Federated Reinforcement Learning: Linear Speedup Under Markovian Sampling

Sajad Khodadadian

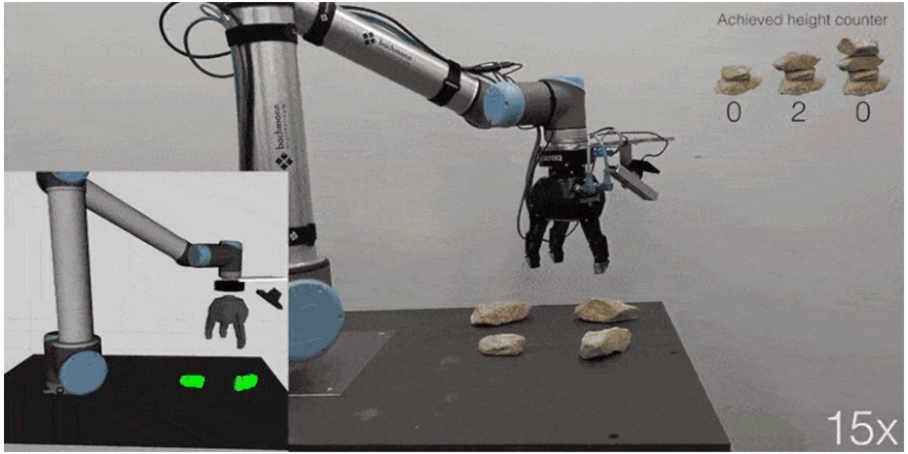
June, 2022

Pranay Sharma (CMU)

Gauri Joshi (CMU)

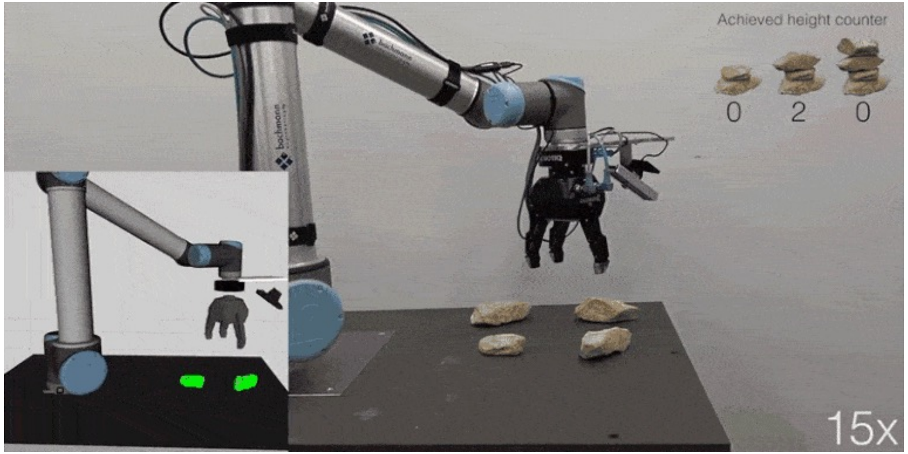
Siva Theja Maguluri (Gatech)

