– Supplementary Material – Hijacking Robot Teams Through Adversarial Communication

Anonymous Author(s)
Affiliation
Address
email

Appendix A Real-World Demonstrations: Robotarium

We use Robotarium [20], a free remotely accessible swarm robotics research platform, to do real-world demonstrations. It is equipped with a group of miniature differential drive robots 'GRITSBots' on a testbed measuring 130×90×180 cm, with a projector and an automatic overhead tracking system. The GRITSBot's main board has WiFienabled 160 MHz ESP8266 chip as the controller and communication (54 MBit/s WiFi) and the stepper motors droven by Atmega 168 microcontroller [20]. The global position is tracked using an overhead camera and then used down-stream for safety checking and feedback control. Features of the robot environment are displayed by the projector for visualization.

First, we need to run our algorithm in the robotarium simulator before implementing it on the real platform. However, physical collisions are strictly prohibited when using actual robots. To overcome this limitation, we record the trajectories of each agent in the real environment and perform post-analysis to determine if there are any instances where two robots collide. This analysis is based on the relative distance between the robots, following our predefined criteria. A collision between two robots is defined as when the circles centered on each robot intersect. The radius of each circle is defined according to the environment specifications [23]. The reward for the environments is defined as the L2 distance between the robot and its target destination. In PP and PCP, the target is the prey robot. In SL, the target is the designated goal location. We show the trajectories we collected in each environment Fig 1.

We also include the average reward and collision numbers of each robot in Tables 1-6. It shows that our adversarial method universally outperforms the random flipping one for each agent since it makes the attacked agents receive less reward and has fewer collisions with their targets. Moreover, we find that our method is even more stable than the random flipping one, with standard deviation only decreasing by 22.97%, 53.33%, and 40.89% on average in the three environments.

Submitted to the 7th Conference on Robot Learning (CoRL 2023). Do not distribute.

Table 1: PCP Reward

	Capture Agent 1	Capture Agent 2	Average
Adv[Ours]	**-0.76±0.32**	**-0.66±0.36**	**-0.71±0.34**
Random	-0.27±0.29	-0.090±0.36	-0.18±0.33

Table 2: PCP Collisions

	Capture Agent 1	Capture Agent 2	Average
Adv[Ours]	**0.02±0.14**	**0.09±0.28**	**0.05±0.23**
Random	0.17±0.37	0.40±0.49	0.29±0.45

Table 3: SL Reward

	Listener
Adv[Ours]	**-0.43±0.14**
Random	-0.21±0.15

Table 4: SL Collisions

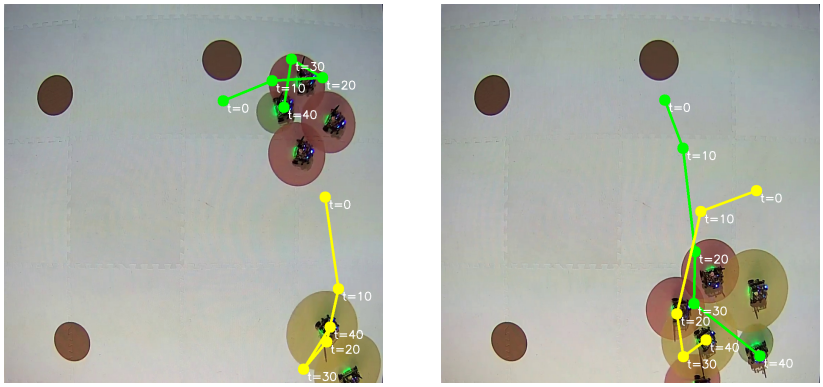
	Listener
Adv[Ours]	**0.00±0.00**
Random	0.34±0.47

Table 5: PO-PP Reward

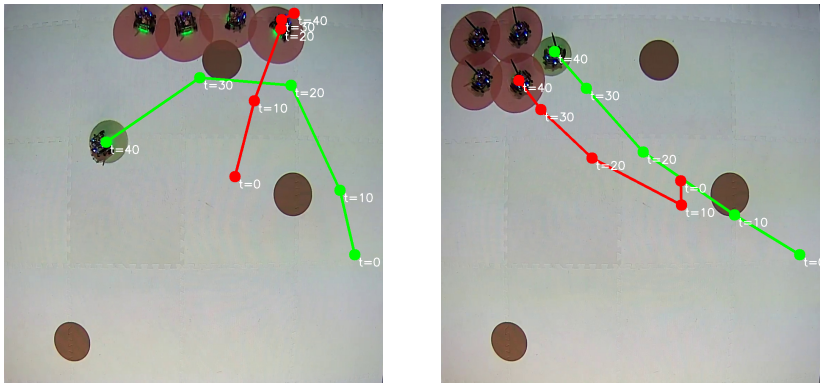
	PO Agent 1	PO Agent 2	PO Agent 3	PO Agent 4	Average
Adv[Ours]	**-0.81±0.34**	**-0.80±0.27**	**-0.71±0.31**	**-0.80±0.30**	**-0.71±0.34**
Random	0.01±0.56	-0.06±0.54	-0.00±0.60	-0.03±0.54	0.02±0.56

