25th HumanTech Paper Award

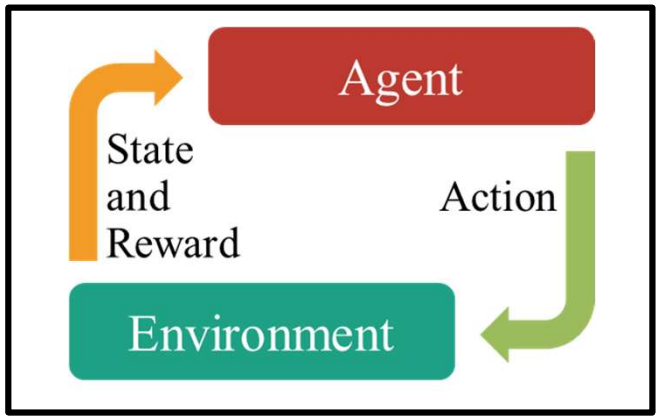
Learning to Schedule Communication in Multi-agent Reinforcement Learning

2019.1.22

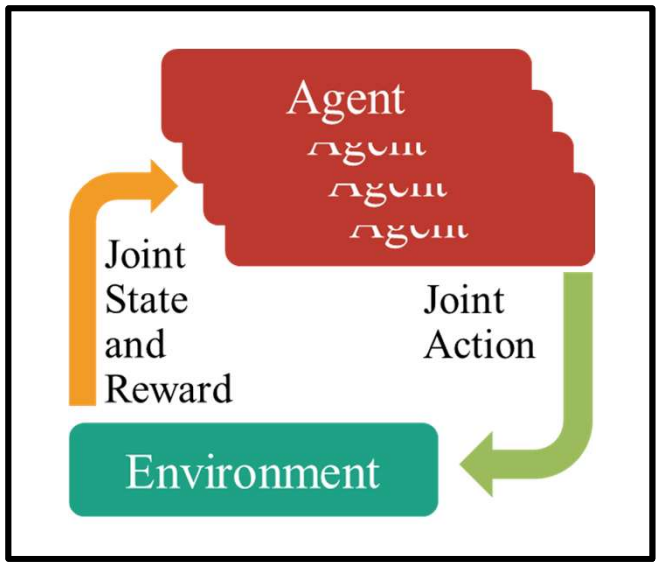
Wan Ju Kang
강완주

Inducing Cooperation among Multiple Agents

- Reinforcement Learning (RL) can model many real-world tasks
  - *e.g.*, drone control for human tracking
- Some multi-agent extensions still remain unconquered
  - Inducing cooperation is non-trivial
  - *e.g.*, cooperative search and rescue robots
- Want to better coordinate multiple agents
  - By means of inter-agent communication



