SC '24, Nov 17–22, 2024, Atlanta, USA

★ Best Student Paper Finalist ★

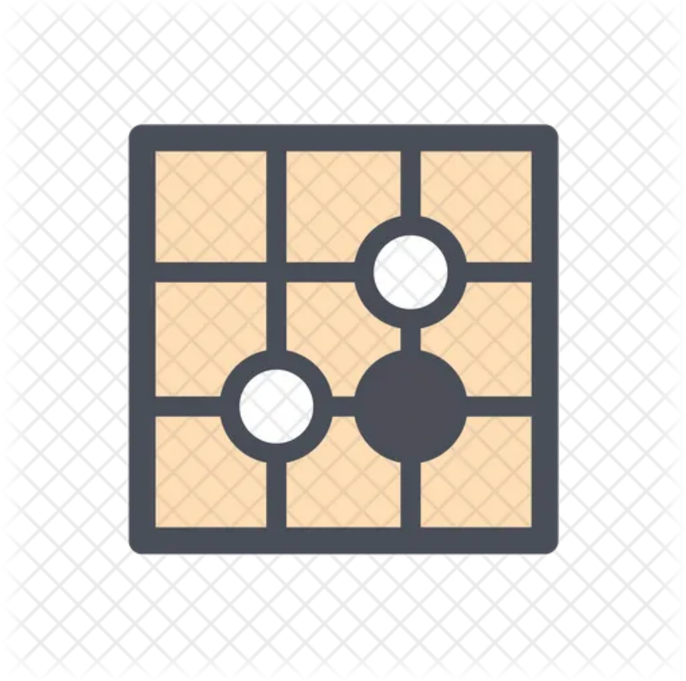
Stellaris: Staleness-Aware Distributed Reinforcement Learning with Serverless Computing

Hanfei Yu[1], Hao Wang[1], Devesh Tiwari[2], Jian Li[3], Seung-Jong Park[4]

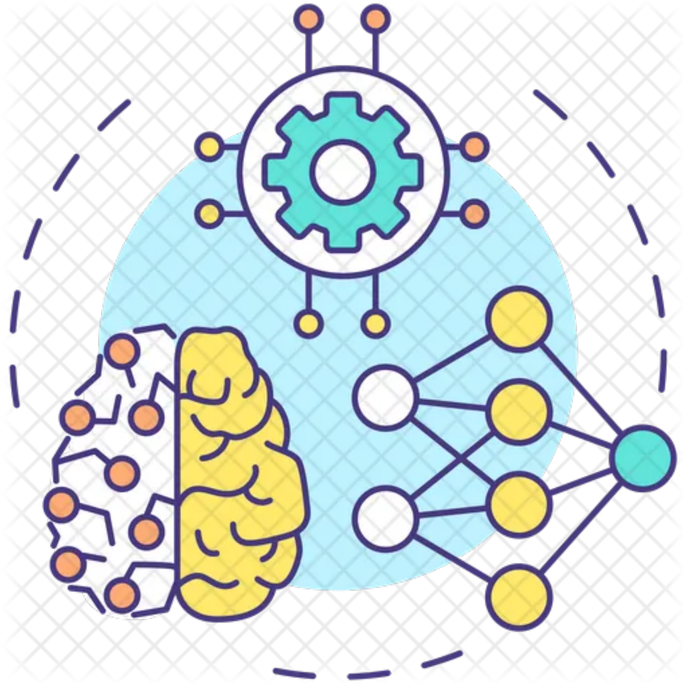
Stevens Institute of Technology[1], Northeastern University[2], Stony Brook University[3], Missouri University of Science & Technology[4]

IntelliSys Lab

Deep Reinforcement Learning (DRL)



