Sample and Communication-Efficient Decentralized Actor-Critic Algorithms

Ziyi Chen **Yi Zhou** **Rong-Rong Chen** **Shaofeng Zou**

Department of ECE
University of Utah

Accepted by ICML 2022.

Cooperative Multi-agent Reinforcement Learning

❖ Cooperative MARL:

$$\underbrace{s_t \xrightarrow{\pi} a_t = \left\{a_t^{(m)}\right\}_{m=1}^{M} \xrightarrow{P} s_{t+1}}_{\displaystyle\downarrow\ \left\{R_t^{(m)}\right\}_{m=1}^{M}}$$

❖ Cooperative goal: Find the optimal policy π_ω that maximizes

$$J(\omega) := (1-\gamma)\mathbb{E}_{\pi_\omega}\left[\sum_{t=0}^{\infty} \gamma^t \left(\frac{1}{M}\sum_{m=1}^{M} R_t^{(m)}\right)\right]$$

❖ Broad applications:

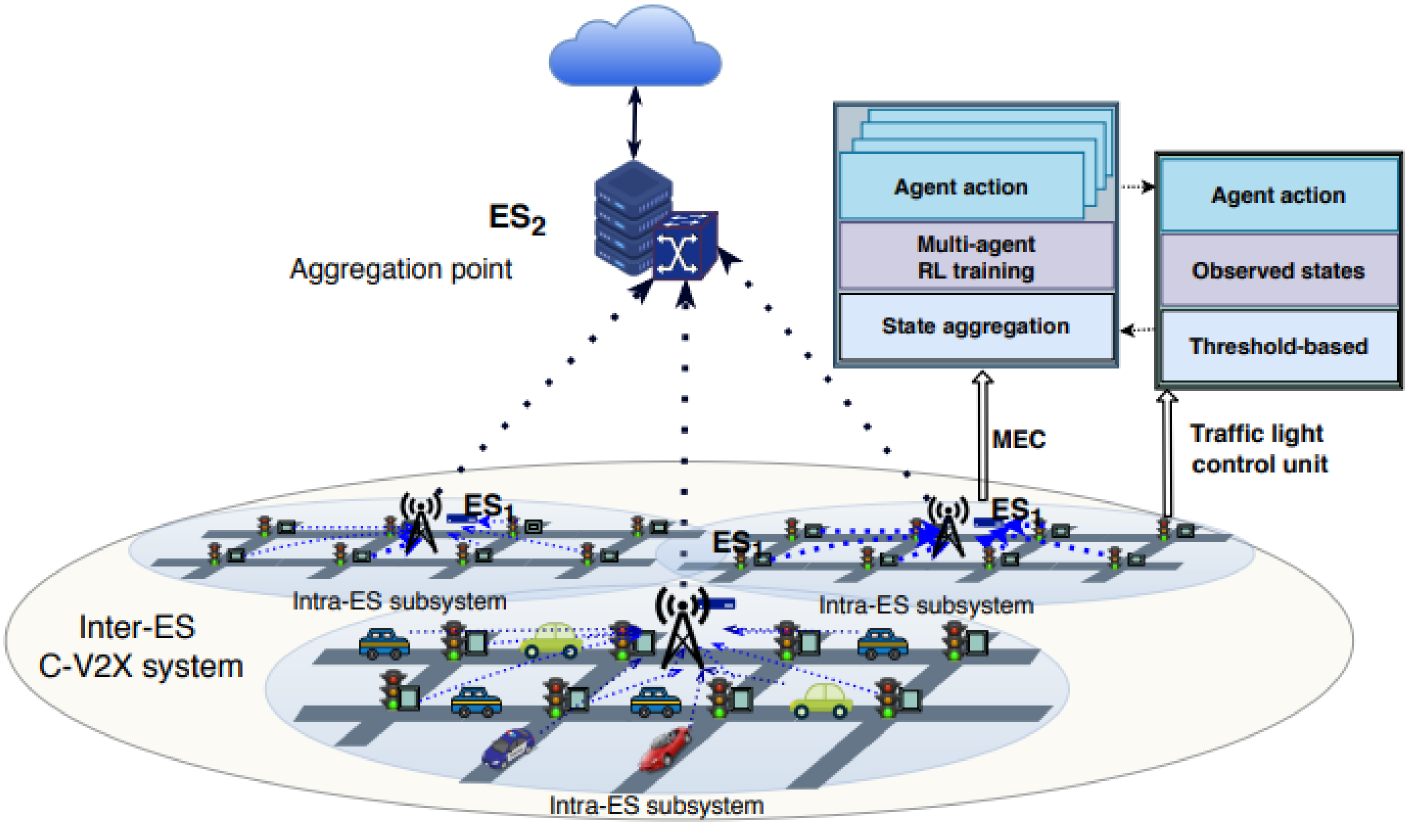
- Traffic signal control
- Dynamic web service composition problem
- Stock market
- ……

