

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).

The diagram illustrates the Envero smart grid system. At the top left, the Envero logo is shown. The system is a central hub-and-spoke model. The central hub is a blue circle labeled "Data centre". Radiating from this hub are several lines representing different energy and data flows. On the left, a "Fossil fuel power plants" (industrial facility) and "Renewable power plants" (wind turbines) are connected to the grid. Above the hub, "Solar farms" (panels) and a "Transmission and distribution network" (power lines) are shown. To the right, "Smart buildings" (tall office blocks) and "Smart homes" (residential houses) are connected. Below the hub, "Combined heat and power" (industrial facility) is shown. A legend at the bottom identifies the symbols: a blue line for "Data network (DPI)", a green line for "Power grid", and a blue square with a white dot for "Smart metering systems".

A vibrant, isometric illustration of a smart city intersection. The scene is filled with various vehicles, including a red truck, a white car, a yellow taxi, a black car, a white bus, and a red car. Pedestrians are walking across the crosswalks. Yellow concentric circles around the vehicles represent their communication range. A green pedestrian signal is visible on the right side of the intersection.

Key challenges

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

The first illustration on the left, titled 'envio', depicts a smart energy network. It shows a central 'Data centre' connected to various energy sources: 'Fossil fuel power plants', 'Solar farms', and 'Wind turbines'. The network also includes 'Transmission and distribution network' lines, 'Smart buildings', 'Smart homes', and 'Combined heat and power' units. A legend at the bottom identifies the 'Data network (SDN)' in blue and the 'Power grid' in green.

The middle illustration shows a smart transportation scene. It features a busy street with 'Autonomous vehicles', 'Smart traffic lights', and 'Smart parking' spaces. Pedestrians and cyclists are also shown interacting with the smart infrastructure.

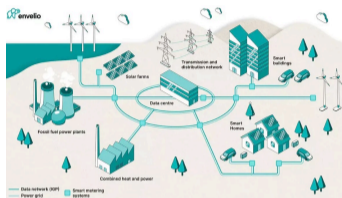
The third illustration on the right, titled 'SMART CITY', shows a cityscape with various smart services. These include 'Waste Management', 'Public Safety', 'Smart Energy', 'Water Quality', 'Smart Parking', 'Electric Vehicle Charging', 'Waste Management', 'Smart Buildings', 'Smart Homes', 'Smart Street Lights', 'Smart Environment', 'Intelligent Shopping', 'Smart Health', 'Public Safety', 'Gas & Water Leak Detection', 'Smart Energy', 'Water Quality', 'Smart Parking', 'Electric Vehicle Charging', and 'Waste Management'. A legend at the bottom identifies the 'Data network (SDN)' in blue and the 'Power grid' in green.

(b) Transportation

(c) Smart city

- **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 Scalable Actor-Critic (GSAC)

Central Research Problem

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

Our answer: Yes!

Generalizable and Scalable Actor-Critic (GSAC)

Key insights

1) **Locality + Causality** → 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 Scalable Actor-Critic (GSAC)

Key insights

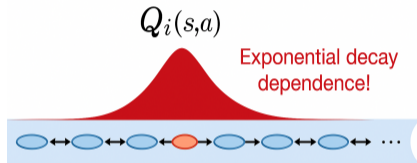
1) **Locality + Causality** → Scalability

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

2) Meta-training → 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

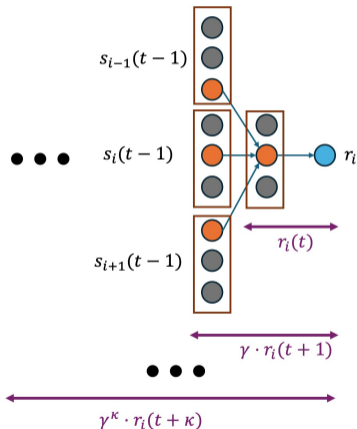
$$|Q_i^\pi(\mathbf{s}_{\mathcal{N}_i^\kappa}, \mathbf{a}_{\mathcal{N}_i^\kappa}) - Q_i^\pi(\mathbf{s}, \mathbf{a})| \leq \mathcal{O}(\gamma^\kappa)$$

¹Guannan 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: $\mathbf{s}_{\mathcal{N}_i^\kappa}^\circ$



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)	$\mathbf{s}_{\mathcal{N}_i^\kappa}^o$	$\leq \sum_{j \in \mathcal{N}_i^\kappa} d_j^s$	$\mathcal{O}(\gamma^\kappa)$	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 $\hat{\omega}'$, 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!

²Biwei 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:

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

Theorem (Policy gradient convergence)

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

Adaptation Guarantee

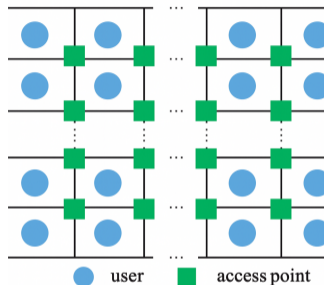
Theorem (Adaptation gap)

For new domain, after collecting T_a adaptation trajectories:

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

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

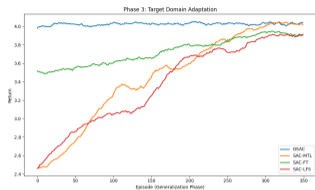
Empirical Validations



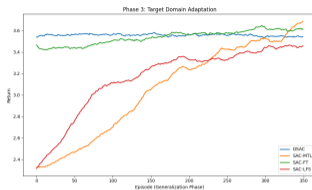
Wireless communication network [Zoc19]³

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

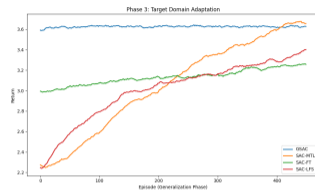
Adaptation Performance



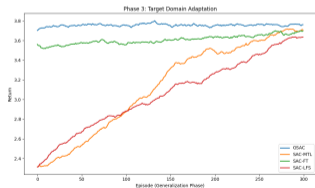
(a) grid size 3 (16 agents)



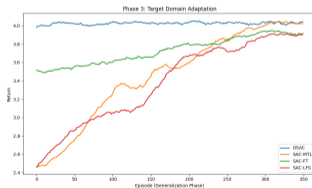
(b) grid size 4 (25 agents)



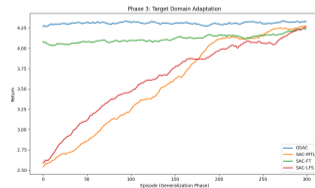
(c) grid size 5 (36 agents)



(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.