Towards Efficient Multi-Agent Learning Systems

Kailash Gogineni, Peng Wei, Tian Lan, Guru Venkataramani

Department of Electrical & Computer Engineering
School of Engineering & Applied Science

National Science Foundation

THE GEORGE WASHINGTON UNIVERSITY
WASHINGTON, DC

MLArchSys, ISCA 2023

Outline

Multi-Agent RL : Background

Workload Characterization

Neighbor Sampling Strategy

Conclusion & Future work

Outline

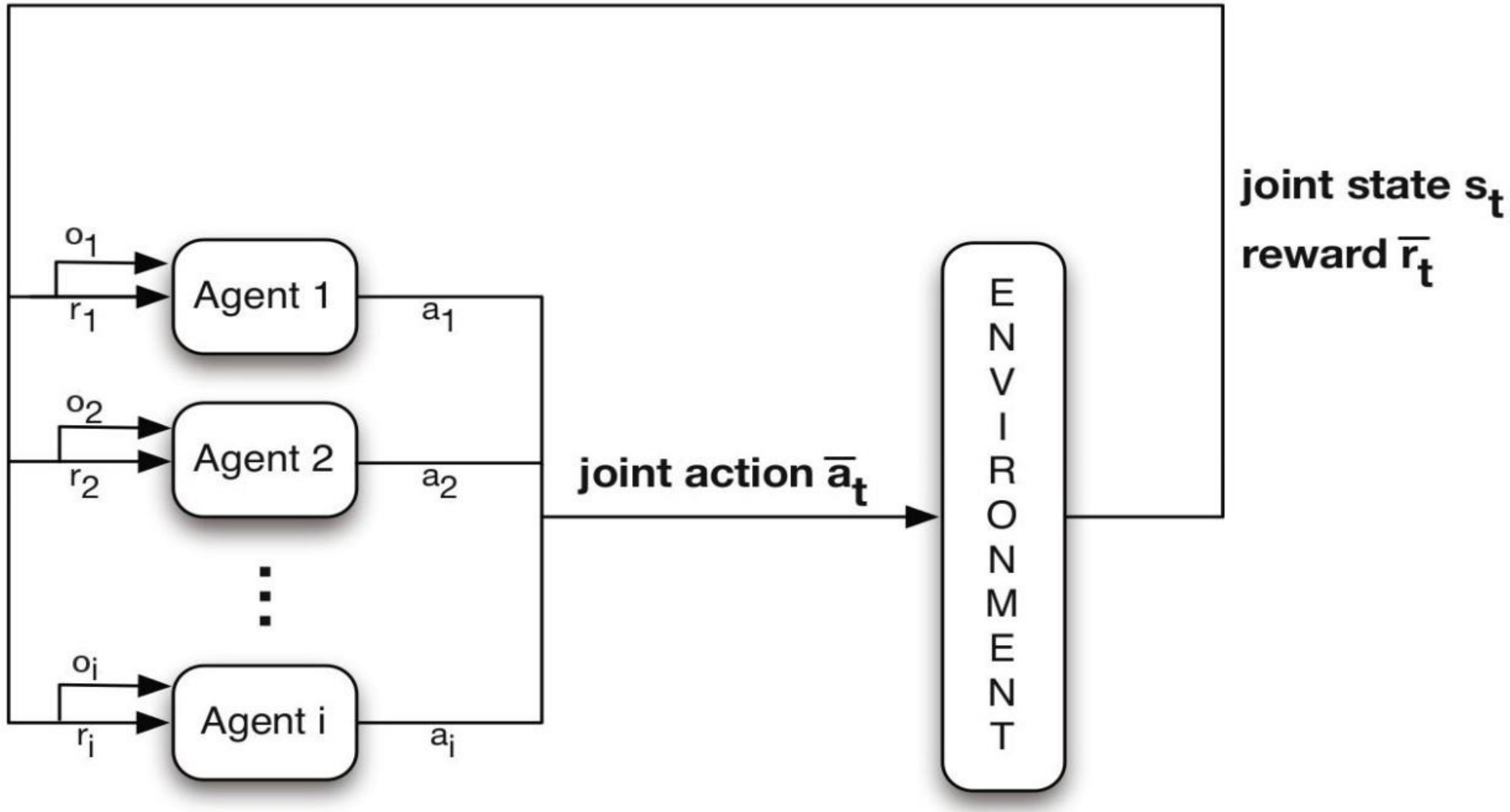
Multi-Agent RL : Background

Workload Characterization

Neighbor Sampling Strategy

Conclusion & Future work

Multi-Agent RL

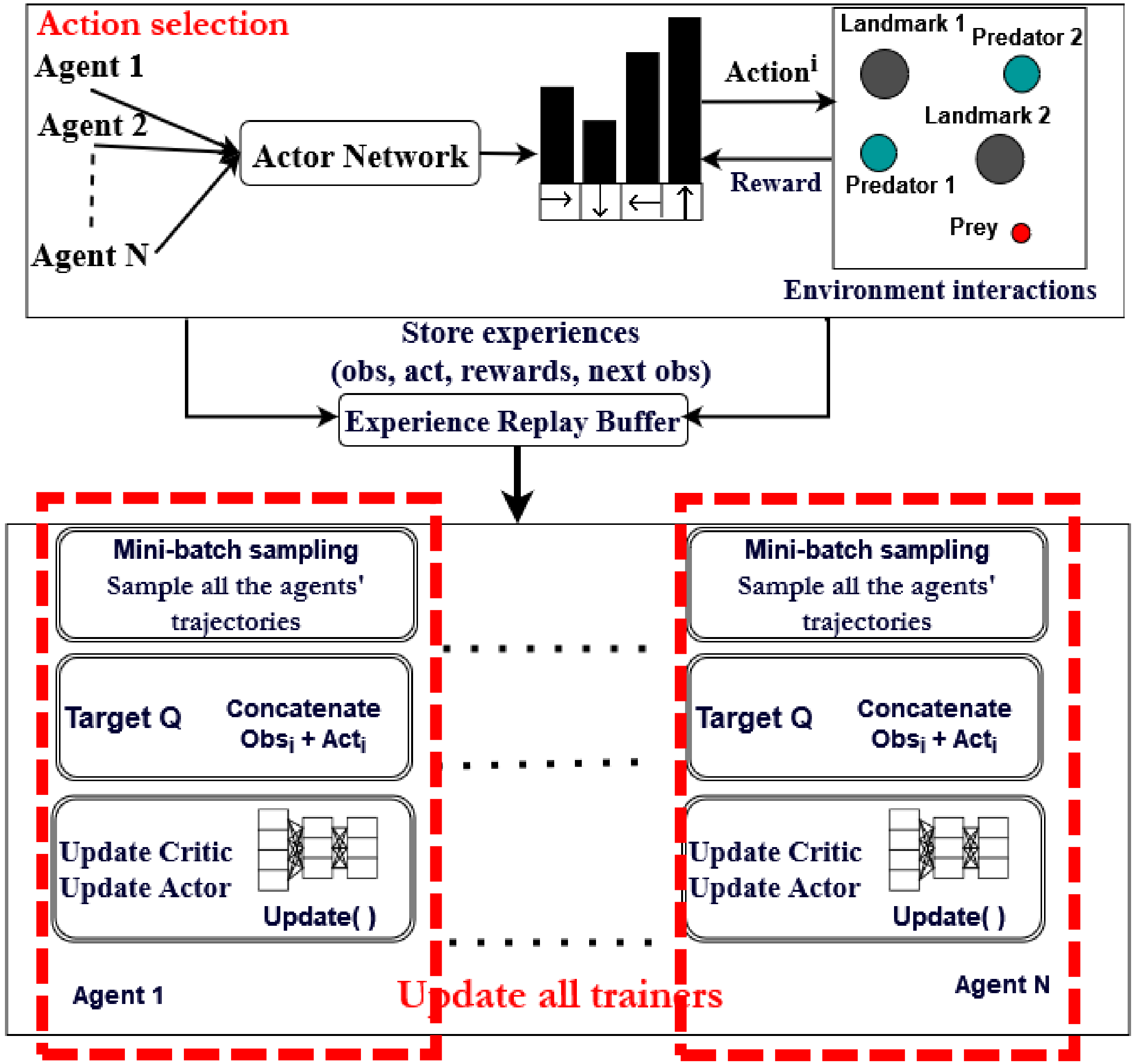


Source: Nowe, Vrancx & De Hauwere 2012

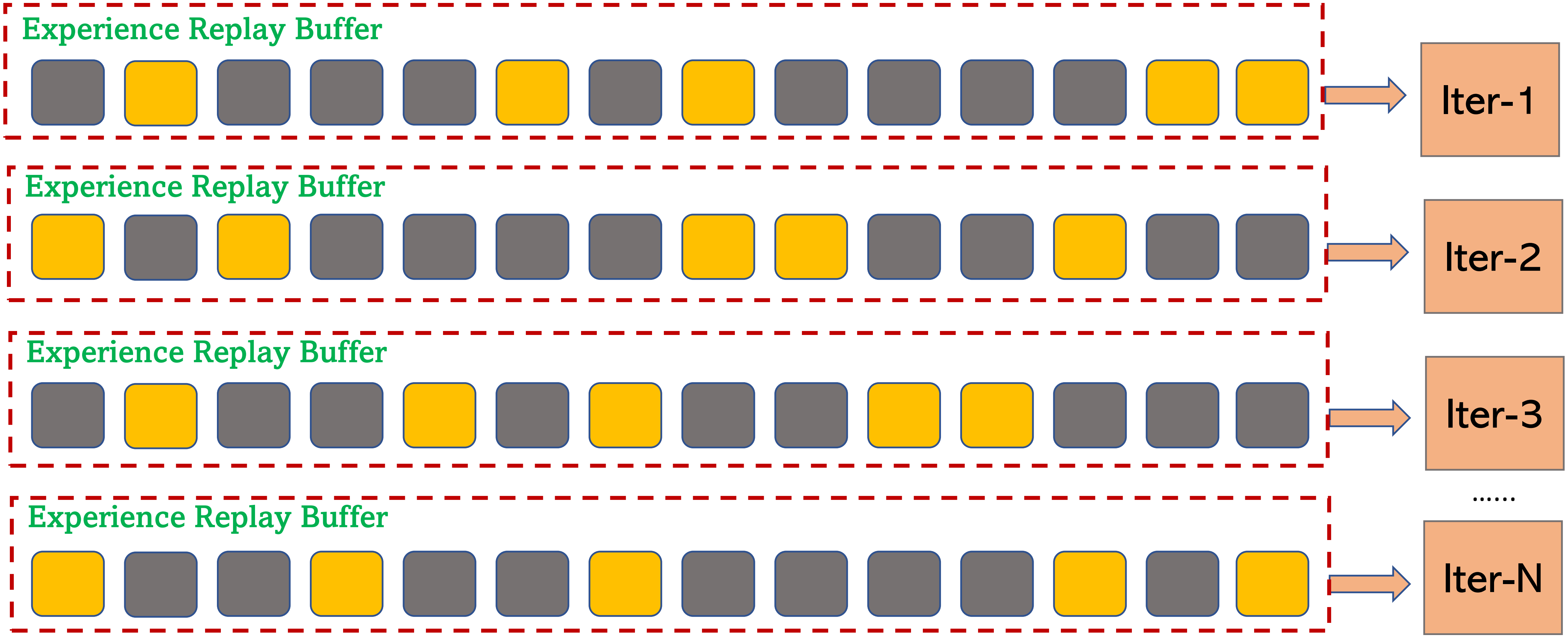
Outline

- Multi-Agent RL : Background
- Workload Characterization
- Neighbor Sampling Strategy
- Conclusion & Future work