Reinforcement Learning

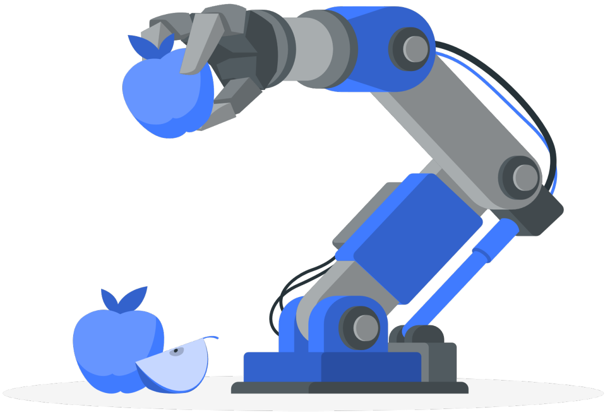


freecodecamp.org

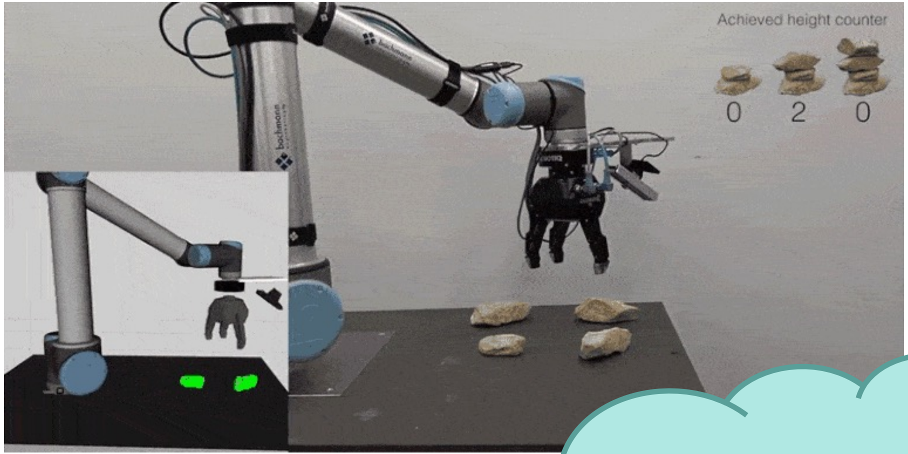
Reinforcement Learning



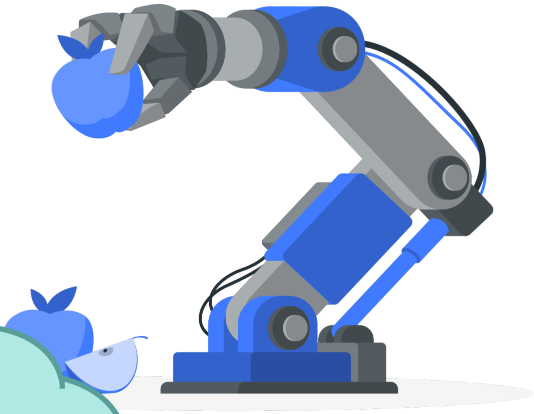
freecodecamp.org



Reinforcement Learning



freecodecamp.org



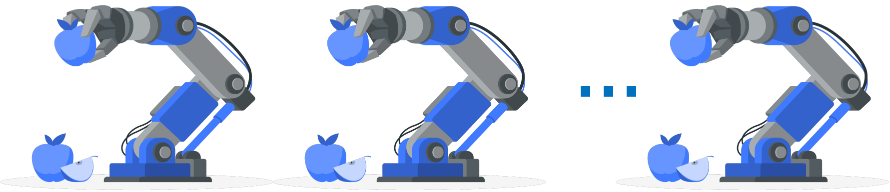
But RL is data intensive!

Federated Reinforcement Learning

Google AI Blog

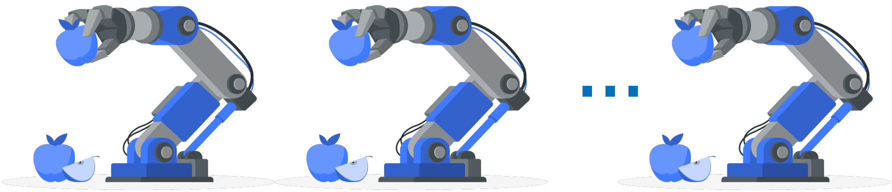
Federated Reinforcement Learning

Google AI Blog



Federated Reinforcement Learning


Google AI Blog



Multiple data collecting agents

Background on MDP Theory

- Discounted Markov Decision Process (MDP)
- Q-function

$$\boxed{Q^{\pi}(s,a) = \mathbb{E}\{\textstyle\sum_{t=0}^{\infty} \gamma^t \mathcal{R}(S_t, A_t) \,|S_0 = s, A_0 = a, A_t \sim \pi(\cdot\,|S_t)\}}$$

Background on MDP Theory

- Discounted Markov Decision Process (MDP)

- Q-function

$$\boxed{Q^{\pi}(s, a) = \mathbb{E}\{\textstyle\sum_{t=0}^{\infty} \gamma^{t} \mathcal{R}(S_t, A_t)\,|S_0 = s, A_0 = a, A_t \sim \pi(\cdot\,|S_t)\}}$$

initial state

Background on MDP Theory

- Discounted Markov Decision Process (MDP)