AlphaGo
Gaming AIs



Large Language Models
RL from Human Feedback

HPC System Scheduling

Training DRL Agent in Environments

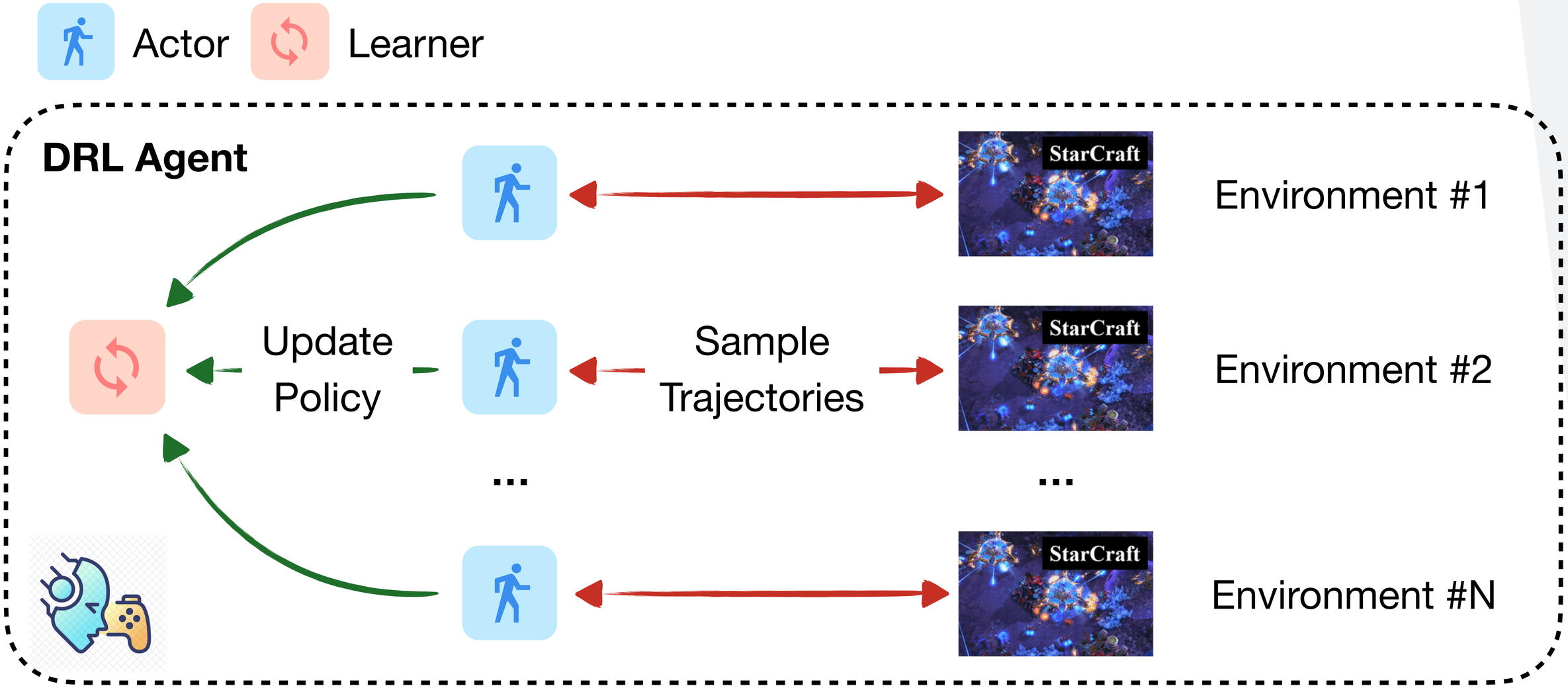
DRL Agent

State
&
Reward

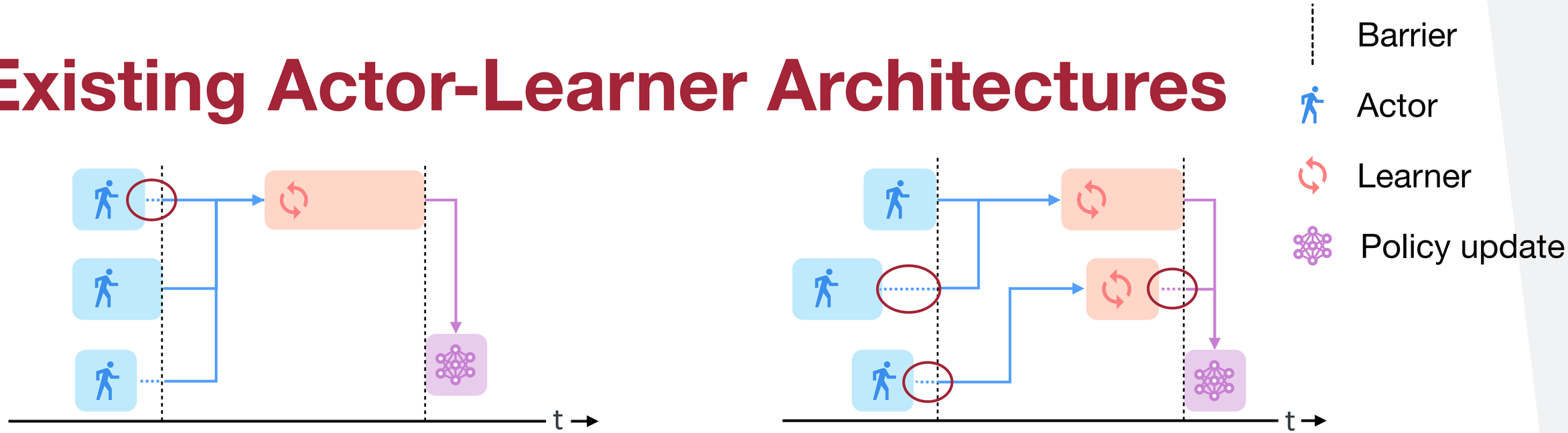
Action

Environment

Scaling DRL Training: Actor-Learner

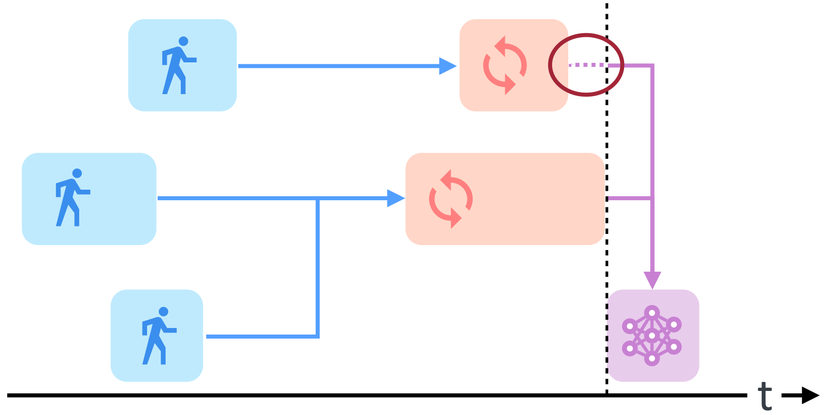


Existing Actor-Learner Architectures

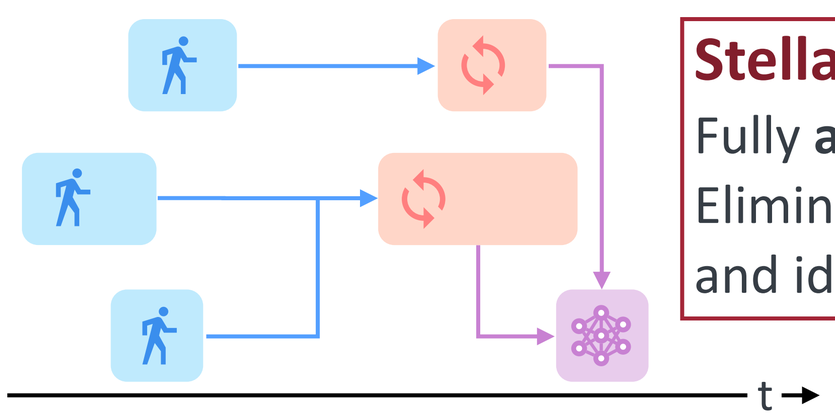


Synchronous Actors + Centralized **Sync**. Learner

Synchronous Actors + **Sync**hronous Learners



Asynchronous Actors + **Sync**hronous Learners



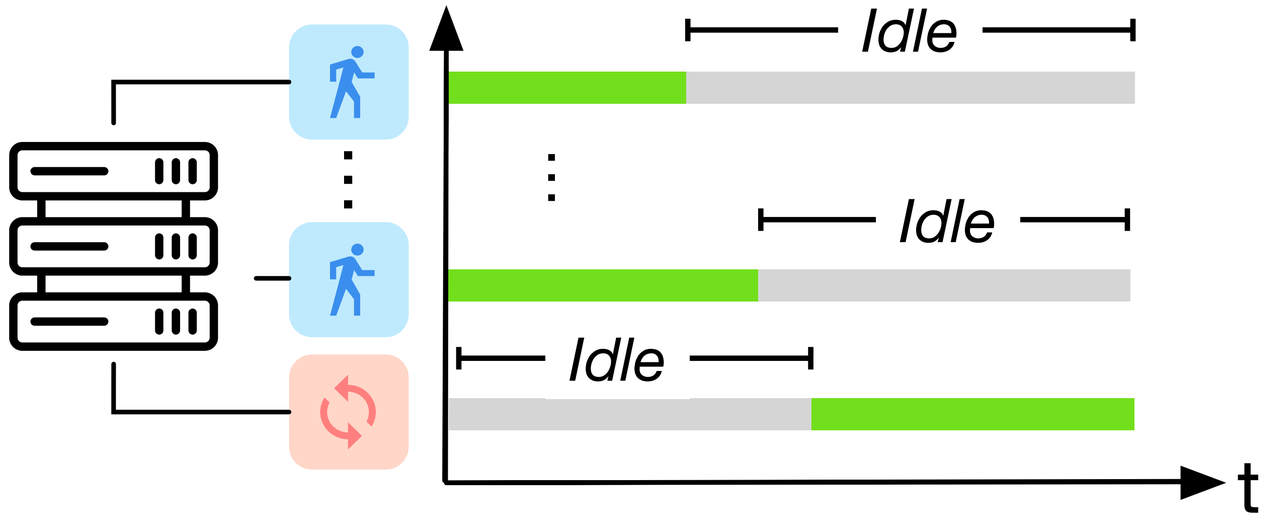
(**A**)**sync**hronous Actors + **Async**hronous Learners