Workload breakdown

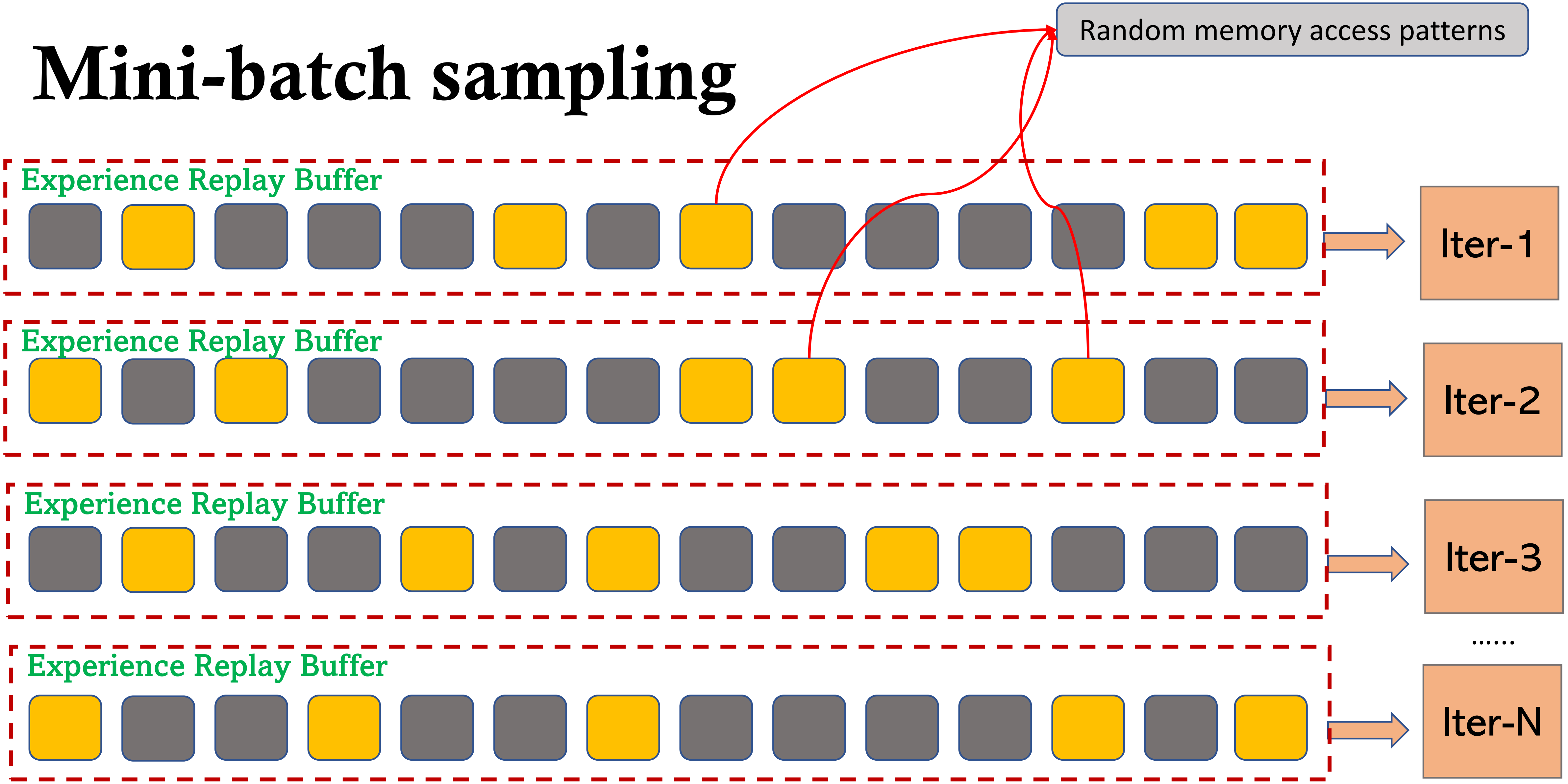


Mini-batch sampling



Layout of Random sampling, batch size = 5

Mini-batch sampling



Layout of Random sampling, batch size = 5

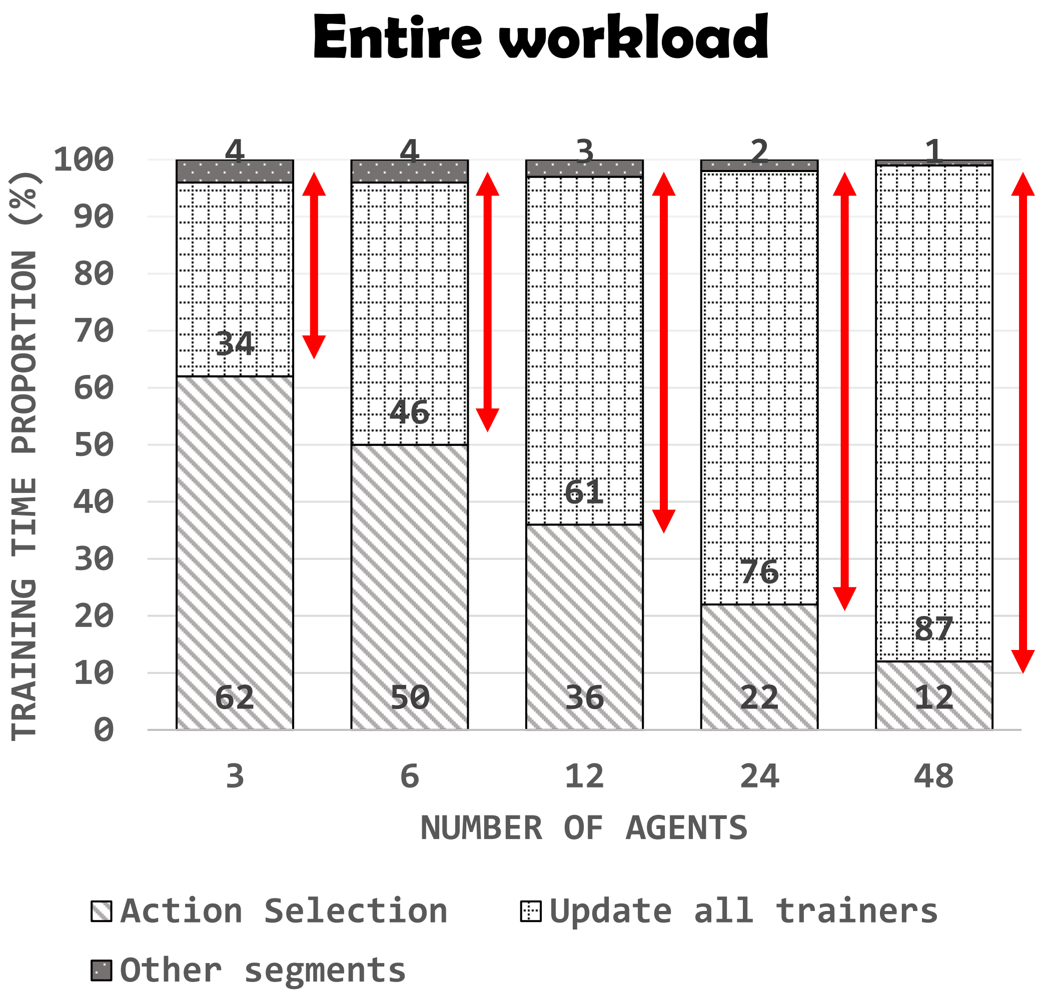
More details in the paper