- Q-function

$$\boxed{Q^{\pi}(s, a) = \mathbb{E}\{\textstyle\sum_{t=0}^{\infty} \gamma^{t} \mathcal{R}(S_t, A_t)\,|S_0 = s, A_0 = a, A_t \sim \pi(\cdot\,|S_t)\}}$$

initial state

initial action

Background on MDP Theory

- Discounted Markov Decision Process (MDP)
- Q-function

$$Q^{\pi}(s, a) = \mathbb{E}\{\textstyle\sum_{t=0}^{\infty} \gamma^t \mathcal{R}(S_t, A_t) \,|S_0 = s, A_0 = a, A_t \sim \pi(\cdot\,|S_t)\}$$

initial state

initial action

policy

Background on MDP Theory

- Discounted Markov Decision Process (MDP)
- Q-function

state at time t

$$Q^\pi(s, a) = \mathbb{E}\{\sum_{t=0}^{\infty} \gamma^t \mathcal{R}(S_t, A_t) \,|S_0 = s, A_0 = a, A_t \sim \pi(\cdot\,|S_t)\}$$

initial state

initial action

policy

Background on MDP Theory

- Discounted Markov Decision Process (MDP)
- Q-function

state at time t

$$Q^{\pi}(s, a) = \mathbb{E}\{\textstyle\sum_{t=0}^{\infty} \gamma^{t} \mathcal{R}(S_t, A_t) \mid S_0 = s, A_0 = a, A_t \sim \pi(\cdot \mid S_t)\}$$

initial state initial action action at time t policy

Background on MDP Theory

- Discounted Markov Decision Process (MDP)
- Q-function

Reward function

state at time t

$$Q^{\pi}(s, a) = \mathbb{E}\{\textstyle\sum_{t=0}^{\infty} \gamma^{t} \mathcal{R}(S_t, A_t) \,|S_0 = s, A_0 = a, A_t \sim \pi(\cdot\,|S_t)\}$$

policy

initial state

initial action

action at time t

Background on MDP Theory

- Discounted Markov Decision Process (MDP)
- Q-function

Reward function

state at time t

$$Q^{\pi}(s, a) = \mathbb{E}\{\textstyle\sum_{t=0}^{\infty} \gamma^t \mathcal{R}(S_t, A_t) \mid S_0 = s, A_0 = a, A_t \sim \pi(\cdot \mid S_t)\}$$

initial state

initial action

discount factor

action at time t

policy

Background on MDP Theory

- Discounted Markov Decision Process (MDP)
- Q-function

Reward function

state at time t

$$Q^\pi(s, a) = \mathbb{E}\{\textstyle\sum_{t=0}^{\infty} \gamma^t \mathcal{R}(S_t, A_t) \mid S_0 = s, A_0 = a, A_t \sim \pi(\cdot \mid S_t)\}$$

policy

discount factor

initial state

initial action

action at time t

- Optimal policy

$$\pi^* \in \operatorname*{argmax}_{\pi} Q^\pi(s, a), \qquad \forall s, a$$

Background on MDP Theory

- Discounted Markov Decision Process (MDP)
- Q-function

Reward function state at time t

$$Q^{\pi}(s,a) = \mathbb{E}\{\textstyle\sum_{t=0}^{\infty} \gamma^{t} \mathcal{R}(S_t, A_t) \mid S_0 = s, A_0 = a, A_t \sim \pi(\cdot \mid S_t)\}$$

initial state initial action discount factor action at time t policy

- Optimal policy

$$\pi^* \in \operatorname*{argmax}_{\pi} Q^{\pi}(s,a), \qquad \forall s, a$$

- Optimal Q-function

$$Q^*(s,a) \equiv Q^{\pi^*}(s,a)$$

Background on MDP Theory

- Discounted Markov Decision Process (MDP)
- Q-function

Reward function   state at time t

$$Q^{\pi}(s,a) = \mathbb{E}\{\textstyle\sum_{t=0}^{\infty} \gamma^t \mathcal{R}(S_t, A_t) \,|S_0 = s, A_0 = a, A_t \sim \pi(\cdot\,|S_t)\}$$