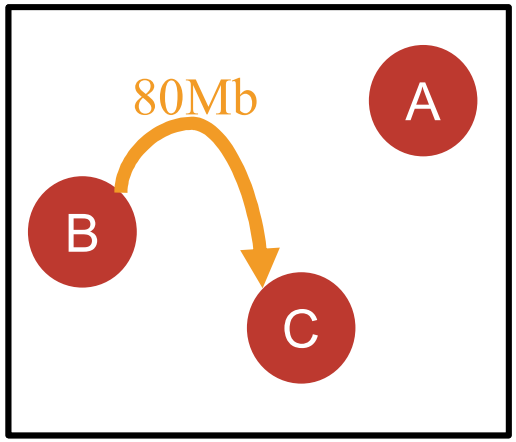
Single-agent RL



Multi-agent RL

Difficulties in Training Communicators

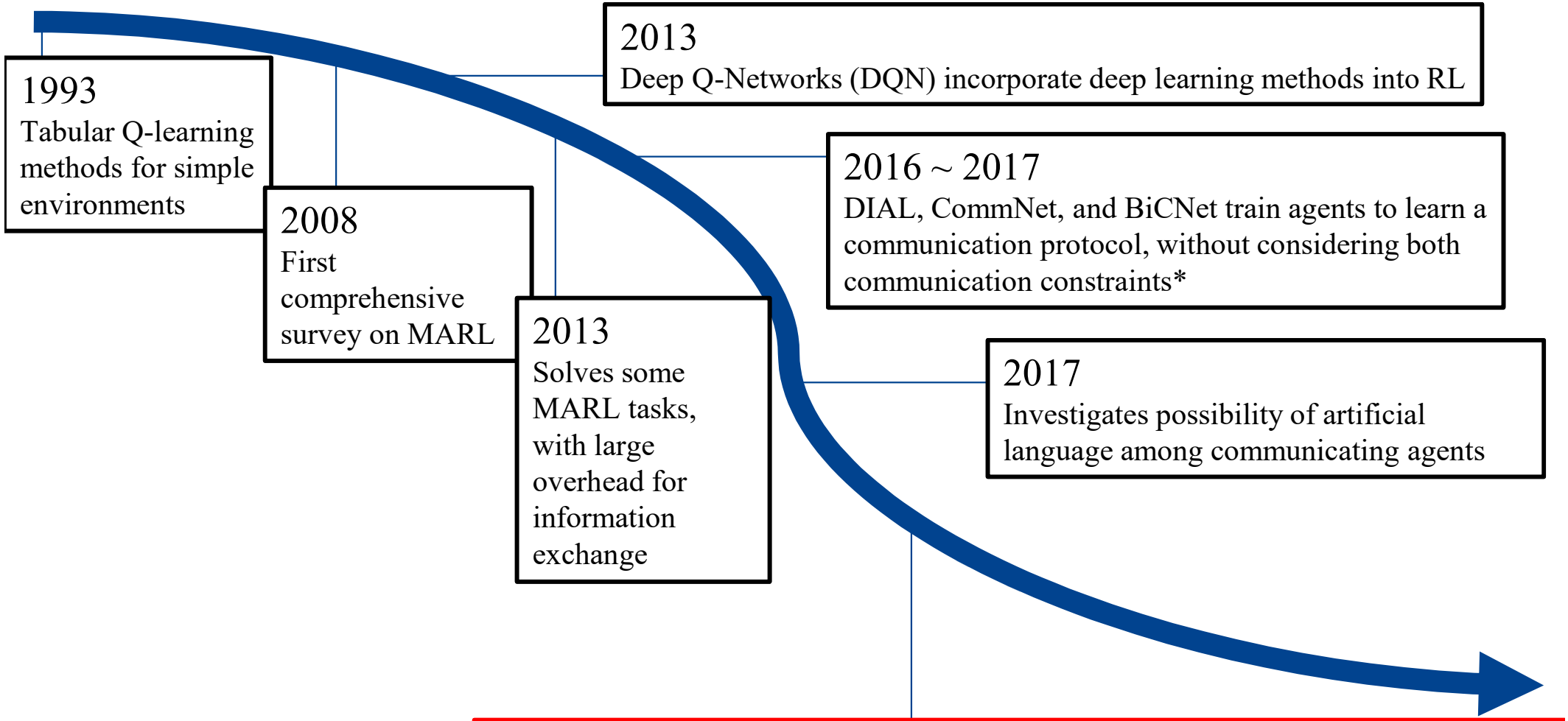
- Bandwidth constraint
  - Need for efficient exchange of succinct information
  - *e.g.*, total capacity of the channel is 100Mbps
  - What messages should be sent over the limited bandwidth?
- Medium access contention
  - Need for efficient allocation of channel resource
  - *e.g.*, only one agent may access the channel at a time
  - Who should be given access to the channel and when?



Agent B is sending a message to Agent C by accessing the channel

- First study to jointly consider both issues

Timeline of Related Work



1993
Tabular Q-learning methods for simple environments

2008
First comprehensive survey on MARL

2013
Solves some MARL tasks, with large overhead for information exchange

2013
Deep Q-Networks (DQN) incorporate deep learning methods into RL

2016 ~ 2017
DIAL, CommNet, and BiCNet train agents to learn a communication protocol, without considering both communication constraints*

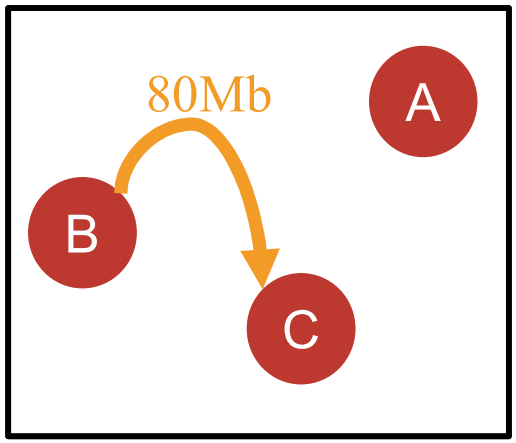
2017
Investigates possibility of artificial language among communicating agents