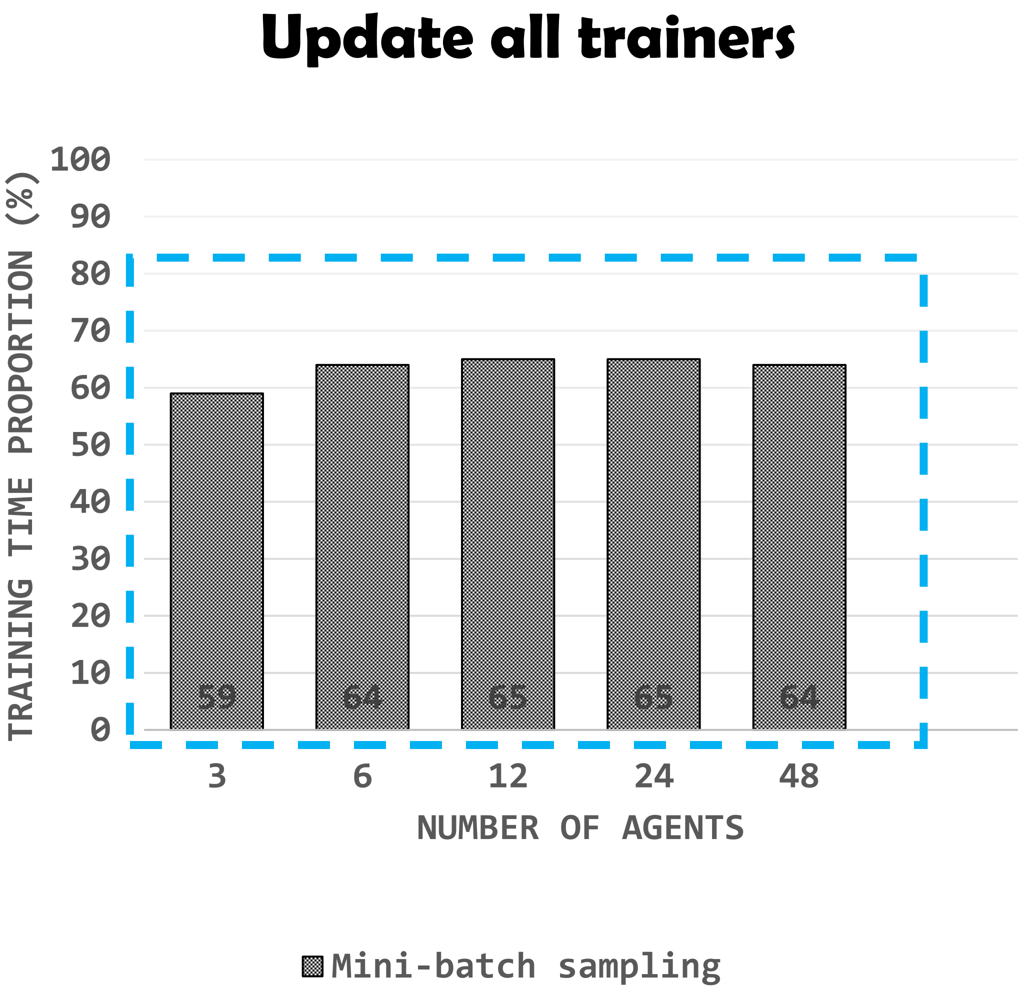
- Target Q calculation
  - State and action spaces grow exponentially
- Back-propagation phases of Actor & Critic networks
  - Sequential updates of Actor-Critic networks
- Mini-batch size limitations in MARL for real-world systems
  - Communication overhead