initial state   initial action   discount factor   action at time t policy

- Optimal policy

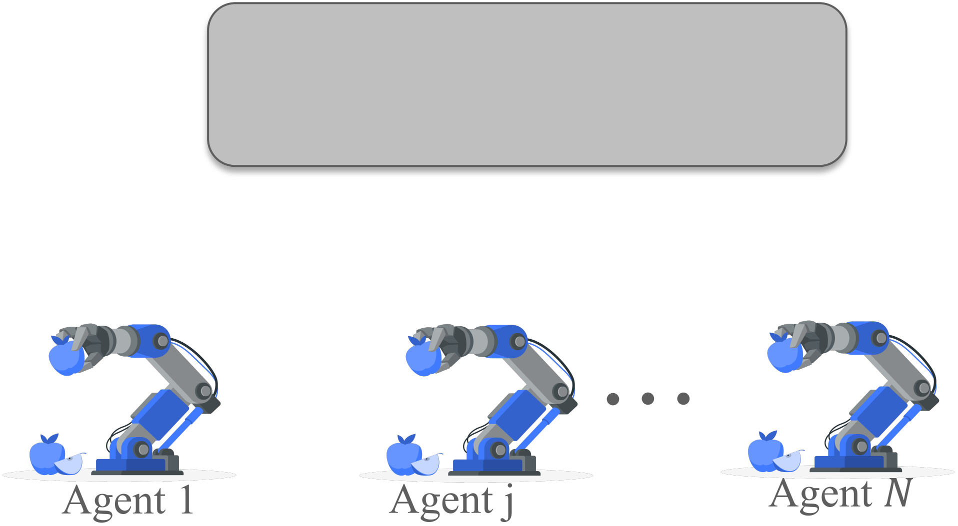
$$\pi^* \in \operatorname*{argmax}_{\pi} Q^{\pi}(s,a)\,, \qquad \forall s, a$$