Stellaris

Fully **async**hronous

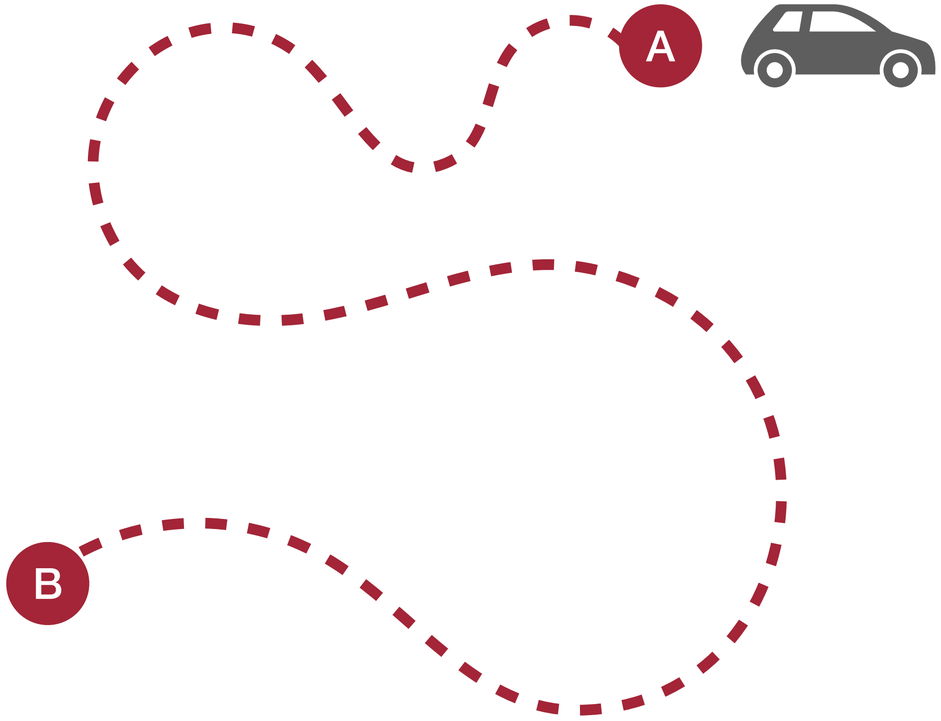
Eliminate blocking and idle waste

Dynamic Usage in Serverful DRL Training



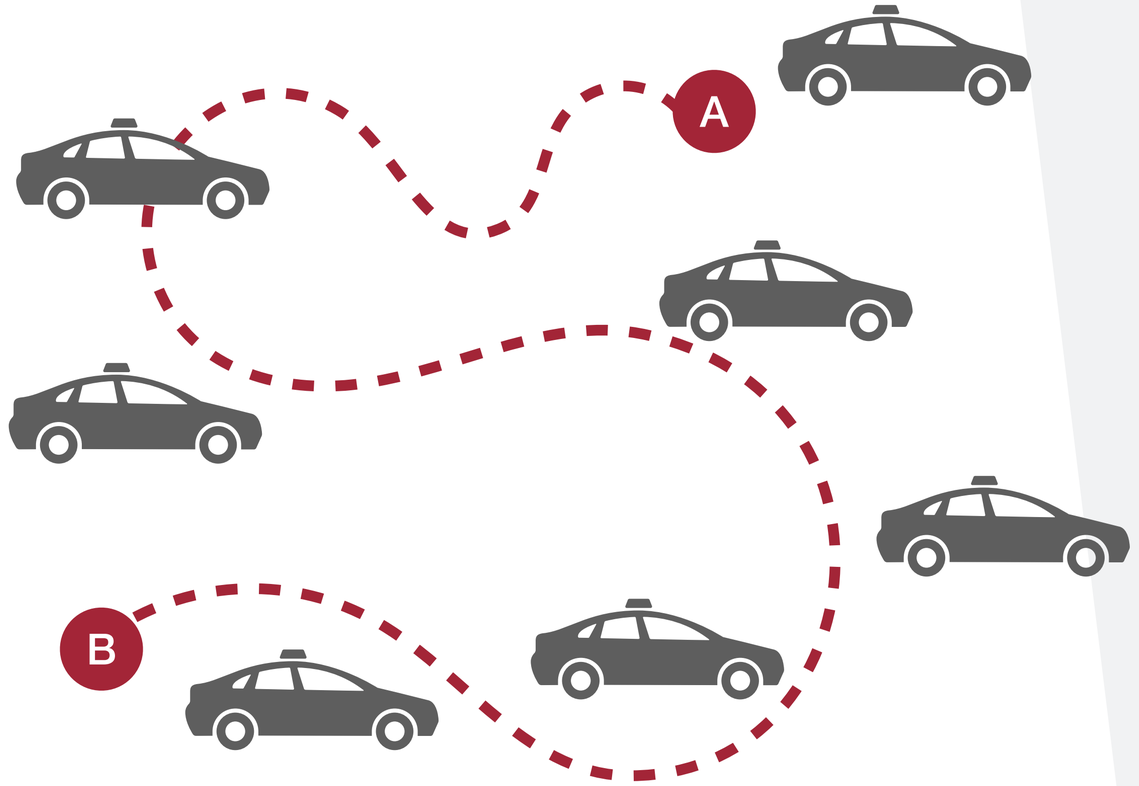
Can DRL training go **serverless**? Yes!

Cloud / HPC



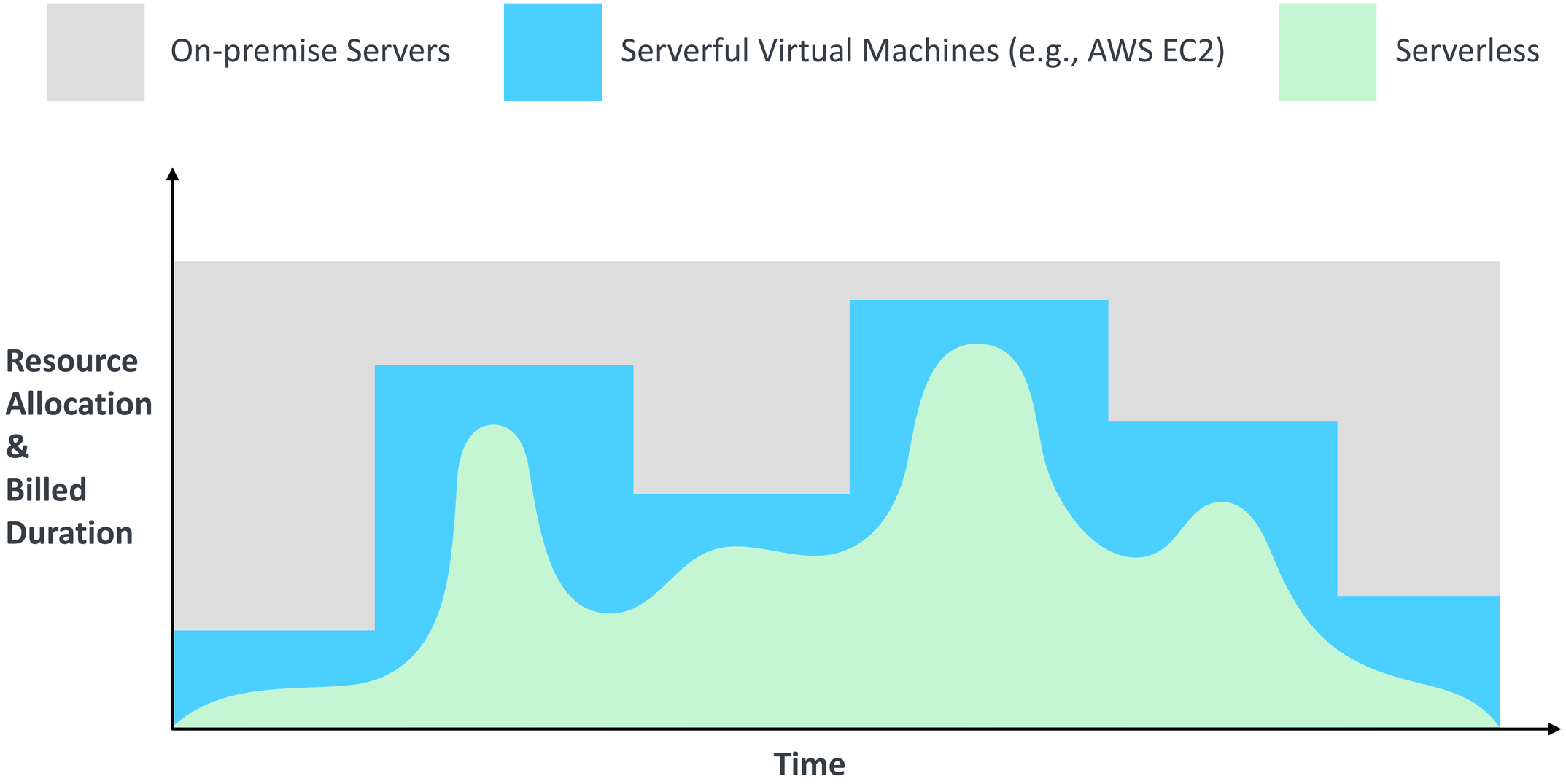
Car rental

Serverless



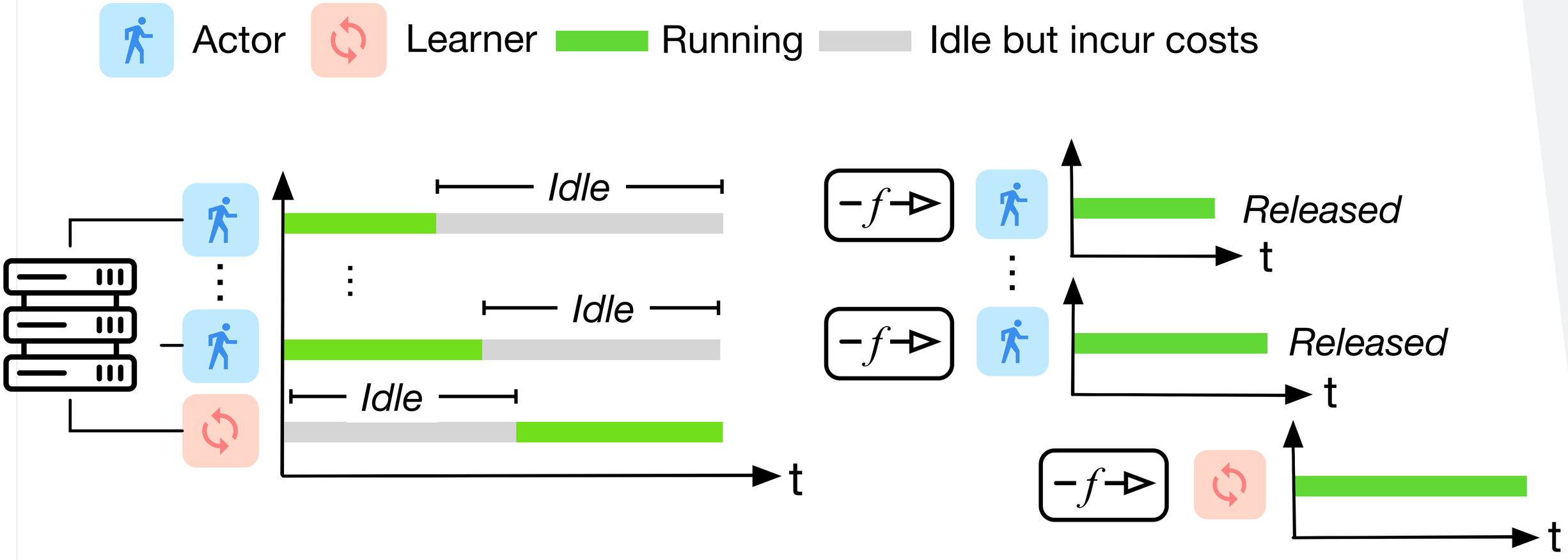
***Cruise* (Self-driving Taxi)**

From *A Serverless Vision for Cloud Computing* by Prof. Ana Klimovic



Schleier-Smith, Johann, et al. "What serverless computing is and should become: The next phase of cloud computing." Communications of the ACM 64.5 (2021): 76-84.