2019
SchedNet solves MARL tasks by inducing cooperation in a distributed manner as the first study to address both communication constraints*
SchedNet trains agents to learn to gauge the importance of their observation

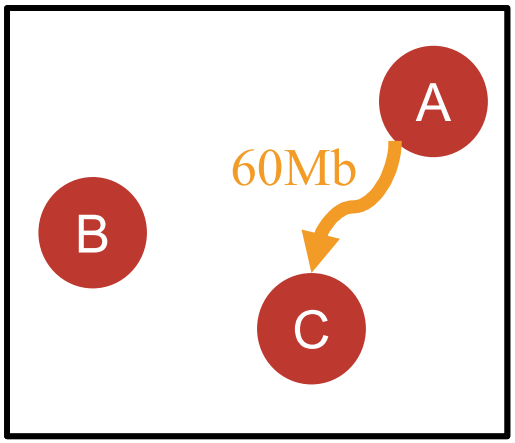
*Communication constraints
- Limited bandwidth
- Medium contention

Communication and Scheduling

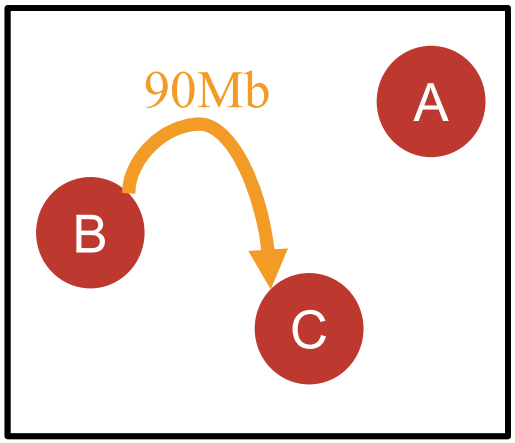
- Bandwidth Constraint → Encoding and Decoding
- Medium Contention → Scheduling
- Effective communicators
  - Know what to send and when to send it
  - *e.g.*, a scenario where three agents must communicate over a 100Mbps channel that allows only one access at a time