Methodology

- MARL Workloads
  - MADDPG
  - MASAC (more details in the paper)
- Multi-Agent Particle Environment
  - Competitive task (Predator-Prey) - N predators work cooperatively to block the way of M fast paced prey agents. The prey agents are environment controlled and they try to avoid the collision with predators
- Hyper-parameters
  - The workloads are trained for 60K episodes, max episode length is 25, size of replay buffer is 1M, mini-batch size is 1024, entropy coefficient for MASAC is 0.05
- Platform (experimental testbed)
  - Ampere Architecture NVIDIA-RTX 3090

MADDPG: Training time from 3 to 48 agents





☞ **MASAC algorithm also exhibits the same pattern!**

Outline

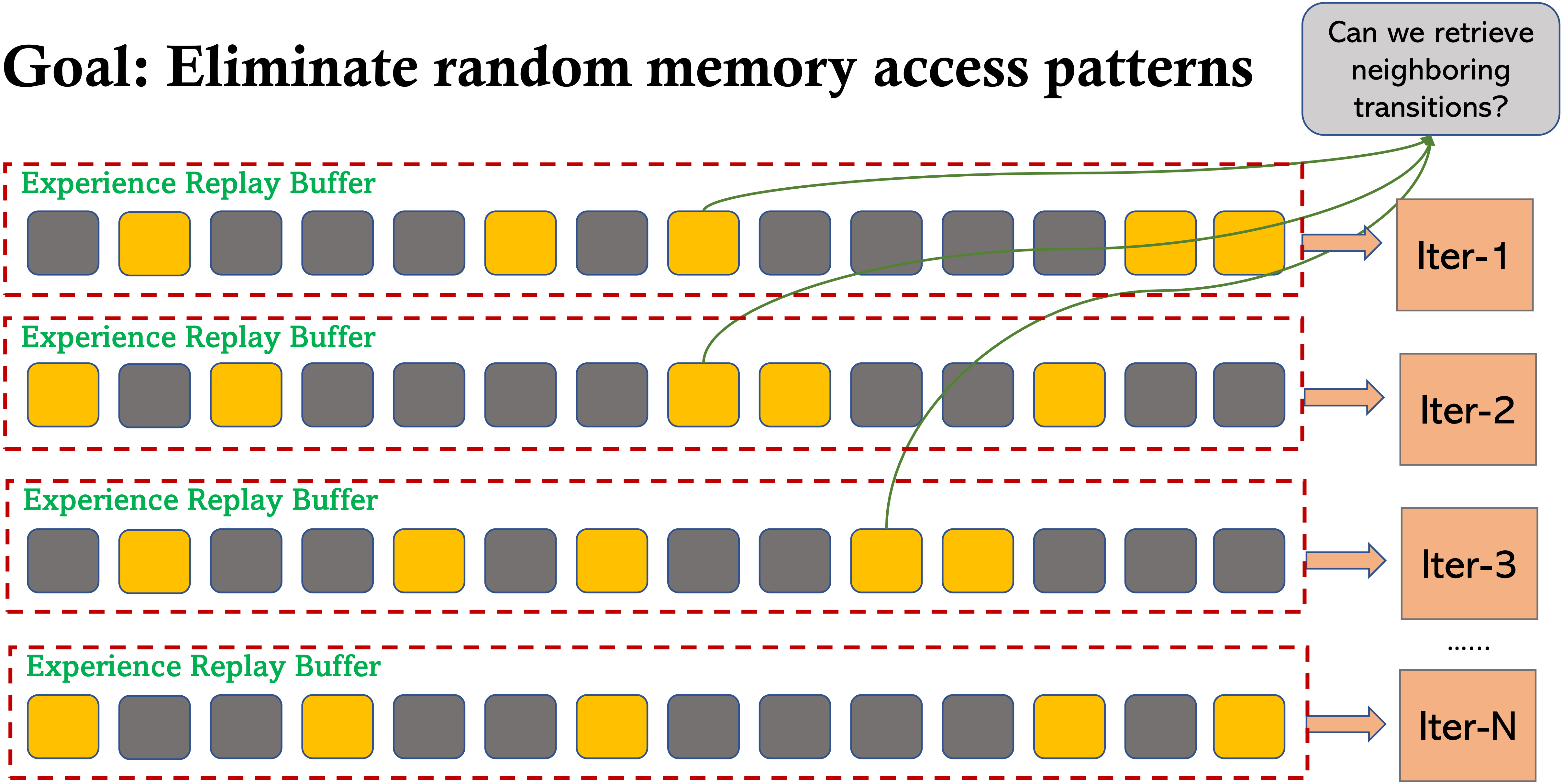