Serverful vs. Serverless DRL Training

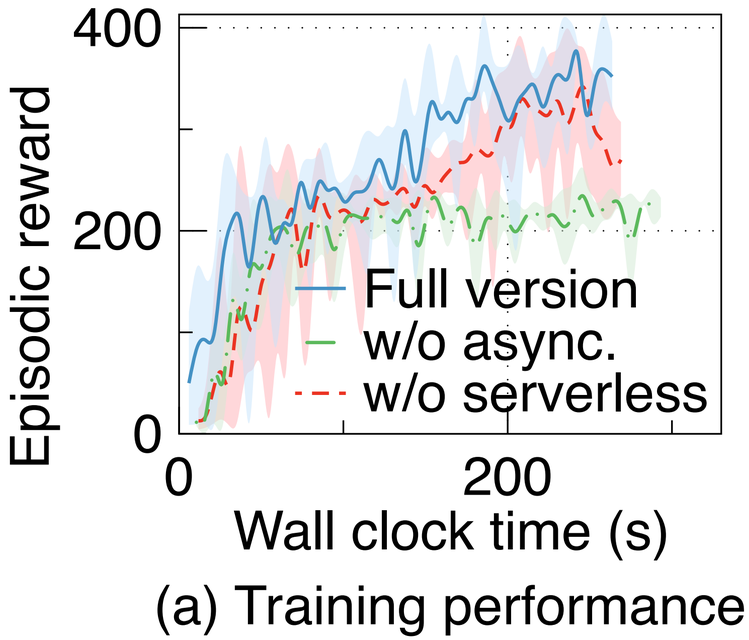


Serverful Training

Serverless Training

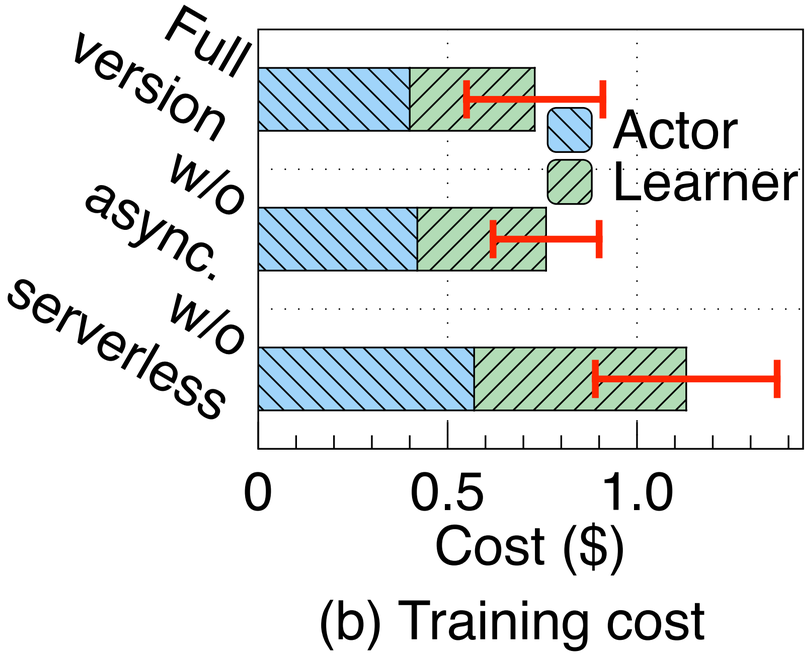
Asynchronous and Serverless DRL Training

High training efficiency



(a) Training performance

Low training cost



(b) Training cost

Proximal Policy Optimization (PPO) on MuJoCo Hopper-v4

Existing Works

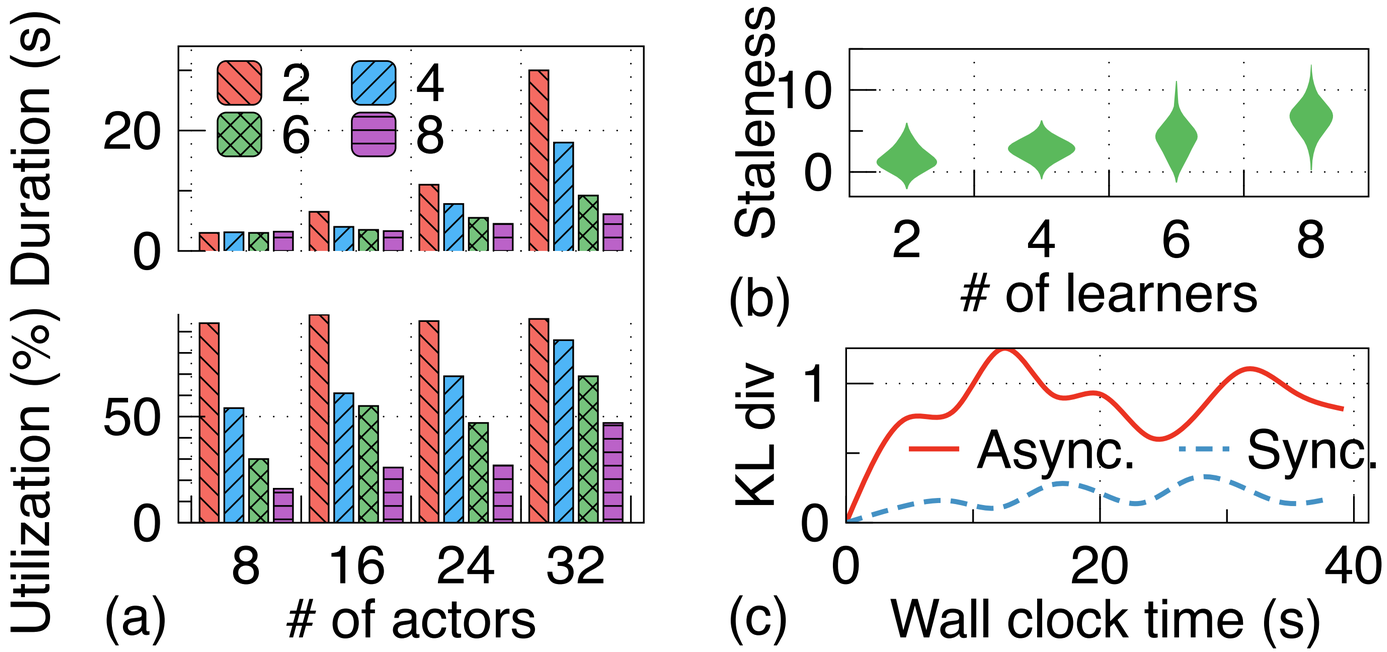
Framework	Asynchronous Learners	Scalable Actors	On-policy and Off-policy	Serverless
Ray RLlib (ICML 2018)	✗	✗	✓	✗
MSRL (ATC 2023)	✗	✗	✓	✗
SEED RL (ICLR 2020)	✗	✗	✓	✗
SRL (ICLR 2024)	✗	✗	✗	✗
MinionsRL (AAAI 2024)	✗	✓	✗	✓
Stellaris (SC 24)	✓	✓	✓	✓