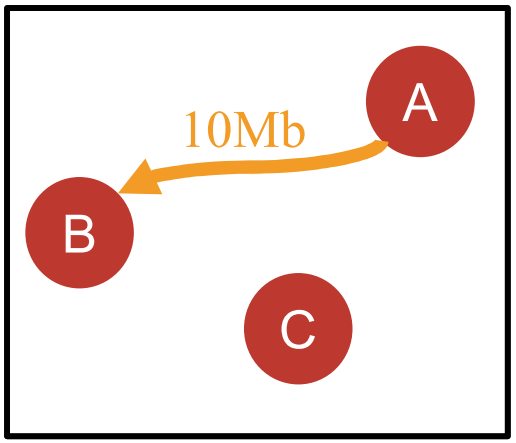
$t = 0$s



$t = 1$s



$t = 2$s

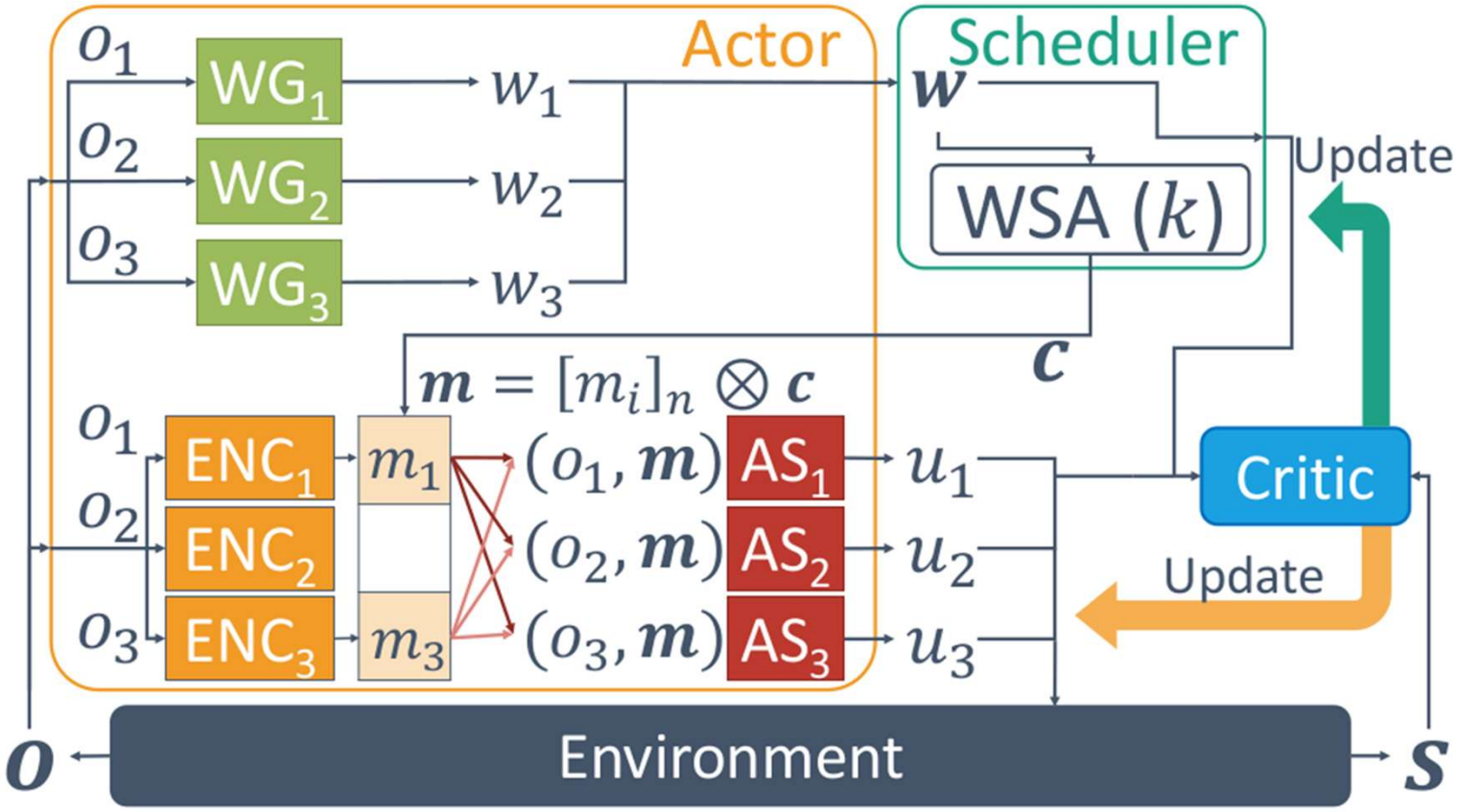


$t = 3$s

Time t (s)	0	1	2	3
Scheduled agent	B	A	B	A

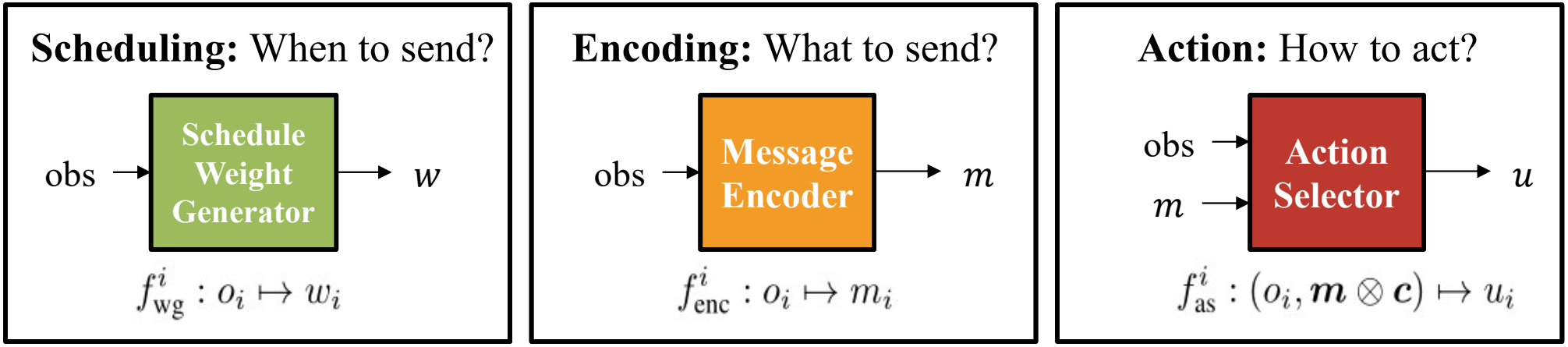
SchedNet: Centralized Training

- Centralized Training and Distributed Execution
  - Allows for the learning of decentralized policies, in a centralized manner
    - multiple actors
    - single critic
  - Popularized in recent works for its scalability and stability in training