Table 6: PO-PP Collisions

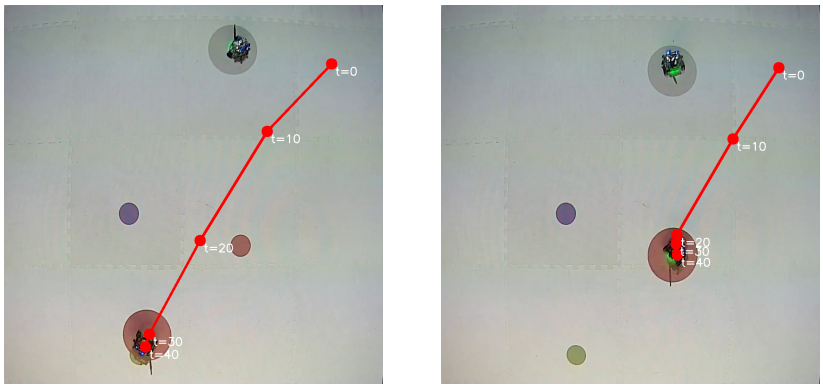
	PO Agent 1	PO Agent 2	PO Agent 3	PO Agent 4	Average
Adv[Ours]	**0.00±0.06**	**0.00±0.00**	**0.00±0.00**	**0.00 ±0.05**	**0.05±0.23**
Random	0.26±0.44	0.06±0.33	0.33±0.47	0.15±0.35	0.20±0.40



(a) Predator Capture Prey



(b) Partially Observable PP



(c) Speaker Listener

Figure 1: Comparison of Environment Trajectories: All three environments are shown, where the left images are the adversarial communication policy rollouts and the right images are the random flipping rollouts.

Appendix B Simulation Experiment Details

Here we show the hyperparameters used in each environment training (Table 7) and qualitative results (Figure 2).

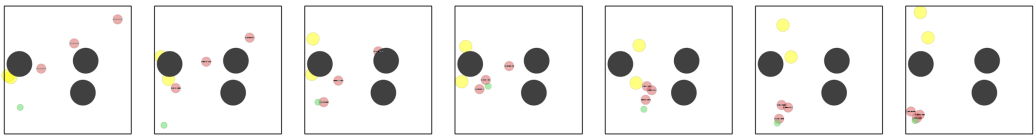
In the PCP environment (a, b), the predators (also called perception agents) are shown as red which can observe all other agents, however, the yellow capture agent (also called action agents) are blind and can only know where the prey (green) is by receiving the messages from the predators. Therefore, communication is the only useful information based on which the capturers can make decisions. Each capture agent will receive a 16-bit communication from each of the three predators so we intercept 48 bits and modify them with our adversarial policy. Compare with Figure 2(a) and 2(b), we find that our adversarial policy can successfully push the captures agents away from the prey but the random flipping one cannot stop the capture agents from pursuing the prey with the same number of bits flipped.

We observe similar behaviors in PO-PP when we compare Figure 2(c) and 2(d), in which the predators and prey are shown with red and green. The difference between PO-PP and PCP environments is that we remove the capture agents but change the predators to be partially observable agents which can only see the prey within a certain distance. Predators change color from red to grey if they observe the prey. If one predator observes the prey, it can broadcast this information to others with its 16-bit communication so that the team can cooperate with each other to achieve higher rewards. When we apply the adversarial policy (see Figure 2), we find that the predators just ignore the prey even though they see it and never collaborate to collide with the prey compared with the random flipping one in Figure 2.

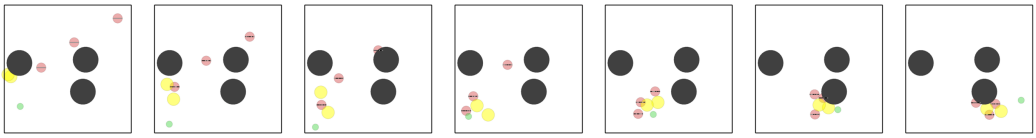
In the speaker-listener environment (Figure 2(e, f)), the speaker knows the colored goal the listener should go to but the listener does not. However, the listener knows the position of the three colored goals. Therefore, the speaker needs to learn to communicate the correct color within its 16-bit communication and the listener should learn which color it needs to go to from the message. Our adversarial method (2) can make the listener go to a completely wrong colored destination, while the random flipping method cannot because it cannot attack the crucial bits of the communication.