Challenges

Dynamic learner orchestration

Dynamic staleness

Unstable policy updates



KL div: Kullback–Leibler divergence

Design Goals

Dynamic learner orchestration

On-Demand Serverless Learners

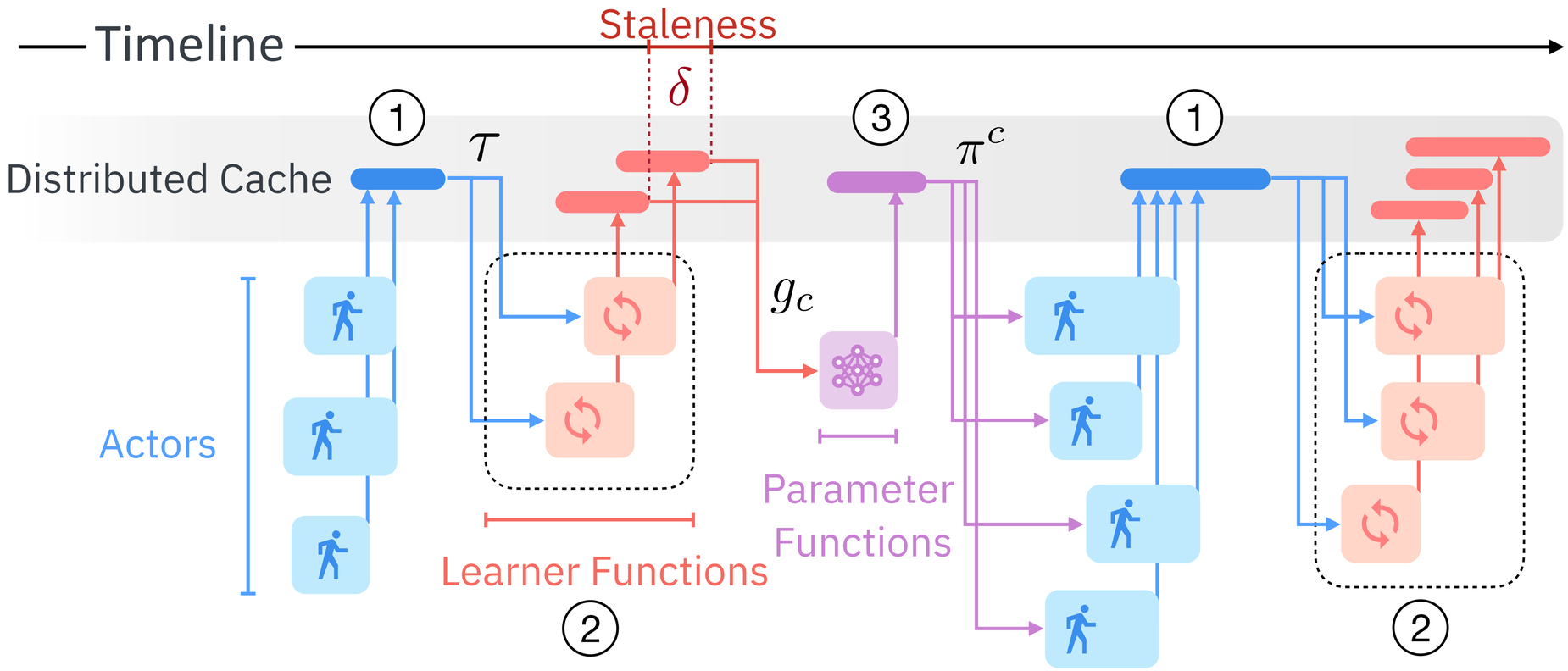
Dynamic staleness

Staleness-Aware Gradient Aggregation

Unstable policy updates

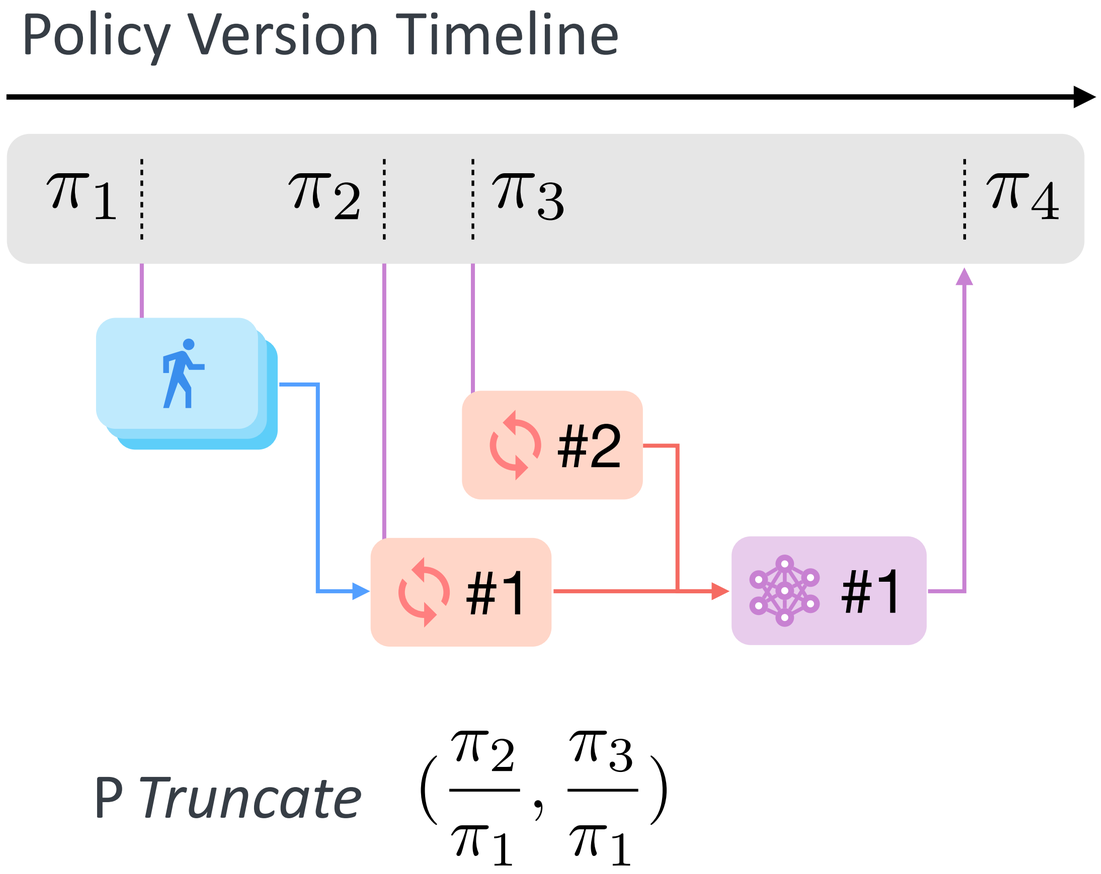
Global Importance Sampling Truncation

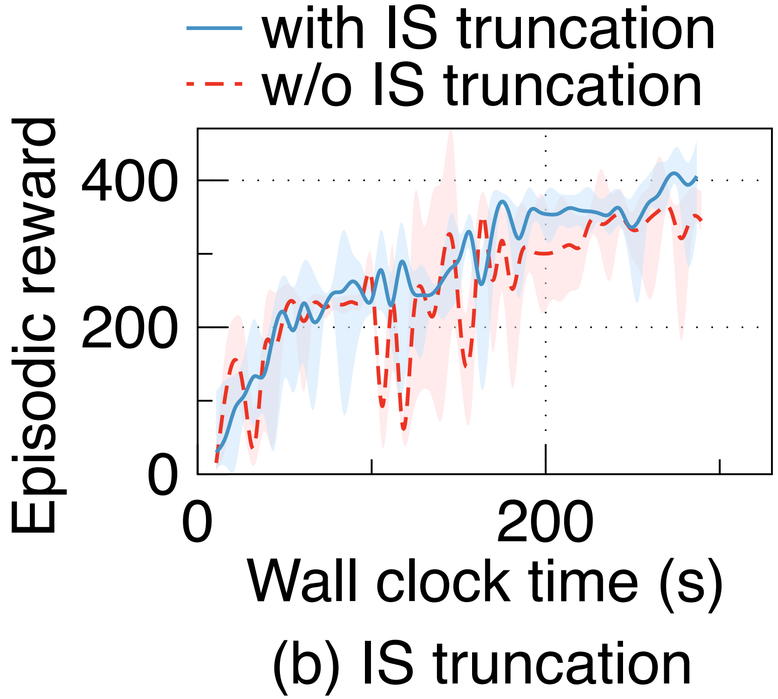
Stellaris



1. Importance sampling driven trajectory collection
2. On-demand gradient calculation
3. Staleness-aware gradient aggregation