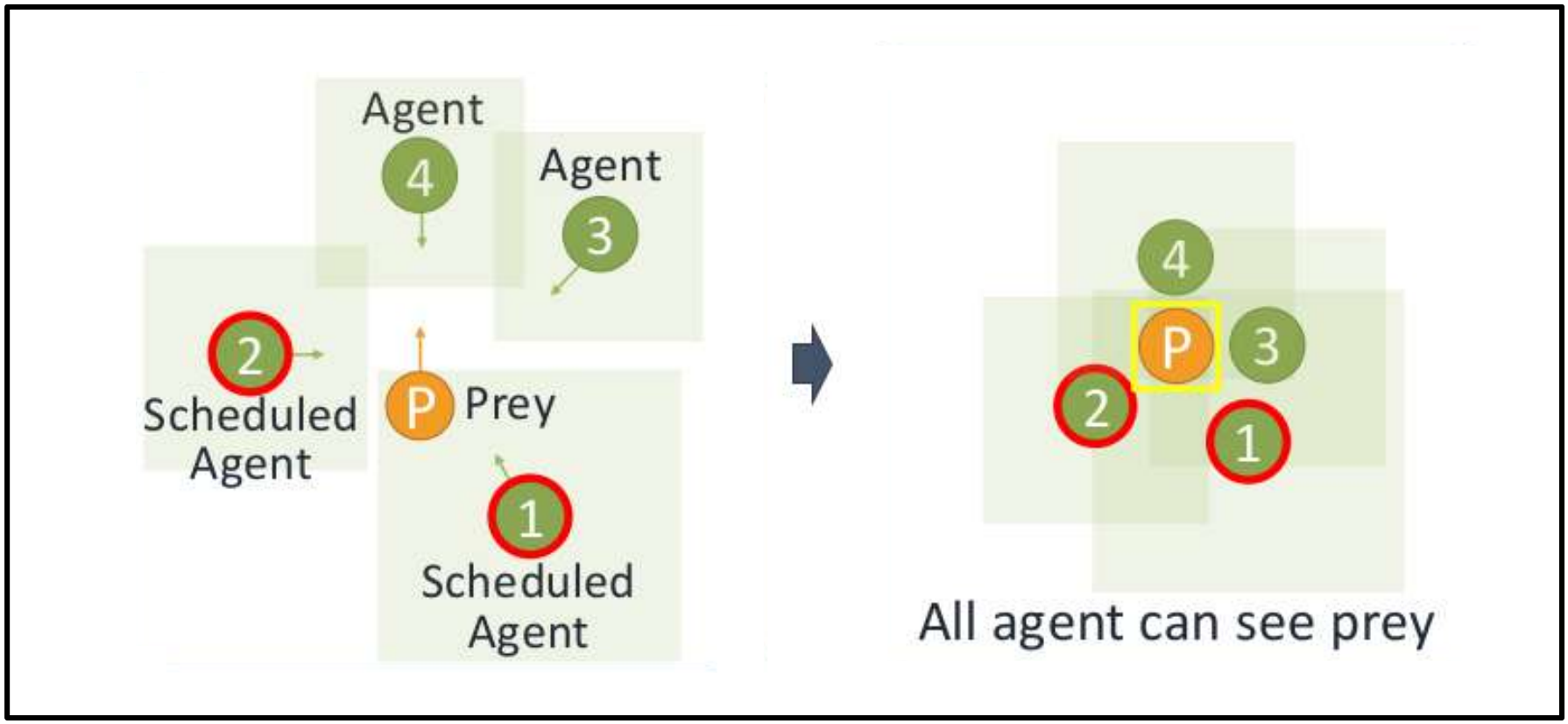
SchedNet: Distributed Execution

- Scheduling Weight Generator
  - Gauges the importance of observation
  - Large weight raises the chance of accessing the channel
    - *e.g.*, Wi-Fi connected devices could be made capable of intelligently accessing the channel
- Encoder
  - Given some observation, compresses it succinctly
- Action Selector
  - Given observation and message from other agent/s, select an action