- Optimal Q-function

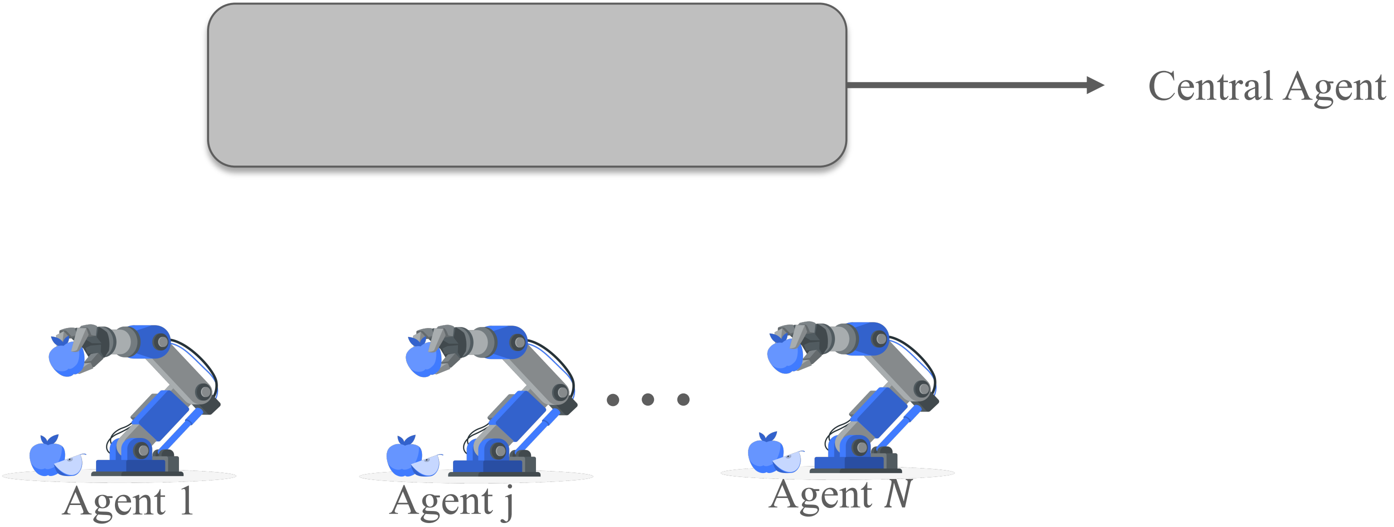
$$Q^*(s,a) \equiv Q^{\pi^*}(s,a)$$

→ Q-learning

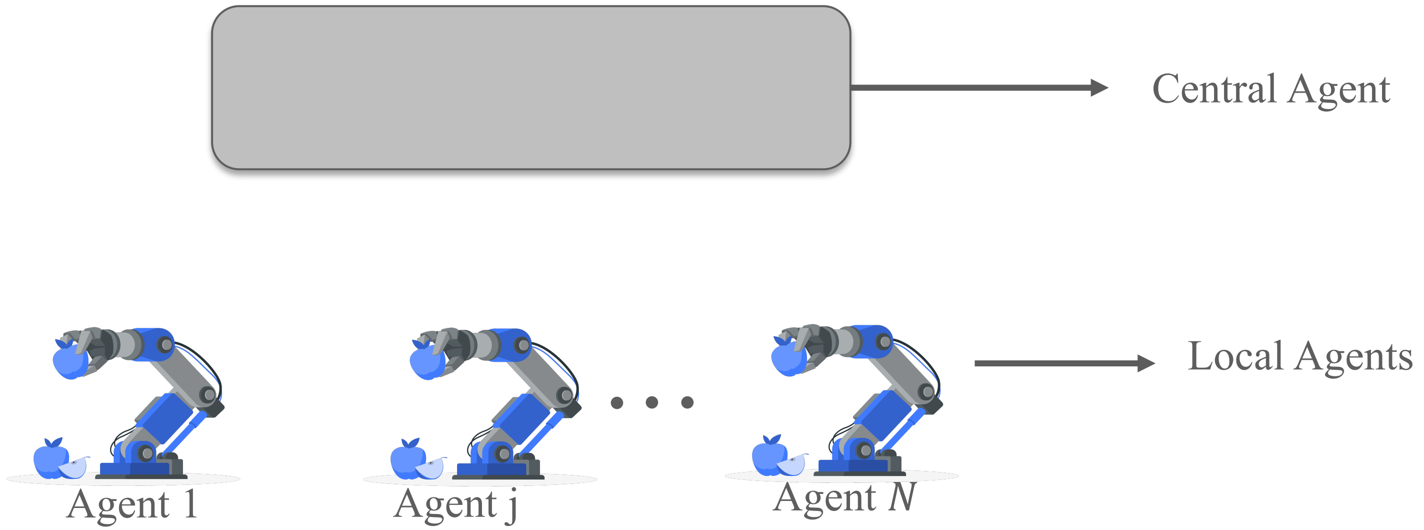
Vanilla Distributed Reinforcement Learning



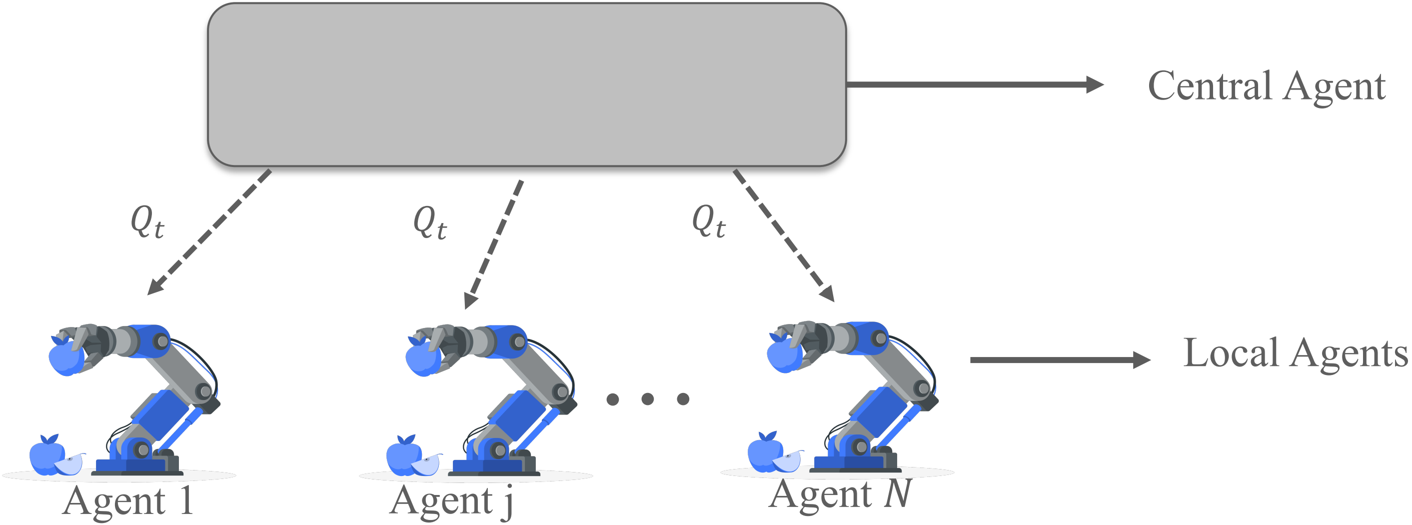
Vanilla Distributed Reinforcement Learning