Multi-Agent RL : Background

Workload Characterization

Neighbor Sampling Strategy

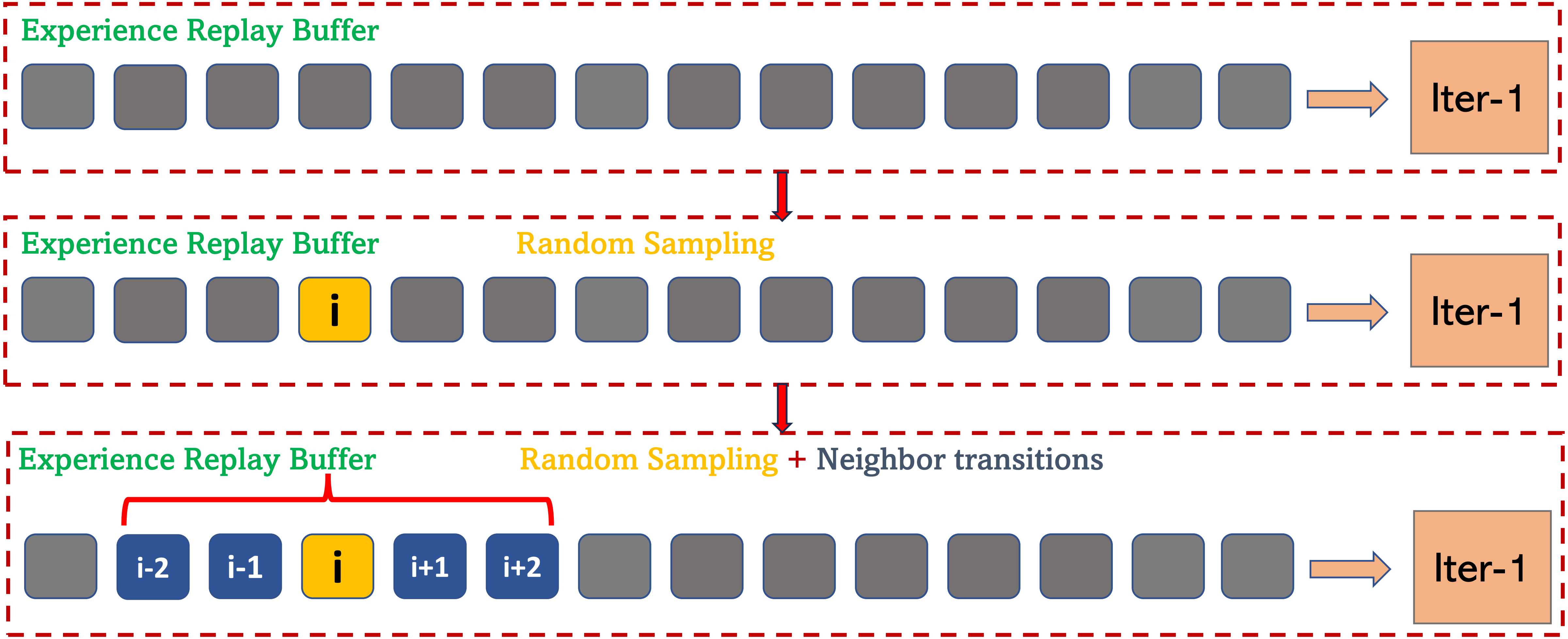
Conclusion & Future work

Goal: Eliminate random memory access patterns



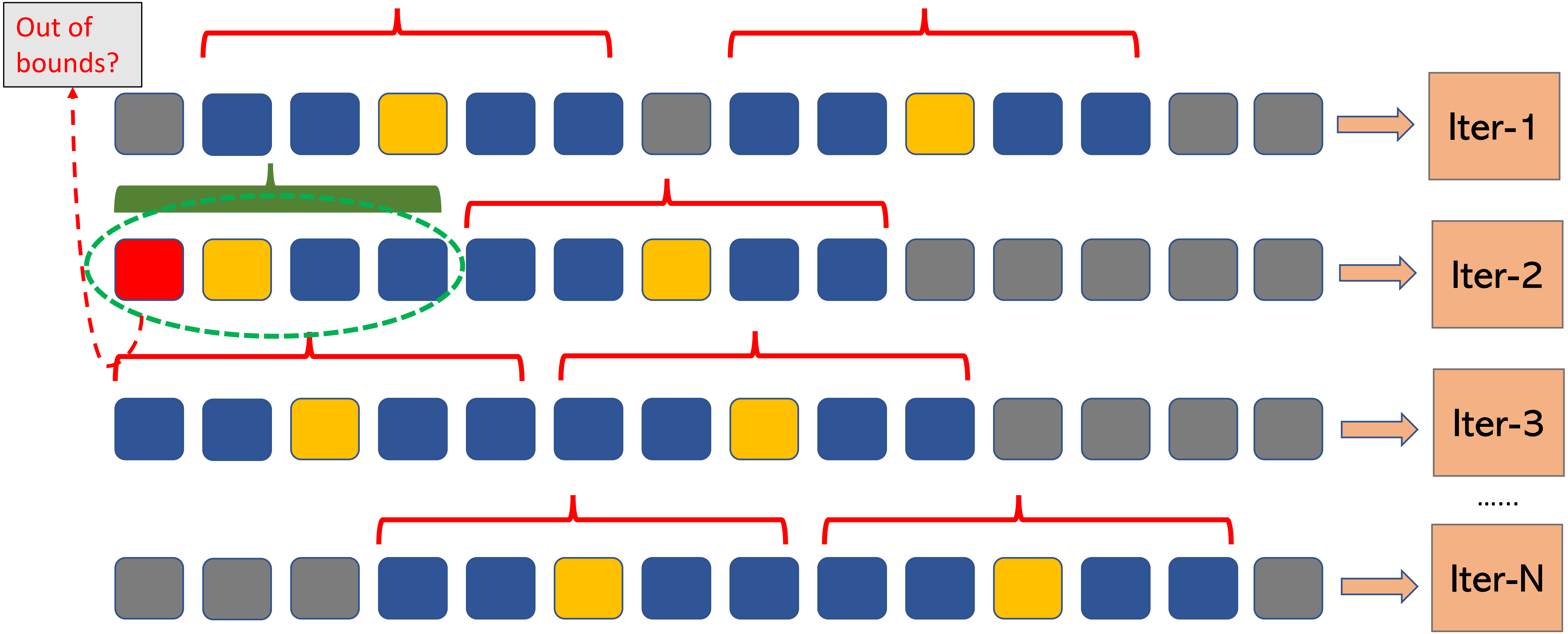
Layout of Random sampling

Effect of neighbor sampling



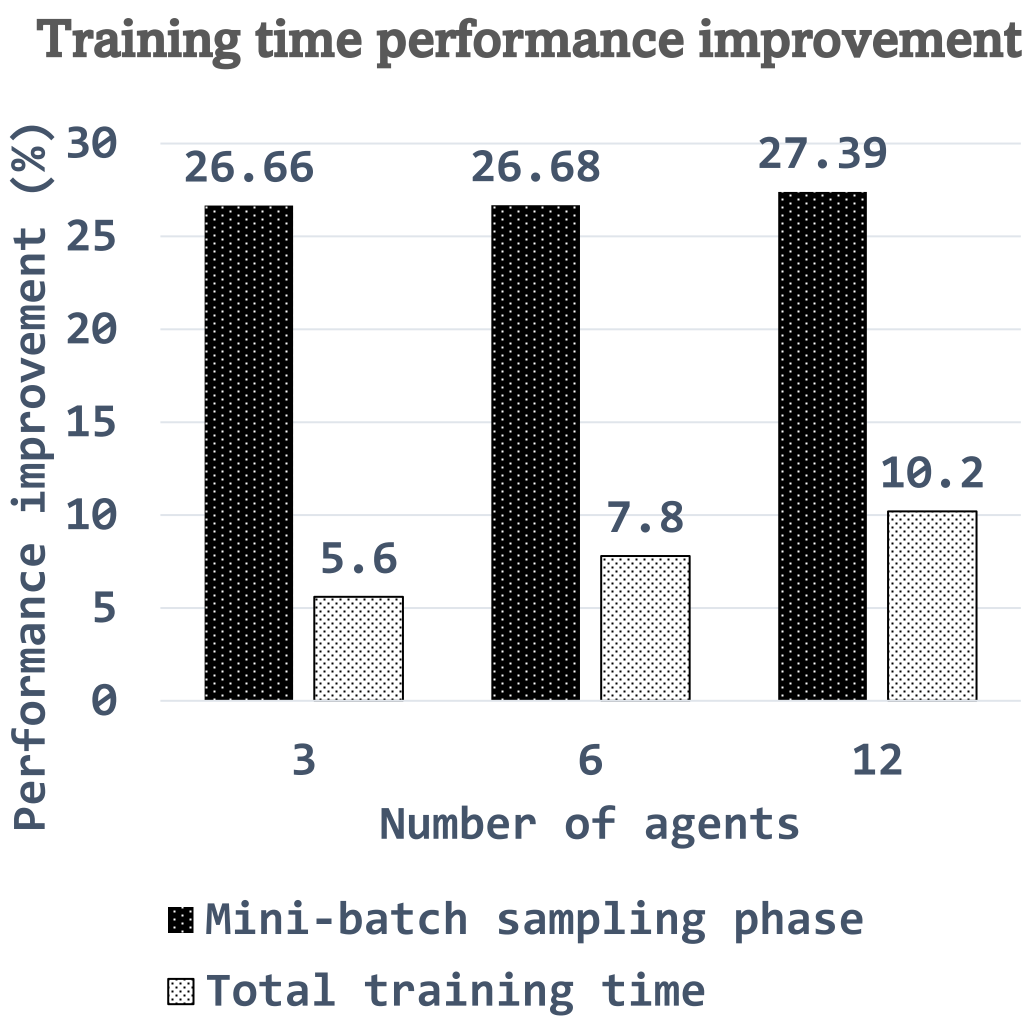
Layout of Neighbor sampling strategy

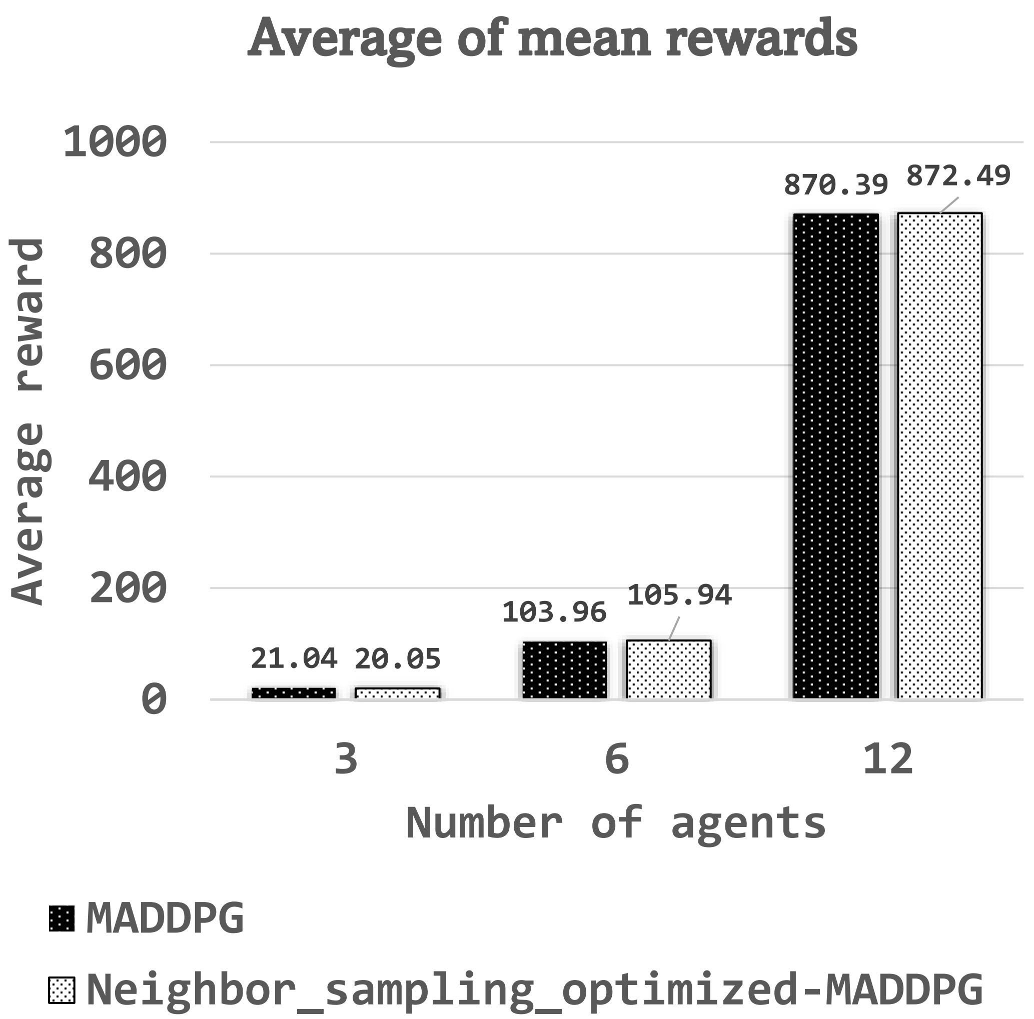
Effect of neighbor sampling



Layout of Neighbor sampling strategy, micro-batch size = 2

Results


Training time performance improvement
Performance improvement (%)
30
25
20
15
10
5
0
26.66
26.68
27.39
5.6
7.8
10.2
3
6
12
Number of agents
Mini-batch sampling phase