Evaluation Setup

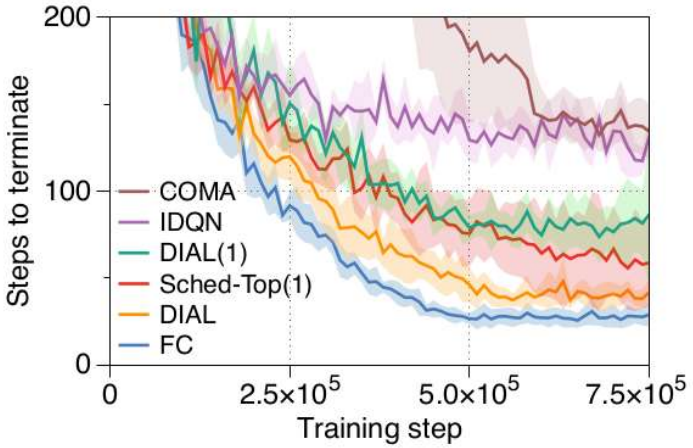
- Predator-prey
  - Multiple predators attempt to catch a randomly moving prey
  - Terminate when the prey is within the observation horizon of all the agents

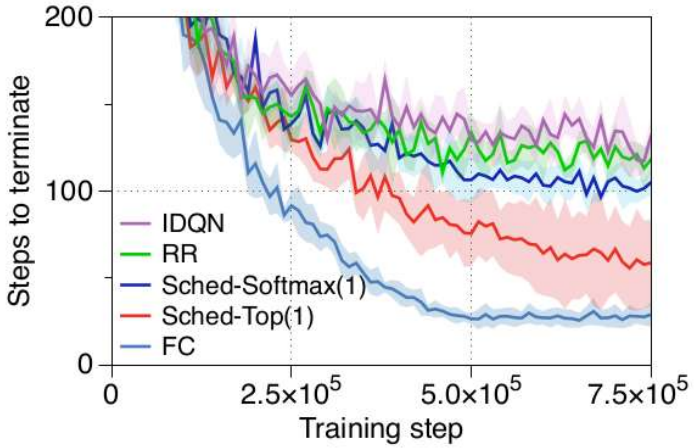


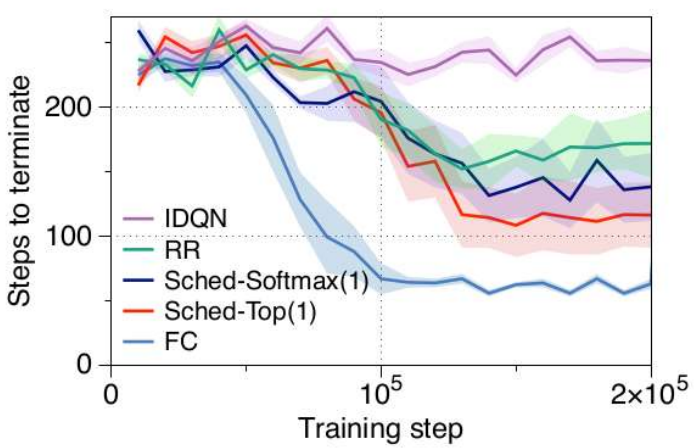
The PP task (left) and its terminating condition (right)

Evaluation Results

- Baselines
  - COMA – no inter-agent communication
  - IDQN – independently trained via Q-learning
  - FC – full communication allowed
  - RR – round-robin scheduling
- SchedNet outperforms most baselines, except
  - DIAL, which ignores medium contention issues and allows all agents to access the channel