Global Importance Sampling Truncation

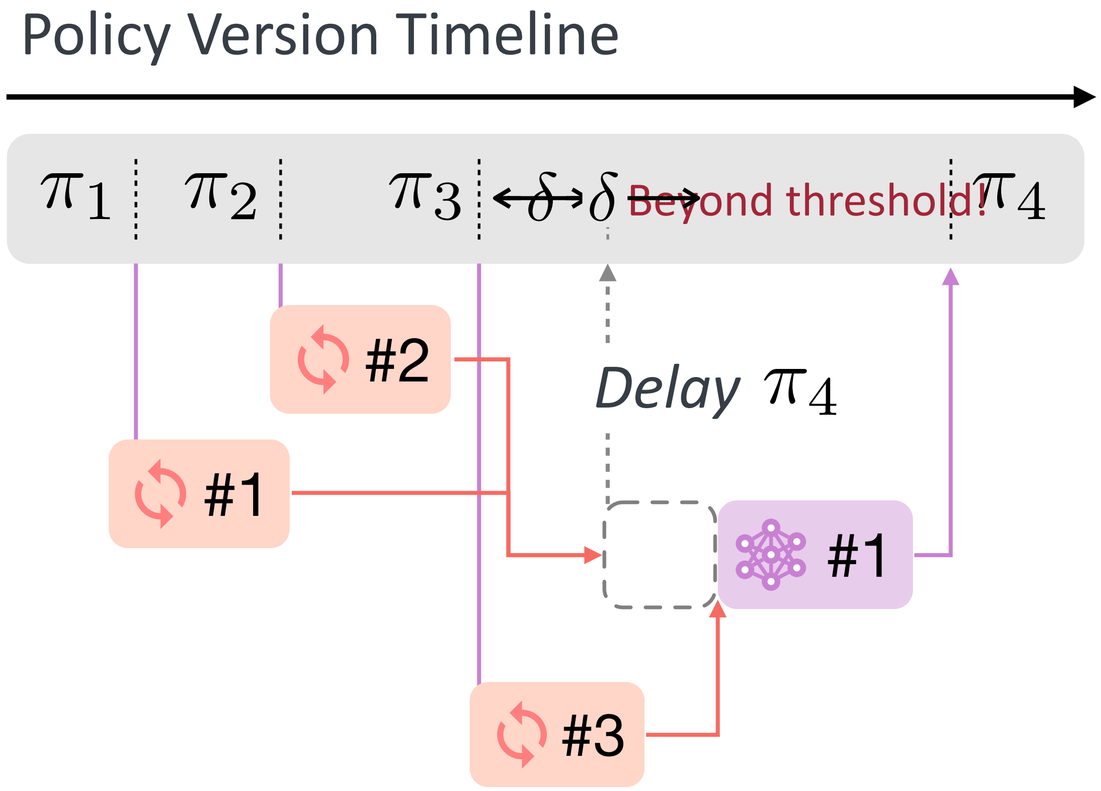


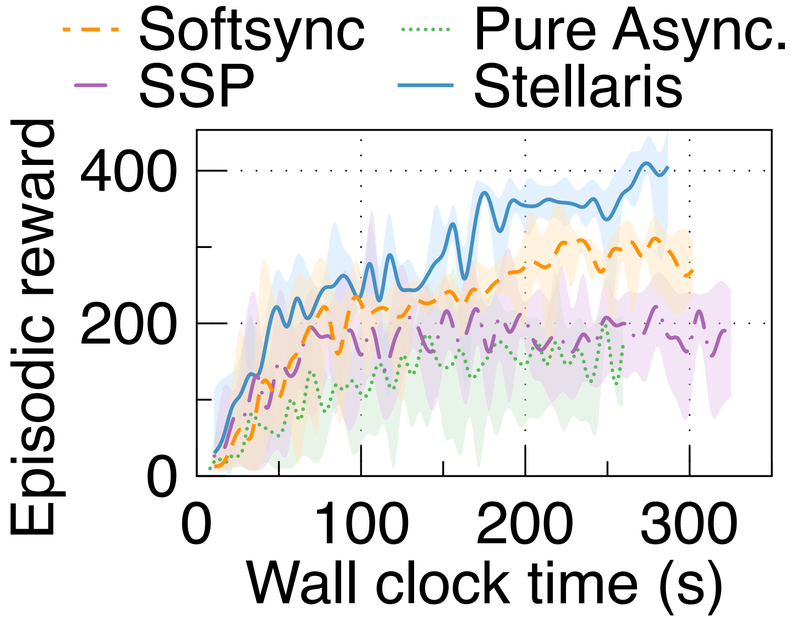


(b) IS truncation

Actor Learner Function Parameter Function Delay

Staleness-Aware Gradient Aggregation





(a) Gradient aggregation

Actor Learner Function Parameter Function Delay

Theoretical Guarantees

Importance Sampling Truncation

Lower bound on
Monotonic reward improvement

$$\frac{1}{T}\sum_{t=1}^{T}\mathbb{E}(||\nabla J(\theta_t)||^2) \leq 2\sqrt{\frac{2C_1C_2}{Tb}}$$

Gradient Aggregation

$\mathcal{O}(\frac{1}{\sqrt{Tb}})$ near-linear convergence rate

$$J(\pi_i) - J(\mu) \geq -\frac{\gamma\epsilon^{\pi_i}\sqrt{2\log\rho}}{(1-\gamma)^2}$$

Please refer details to our paper

Evaluation

Implementation

Ray RLlib
Docker Containers
AWS EC2

Baselines

Ray RLlib [1]
MinionsRL [2]

Metrics

Episodic reward
Training cost

Benchmarks

MuJoCo [3]

Hopper
Humanoid
Walker2d

Atari [4]

Gravitar
SpaceInvaders
Qbert

[1] Liang, E.; et al. RLlib: Abstractions for Distributed Reinforcement Learning. ICML 2018
[2] H. Yu; et al. Cheaper and Faster: Distributed Deep Reinforcement Learning with Serverless Computing. AAAI 2024
[3] Todorov, E.; et al. Mujoco: A Physics Engine for Model-based Control. IROS 2012
[4] Marc, B.; et al. The Arcade Learning Environment: An Evaluation Platform for General Agents. JAIR 2013