Total training time


Average of mean rewards
Average reward
1000
800
600
400
200
0
21.04
20.05
103.96
105.94
870.39
872.49
3
6
12
Number of agents
MADDPG
Neighbor_sampling_optimized-MADDPG

Outline

Multi-Agent RL : Background

Workload Characterization

Neighbor Sampling Strategy

Conclusion & Future work

Conclusion & Future work

- We understand and characterize several MARL algorithms to identify key bottlenecks from a systems perspective
- We implemented a simple heuristic, neighbor sampling strategy to address the mini-batch sampling phase
- For future work, we will investigate:
  - Various efficient sampling strategies and design a hardware-friendly architecture to efficiently fetch the transitions in large-scale MARL
  - Use algorithmic optimizations into systems to reduce the observation-action space

Thank You!

Q & A

References

- CMU AI and http://ai.berkeley.edu
- Google DeepMind – Multi-Agent Systems & AI DeepMind
- Lowe, Ryan, et al. "Multi-agent actor-critic for mixed cooperative-competitive environments." Advances in neural information processing systems 30 (2017).
- Ackermann, Johannes, et al. "Reducing overestimation bias in multi-agent domains using double centralized critics." arXiv preprint arXiv:1910.01465 (2019).
- Haarnoja, Tuomas, et al. "Soft actor-critic: Off-policy maximum entropy deep reinforcement learning with a stochastic actor." International conference on machine learning. PMLR, 2018.