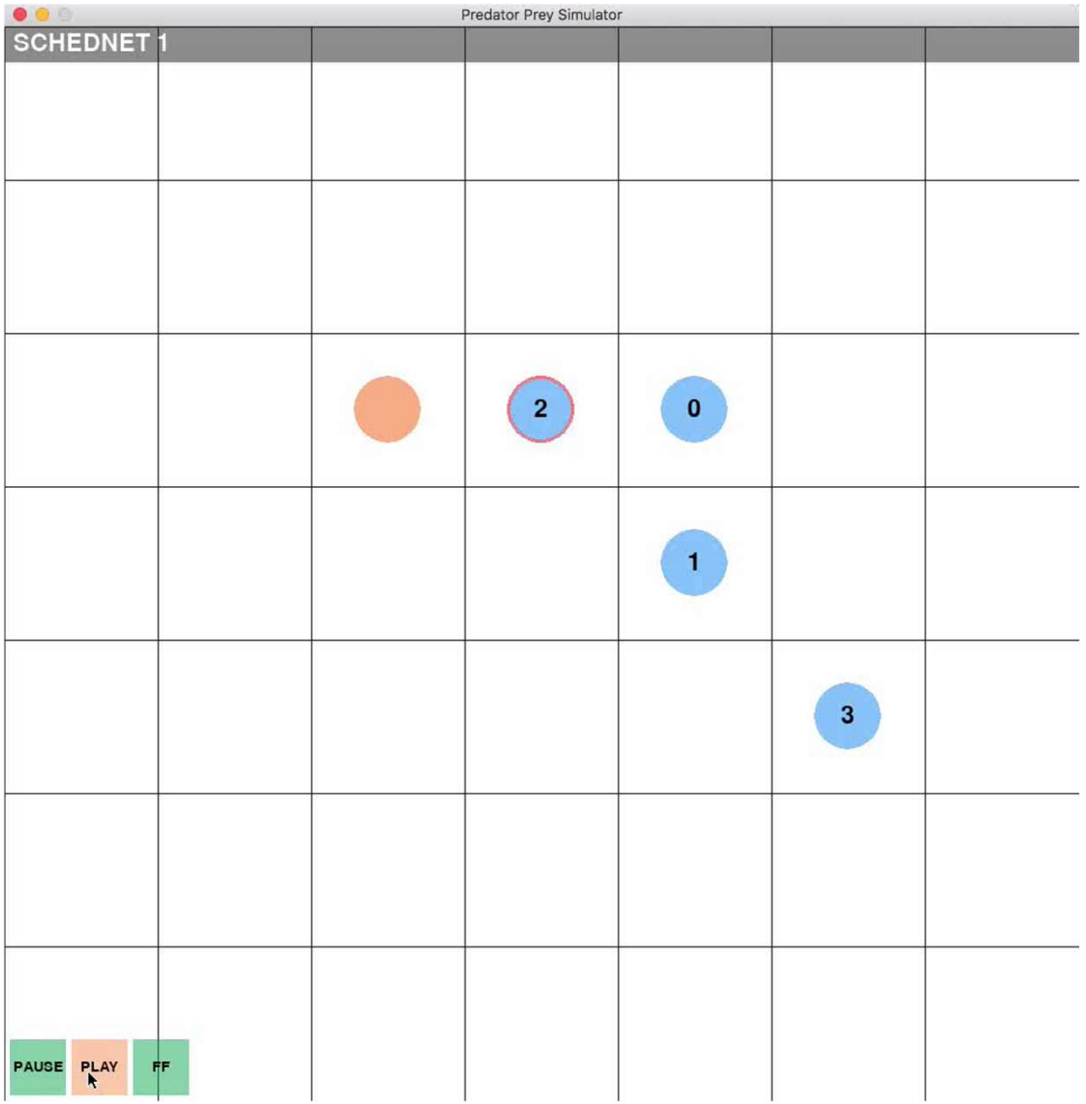






Demonstration

- Blue predators trained for 750k steps
- Orange prey moving according to a uniformly random distribution
- Scheduled predators are circled
  - Messages are transmitted to all other predators
- Predators chase the prey and eventually surround it



Summary and Remarks

- Proposed a new MARL training methodology
- Train multiple agents to take cooperative actions

  Action Selector

  - By exchanging succinct information

    Message Encoder

  - By learning to determine in a distributed manner when to access the channel, based on weights computed to measure the importance of the observations

    Schedule Weight Generator

- Accepted at ICLR 2019

Thank you

Appendix

Coupling of Scheduling and Encoding

- How beneficial was the joint optimization of scheduling and encoding?
- With a pre-trained encoder, agents took a longer time to complete the given task

Average normalized number of steps taken to complete the PP task

FC	SchedNet -Top(1)	Schedule w/ auto-encoder
1	2.030	3.408

*Lower is better

Scheduling in the PP task

- Agent 1 has the widest observation horizon
- Agents 2, 3, 4 have the same observation horizon