Table 7: Hyperparameters for training and testing PCP, PO-PP and SL

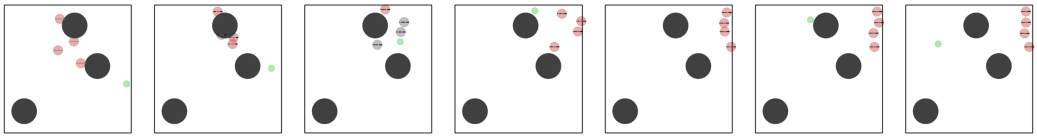
Hyperparameter	Environment	Value
Buffer Length	PCP, PO-PP, SL	1048576
Episode Number	PCP, PO-PP, SL	50001
Episode Length	PCP, PO-PP, SL	100
Batch Size	PCP, PO-PP, SL	1024
Discount Factor γ	PCP, PO-PP, SL	0.9
Learning Rate	PCP, PO-PP, SL	0.0001
Regularizer Coefficient α_0	PCP, SL	0.1
Regularizer Coefficient α_0	PO-PP	0.004
Regularizer Intercept β	PCP, PO-PP, SL	3000
Regularizer Slope ϵ	PCP, PO-PP, SL	20000
Perception Threshold η	PCP, PO-PP, SL	3



(a) Predator-Capture-Prey Adversarial Communication

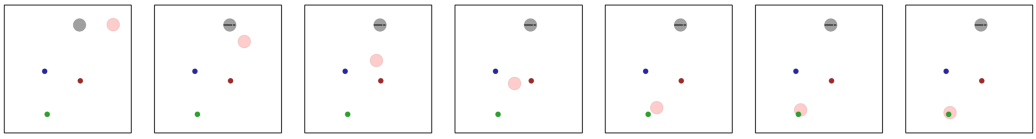


(b) Predator-Capture-Prey Random Flipping

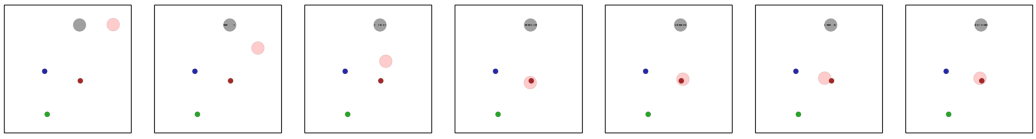


(c) Partial Observability Predator-Prey Adversarial Communication

(d) Partial Observability Predator-Prey Random Flipping



(e) Speaker-Listener Adversarial Communication



(f) Speaker-Listener Random Flipping

Figure 2: These image series show the performance of agents when applying our adversarial communication and random flipping strategy in three environment: Predator-Capture-Prey (a, b), Partial Observability Predator-Prey (c, d) and Speaker-Listener (e, f).

Appendix C Adversarial Message Parameterization

Appendix C.1 Normalized Scores Formulations

As we are making comparisons across different methods, we utilize the score

$$S = \frac{RC_{no\_adv} - RC_{adv}}{\max(N_f - \eta)} \tag{1}$$

where RC_{adv} and RC_{no_adv} represent the reward or the collision number with or without applying the adversarial policy and their difference evaluates how much the adversarial communication channel degrades the agent performance. We normalize by the performance score with the number of bits flipped with a perception threshold η which increases the numeric stability in case of extremely small flipping number N_f.

Appendix C.2 Normalized Score Tables for Attacked Agents

We show detailed tables that quantify the reward and number of collisions for each agent here for our adversarial communication with flipping mode, direct mode and the random flipping. Our proposed method is uniformly better than all other strategies across all attacked agents.

Table 8: PCP Reward Scores

	Capture Agent 1	Capture Agent 2	Average
Adv[Ours]	**0.10±0.06**	**0.12±0.05**	**0.11±0.05**
Adv[Direct]	0.01±0.01	0.01±0.01	0.01±0.01
Random	0.01±0.01	0.01±0.01	0.01±0.01

Table 9: PCP Collision Scores

	Capture Agent 1	Capture Agent 2	Average
Adv[Ours]	**4.53±2.23**	**4.63±2.23**	**4.58±2.23**
Adv[Direct]	1.07±0.14	1.07±0.14	1.07±0.14
Random	0.93±0.29	0.99±0.29	0.96±0.29

Table 10: SL Reward Scores

	Listener
Adv[Ours]	**0.13±0.06**
Adv[Direct]	0.11±0.05
Random	0.04±0.02

Table 11: SL Collision Scores

	Listener
Adv[Ours]	**31.38±6.51**
Adv[Direct]	30.72±6.19
Random	13.68±3.19

Table 12: PO-PP Reward Scores

	PO Agent 1	PO Agent 2	PO Agent 3	PO Agent 4	Average
Adv[Ours]	**0.04±0.02**	**0.04±0.02**	**0.04±0.02**	**0.04±0.02**	**0.04±0.02**
Adv[Direct]	0.02±0.01	0.02±0.01	0.02±0.01	0.02±0.00	0.02±0.01
Random	0.00±0.01	0.00±0.01	0.00±0.01	0.01±0.01	0.00±0.01

Table 13: PO-PP Collision Scores

	PO Agent 1	PO Agent 2	PO Agent 3	PO Agent 4	Average
Adv[Ours]	**1.41±0.55**	**1.44±0.51**	**1.52±0.51**	**1.42±0.65**	**1.45±0.55**
Adv[Direct]	0.76±0.07	0.78±0.07	0.82±0.08	0.84±0.08	0.80±0.07
Random	0.10±0.87	0.34±0.86	0.33±0.92	0.43±0.88	0.30±0.88