Testbed Clusters

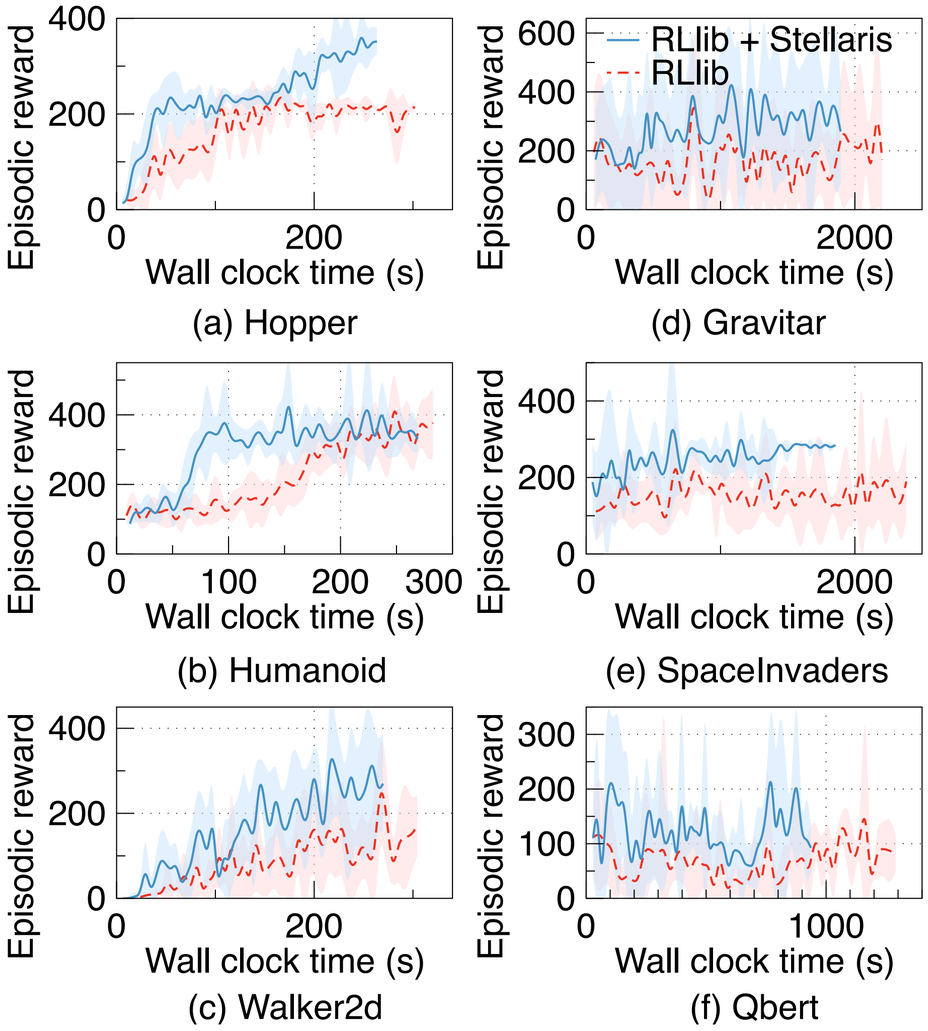
GPU Testbed

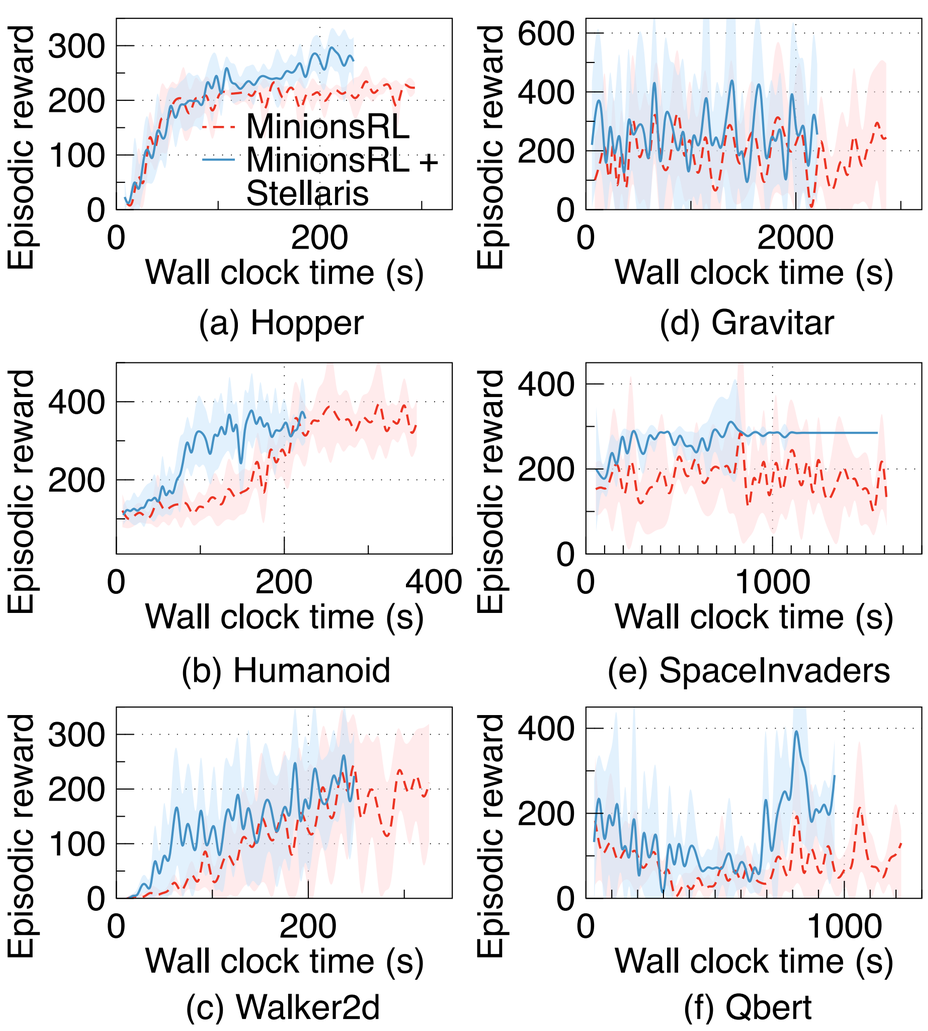
3 nodes
128 AMD EPYC 7R13 CPU cores
2 V100 GPUs

HPC Testbed

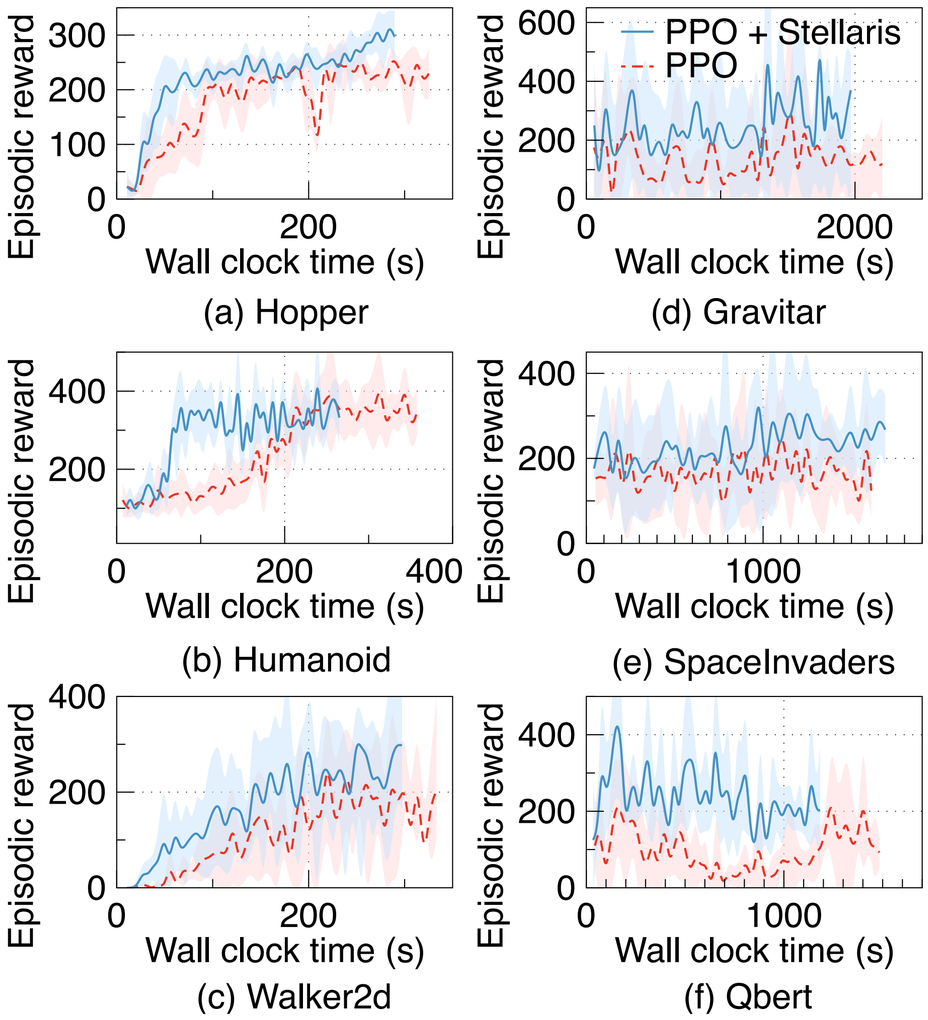
7 nodes
960 AMD EPYC 9R14 CPU cores
16 V100 GPUs