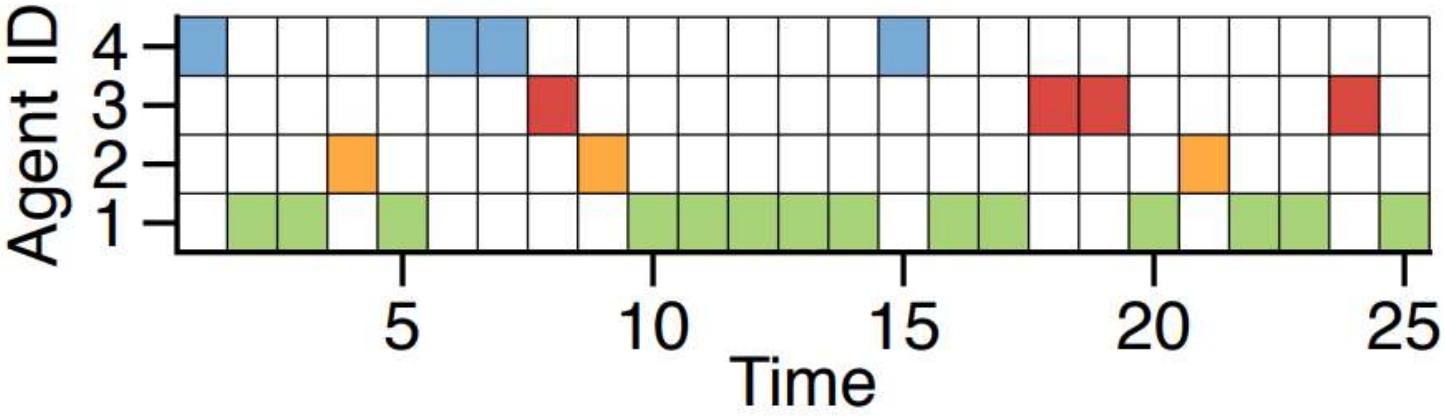


Figure 4: Instances of scheduling results over 25 time steps in PP

Language of the Agents

- 2D projections of the encoded messages
- Upon observing the prey, agents transmit messages with large variance
- This is because they are implicitly embedding some informative content into the outgoing messages

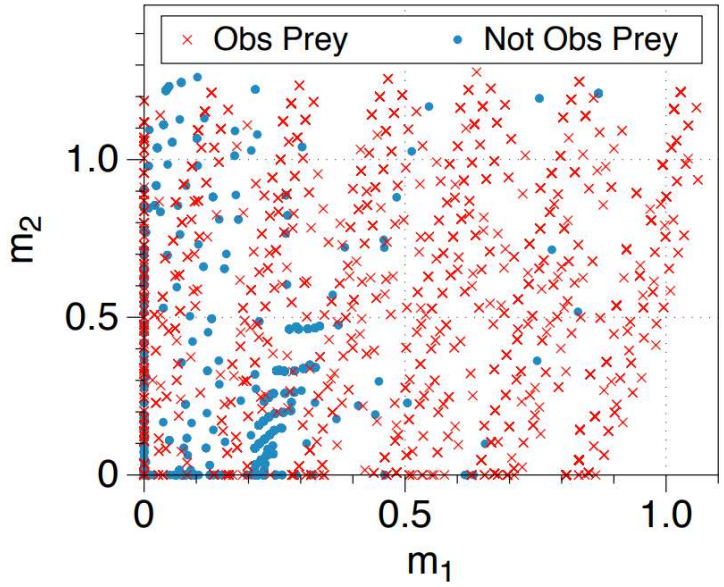


Figure 5: Encoded messages projected onto 2D plane in PP task

Cooperative Communication and Navigation

- Two agents have different observation horizon
- They are start from one state and must reach a goal state
- They are not aware of their own positions, but they are aware of the other agent's position
- They must guide each other to their respective goal states

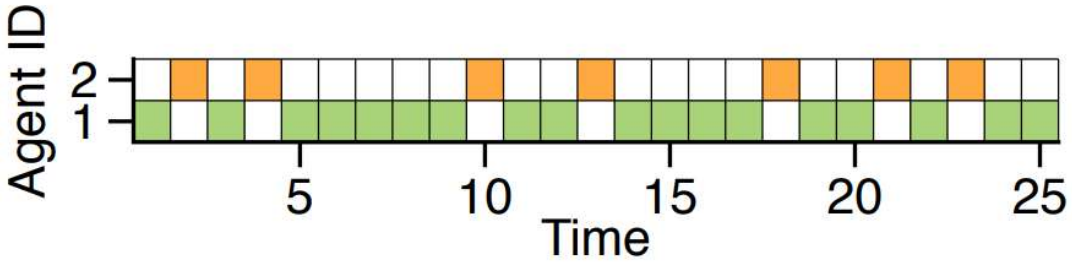


Figure 6: Instances of scheduling results over 25 time steps in CCN