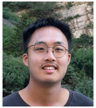
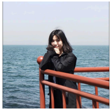
Causality Meets Locality: Provably Generalizable and Scalable Policy Learning for Networked Systems

Hao Liang

King's College London

Joint work with

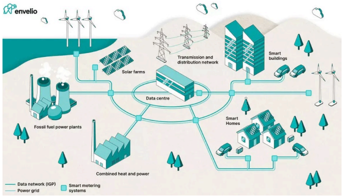




Shuqing Shi (KCL) | **Yudi Zhang** (TU/e) | **Biwei Huang** (UCSD) | **Yali Du** (KCL)

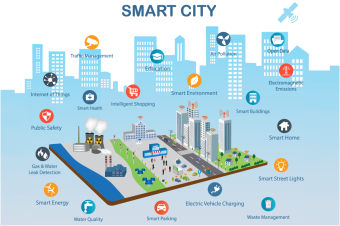
H. Liang*, S. Shi*, Y. Zhang, B. Huang, Y. Du. "Causality Meets Locality: Provably Generalizable and Scalable Policy Learning for Networked Systems." NeurIPS 2025 (Spotlight).

Motivation: Real-World Networked Systems



(a) Grid

(b) Transportation

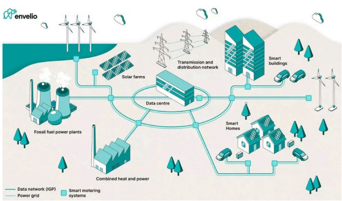


(c) Smart city

Key challenges

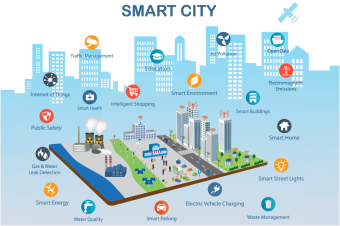
- **Scalability**: Exponential state-action space growth

Motivation: Real-World Networked Systems



(a) Grid

(b) Transportation

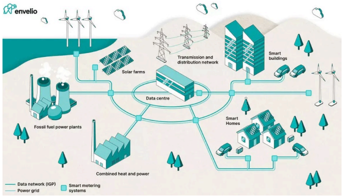


(c) Smart city

Key challenges

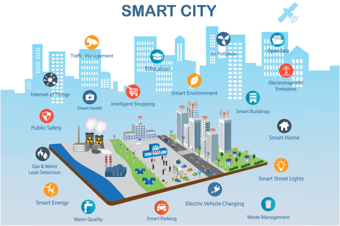
- **Scalability**: Exponential state-action space growth
- **Environment changes**: traffic patterns change, user demands vary

Motivation: Real-World Networked Systems



(a) Grid

(b) Transportation



(c) Smart city

Key challenges

- **Scalability**: Exponential state-action space growth
- **Environment changes**: traffic patterns change, user demands vary

Current methods either scale OR generalize, but rarely both

Central Research Problem

Can we design a **provably** **generalizable** AND **scalable** MARL algorithm for networked systems?

Our answer: Yes!

Generalizable and **S**calable **A**ctor-**C**ritic (GSAC)

Central Research Problem

Can we design a **provably** **generalizable** AND **scalable** MARL algorithm for networked systems?

Our answer: Yes!

Generalizable and **S**calable **A**ctor-**C**ritic (GSAC)

Key insights

1) **Locality + Causality** $\rightarrow$ Scalability

- **Locality**: Exploit local structure
- **Causality**: Identify minimal relevant features

Central Research Problem

Can we design a **provably** **generalizable** AND **scalable** MARL algorithm for networked systems?

Our answer: Yes!

Generalizable and **S**calable **A**ctor-**C**ritic (GSAC)

Key insights

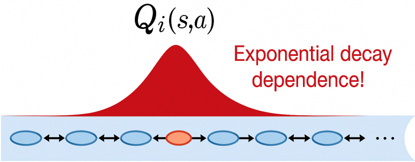
1) **Locality + Causality** $\to$ Scalability

- **Locality**: Exploit local structure
- **Causality**: Identify minimal relevant features

2) Meta-training $\to$ Generalization

Challenge 1: Scalability

- Curse of dimensionality $\#(\mathbf{s}, \mathbf{a}) = |\mathcal{S}_i|^n \times |\mathcal{A}_i|^n$



- κ**-hop truncation** as efficient approximation [QWL22][1]

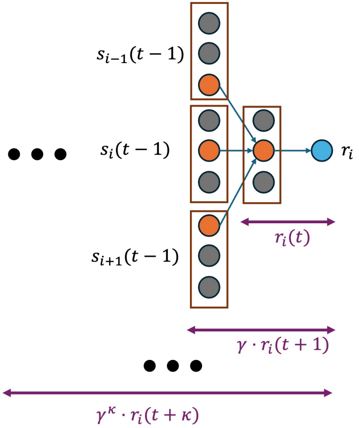
$$\left| Q_i^\pi(\mathbf{s}_{\mathcal{N}_i^\kappa}, \mathbf{a}_{\mathcal{N}_i^\kappa}) - Q_i^\pi(\mathbf{s}, \mathbf{a}) \right| \leq \mathcal{O}(\gamma^\kappa)$$

[1] Guannan Qu, Adam Wierman, Na Li. ”Scalable reinforcement learning for multiagent networked systems.” *Operations Research*, 70(6): 3601–3628, 2022.

