	53.1 ± 6.09	43.3 ± 10.3	0/6/0
	ROMANCE	$\mathbf{81.6 \pm 0.84}$	$\mathbf{89.7 \pm 1.52}$	$\mathbf{90.5 \pm 1.97}$	$\mathbf{86.2 \pm 5.11}$	$\mathbf{84.0 \pm 11.5}$	$\mathbf{66.5 \pm 3.24}$	

Visualization Analysis



- Policy learned without ROMANCE ignores the emergent situation and still tries to assault when the attacked marauder is drawn away.
- Policy learned with ROMANCE learns to wait for the victim Marauder
- to regroup for efficient coordination.