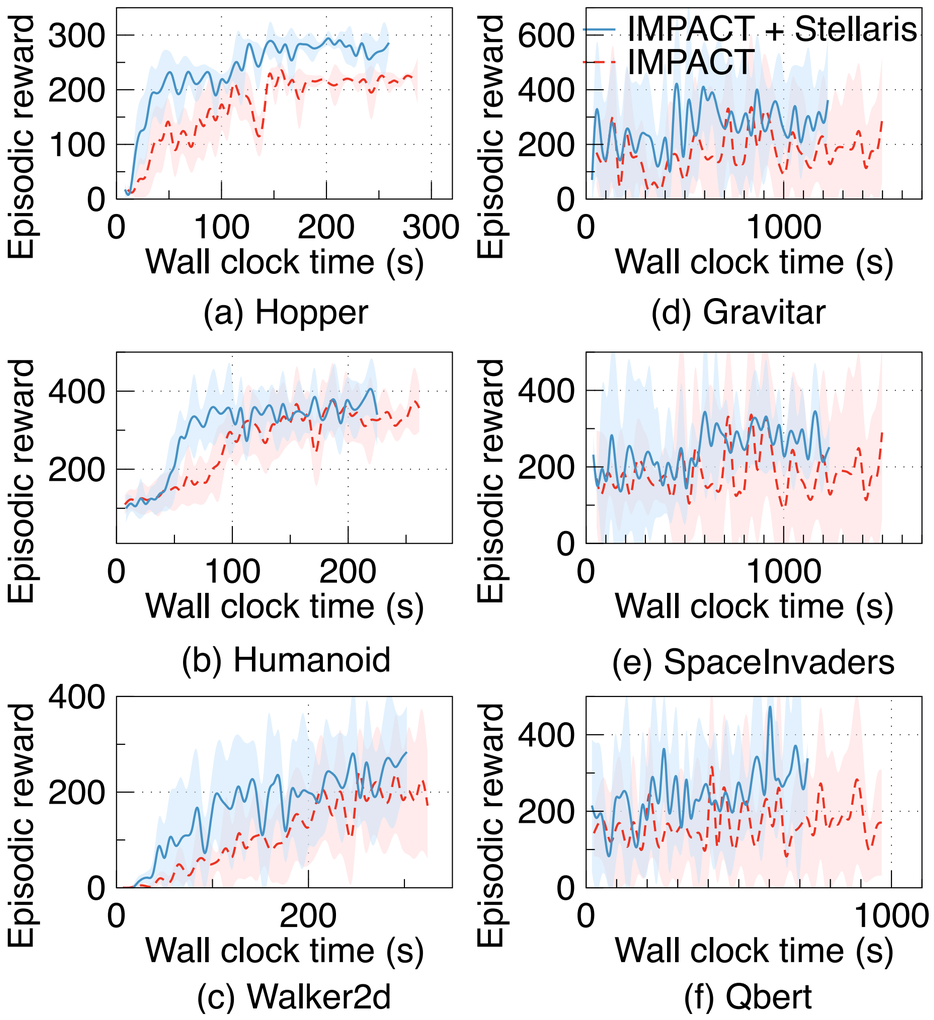
Training Performance



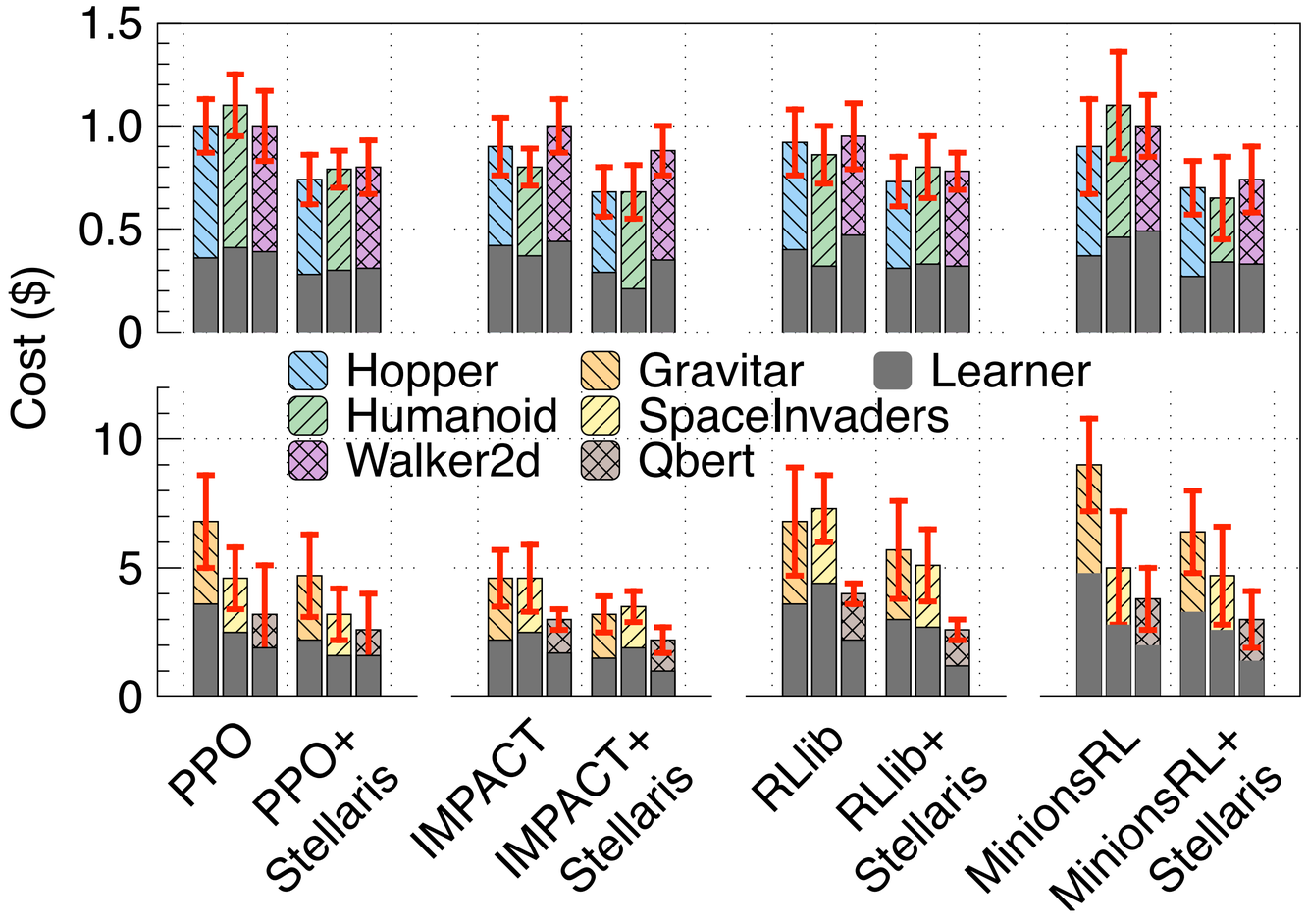


Training Performance

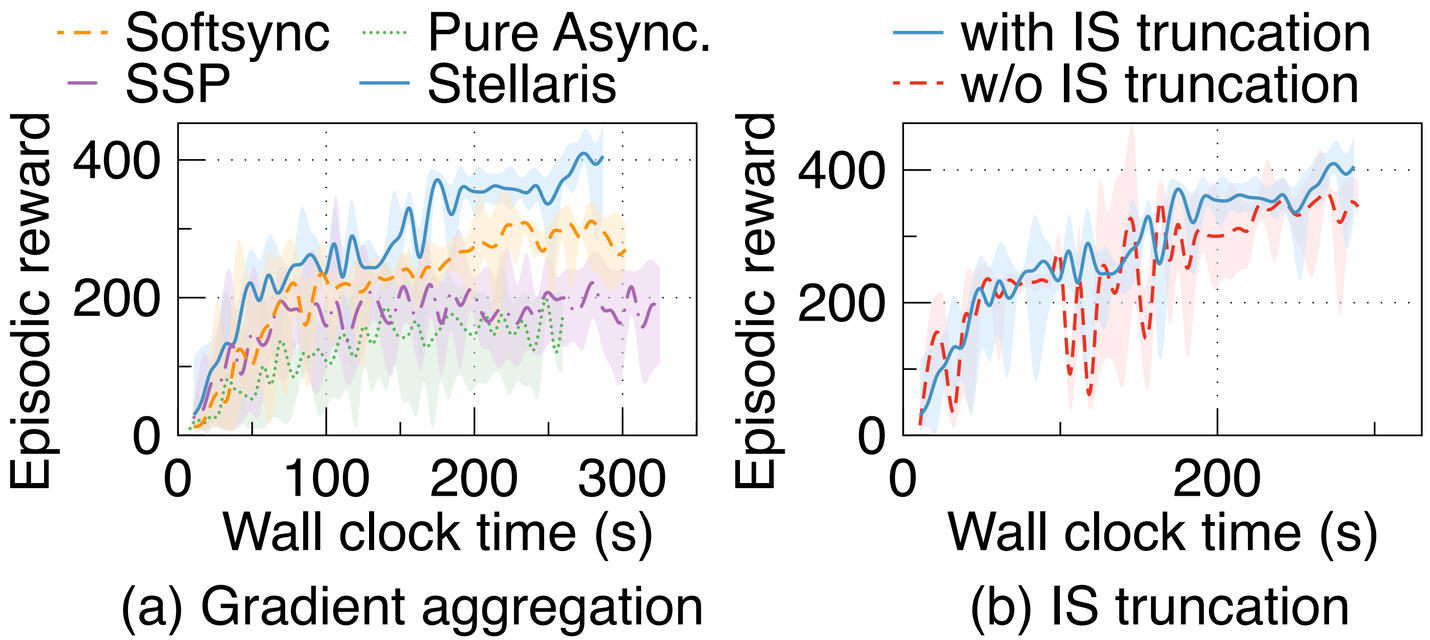




Training Cost



Ablation Study



Asynchronous Serverless Learners

Global Importance Sampling Truncation

Staleness-Aware Gradient Aggregation

Stellaris

2.2x

Training performance improvement

41%

Training cost reduction

STEVENS
INSTITUTE OF TECHNOLOGY
1870

IntelliSys Lab

THANK YOU

Stevens Institute of Technology
1 Castle Point Terrace, Hoboken, NJ 07030