More Scalable via ACRs

Core idea: Recursively Identify minimal variables within $\mathbf{s}_{\mathcal{N}_i^\kappa}$ that influence κ-step rewards

Output: $\mathbf{s}^{\circ}_{\mathcal{N}^\kappa}$



Benefits of ACR

Method	Input	Dimension	Approx. Error	Size
Full State	$\mathbf{s}$	$\sum_{j=1}^{n} d_j^s$	0	all agents ✗
Truncation [QWL22]	$\mathbf{s}_{\mathcal{N}_i^\kappa}$	$\sum_{j \in \mathcal{N}_i^\kappa} d_j^s$	$\mathcal{O}(\gamma^\kappa)$	κ-hop neighbors ▲
GSAC (ACR)	$\mathbf{s}^\circ_{\mathcal{N}_i^\kappa}$	$\leq \sum_{j \in \mathcal{N}_i^\kappa} d_j^s$	$\mathcal{O}(\gamma^\kappa)$	**Much Lower ✓**

Challenge 2: Generalizability

Key idea: **Meta-training** [HFL$^+$22][2]

Training: learn domain-conditioned local policy $\pi_i^{\hat{\theta}_i}(\mathbf{a}_i \mid \mathbf{s}_{\mathcal{N}_i}, \omega_{\mathcal{N}_i})$

Adaptation: few-shot estimate $\hat{\boldsymbol{\omega}}'$, deploy $\pi_i^{\hat{\theta}_i}(\mathbf{a}_i \mid \mathbf{s}_{\mathcal{N}_i}, \hat{\omega}'_{\mathcal{N}_i})$

All computation uses compact ACR inputs!

[2] Biwei Huang, Fan Feng, Chaochao Lu, Sara Magliacane, Kun Zhang, "Adarl: What, where, and how to adapt in transfer reinforcement learning". ICLR 2022.

Convergence

Theorem (Critic error bound)

With high probability, after T inner iterations:

$$\textit{Critic error} \leq \mathcal{O}\left(\underbrace{\frac{1}{\sqrt{T+t_0}}}_{\textit{TD error}} + \underbrace{\frac{2c\rho^{\kappa+1}}{(1-\gamma)^2}}_{\textit{ACR error}} + \underbrace{\sqrt{\frac{1}{T_e}}}_{\textit{Domain estimation error}} \right).$$

Theorem (Policy gradient convergence)

$$\textit{Policy optimization error} \leq \tilde{\mathcal{O}}\left(\underbrace{\frac{1}{\sqrt{K+1}}}_{\textit{optimization error}} + \rho^{\kappa+1} + \sqrt{\frac{1}{T_e}} + \underbrace{\sqrt{\frac{1}{M}}}_{\textit{domain generalization}} \right).$$

Adaptation Guarantee

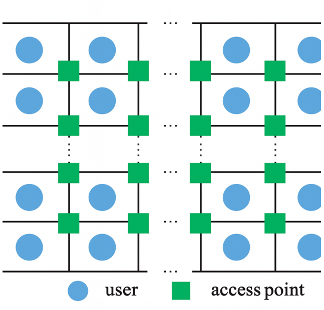
Theorem (Adaptation gap)

For new domain, after collecting T_a adaptation trajectories:

$$\textit{Adaptation gap} \geq \Theta\left(\frac{1}{T_a}\right).$$

Meta-training provides good initialization/zero-shot performance!

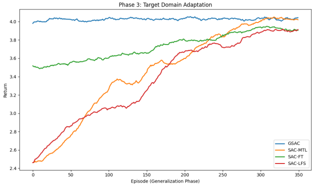
Empirical Validations



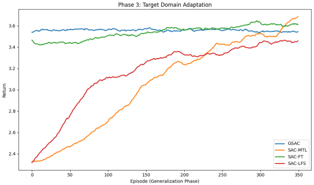
Wireless communication network [Zoc19][3]

[3] Alessandro Zocca. "Temporal starvation in multi-channel csma networks: an analytical framework." ACM SIGMETRICS Performance Evaluation Review (2019).

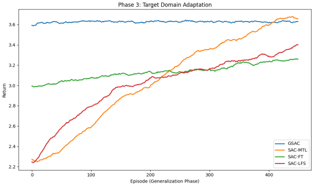
Adaptation Performance



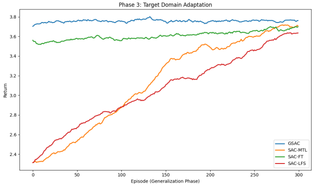
(a) grid size 3 (16 agents)



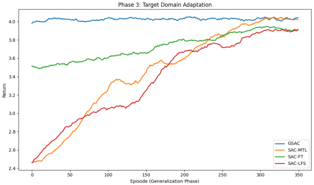
(b) grid size 4 (25 agents)



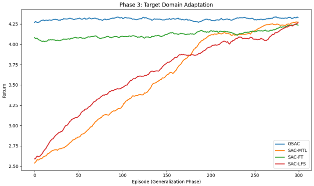
(c) grid size 5 (36 agents)



(a) $p_{\text{target}} = 0.6$



(b) $p_{\text{target}} = 0.65$



(c) $p_{\text{target}} = 0.7$

Thank you!

References I

- Biwei Huang, Fan Feng, Chaochao Lu, Sara Magliacane, and Kun Zhang, *Adarl: What, where, and how to adapt in transfer reinforcement learning*, International Conference on Learning Representations, 2022.
- Guannan Qu, Adam Wierman, and Na Li, *Scalable reinforcement learning for multiagent networked systems*, Operations Research **70** (2022), no. 6, 3601–3628.
- Alessandro Zocca, *Temporal starvation in multi-channel csma networks: an analytical framework*, ACM SIGMETRICS Performance Evaluation Review **46** (2019), no. 3, 52–53.