Application to Traffic Signal Control

❖ Goal: Multiple agents control traffic signals to maximize [1]

$$J(\omega) \coloneqq (1-\gamma)\mathbb{E}_{\pi_\omega}\left[\sum_{t=0}^{\infty} \gamma^t \left(\frac{1}{M}\sum_{m=1}^{M} R_t^{(m)}\right)\right]$$

where $R_t^{(m)}$: measure of traffic fluency in the m-th lane (agent) at time t.



[1] Zhou, P., Chen, X., Liu, Z., Braud, T., Hui, P., & Kangasharju, J. (2020). DRLE: Decentralized reinforcement learning at the edge for traffic light control in the IoV. *IEEE Transactions on Intelligent Transportation Systems*, *22*(4), 2262-2273.

Decentralized Actor-Critic Algorithms

❖ Limitations of existing decentralized actor-critic algorithms:

- Either agents reveal their **sensitive/private information (actions, rewards)**

OR

Learning parameterized reward is **costly** and inaccurate. [2]

- Lack **finite-time** convergence guarantee.

❖ Our goal: Fully Decentralized Actor-Critic (DAC) algorithm:

- DO NOT share agents' **sensitive/private information (actions, rewards)**.

- **Efficient** **finite-time** convergence, sample complexity and communication complexity.

[2] Zhang, K., Yang, Z., Liu, H., Zhang, T., & Basar, T. (2018, July). Fully decentralized multi-agent reinforcement learning with networked agents. In *International Conference on Machine Learning* (pp. 5872-5881). PMLR.

Our Decentralized Actor-Critic Algorithm

- Stochastic policy gradient ascent (actor-step):

$$\hat{\nabla}_{\omega^{(m)}} J(\omega_t) = \frac{1}{N}\sum_{i=tN}^{(t+1)N-1} \left(\bar{R}_i^{(m)} + \gamma\phi(s_{i+1}')^T\theta_t^{(m)} - \phi(s_i)^T\theta_t^{(m)}\right) \nabla_{\omega^{(m)}} \ln \pi_t^{(m)}(a_i^{(m)}|s_i)$$

$$\omega_{t+1} = \omega_t + \alpha\hat{\nabla}_{\omega^{(m)}} J(\omega_t)$$

- Decentralized TD for policy evaluation (critic-step):

$$V(s) \approx \phi(s)^T\theta$$

- Estimate $\bar{R}_i^{(m)} \approx \frac{1}{M}\sum_{m=1}^{M} R_i^{(m)}$:
  - $\tilde{R}_i^{(m)} \approx R_i^{(m)}(1 + e_i^{(m)})$ (**noise** $e_i^{(m)} \sim N(0, \sigma^2)$ **to protect sensitive information**).
  - Obtain agents' average via decentralized local averaging (gossip).

- **Minibatch** $i = tN, \cdots, (t+1)N - 1$ at iteration t:
  - Reduce noise variance by $\frac{1}{N}$.
  - Reduce communication frequency and complexity.

Our Decentralized Natural Actor-Critic Algorithm

❖ We proposed the **first fully decentralized** natural actor-critic (NAC) algorithm for cooperative multi-agent reinforcement learning (MARL).

❖ Key challenge: Computing the natural policy gradient $h(\omega) = F(\omega)^{-1}\nabla J(\omega)$

- is costly
- involves sensitive information of all agents

❖ Our solution: Decentralized SGD to obtain

$$h(\omega) = \underset{h}{\operatorname{argmin}} \left[\frac{1}{2} h^T F(\omega) h - \nabla J(\omega)^T h\right]$$

❖ Actor-step:

$$\omega_{t+1} = \omega_t + \alpha \hat{h}(\omega_t)$$

Theoretical Results

- The **first finite-time** complexity results of decentralized AC/NAC.

	Sample complexity	Communication complexity	Target
AC	$O(\epsilon^{-2}\ln\epsilon^{-1})$	$O(\epsilon^{-1}\ln\epsilon^{-1})$	$\mathbb{E}[\|\nabla J(\omega)\|^2]\le\epsilon$
NAC	$O(\epsilon^{-3}\ln\epsilon^{-1})$	$O(\epsilon^{-1}\ln\epsilon^{-1})$	$\mathbb{E}[J(\omega^*)-J(\omega)]\le\epsilon$

- Match state-of-the-arts for counterparts of centralized AC/NAC for singe-agent RL.

Thank You