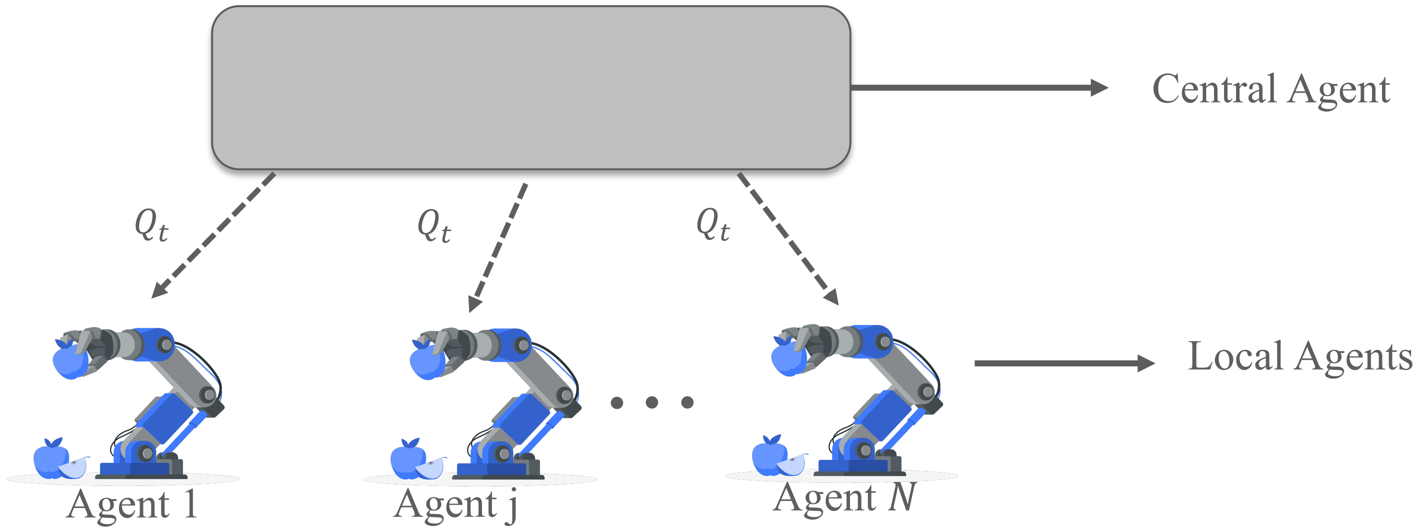
Vanilla Distributed Reinforcement Learning



Vanilla Distributed Reinforcement Learning

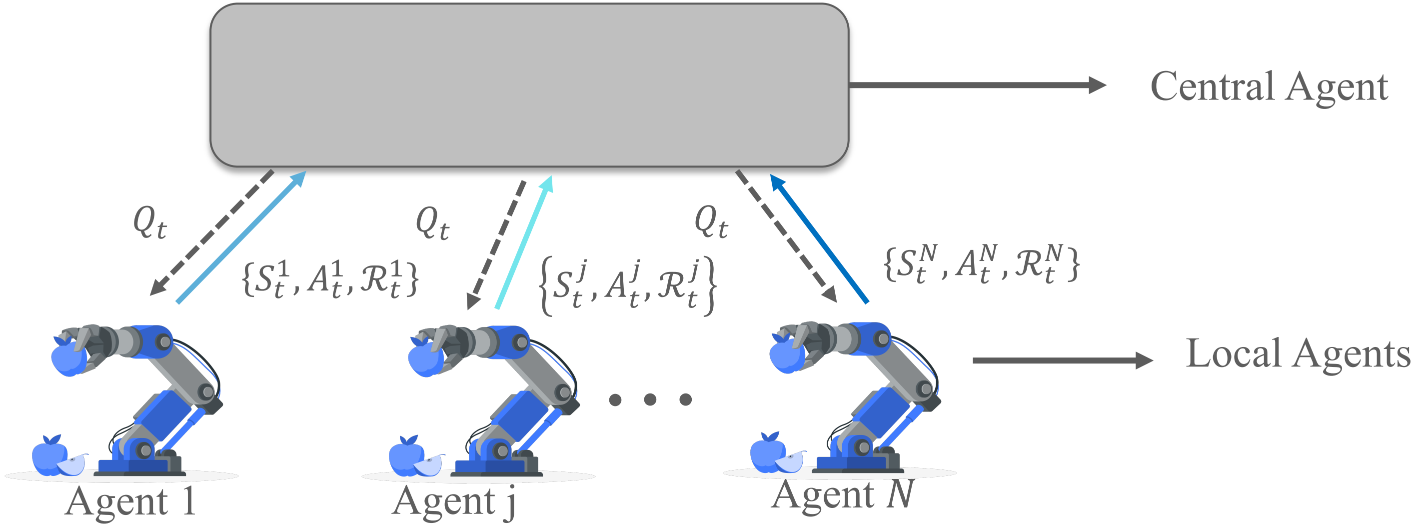


Vanilla Distributed Reinforcement Learning



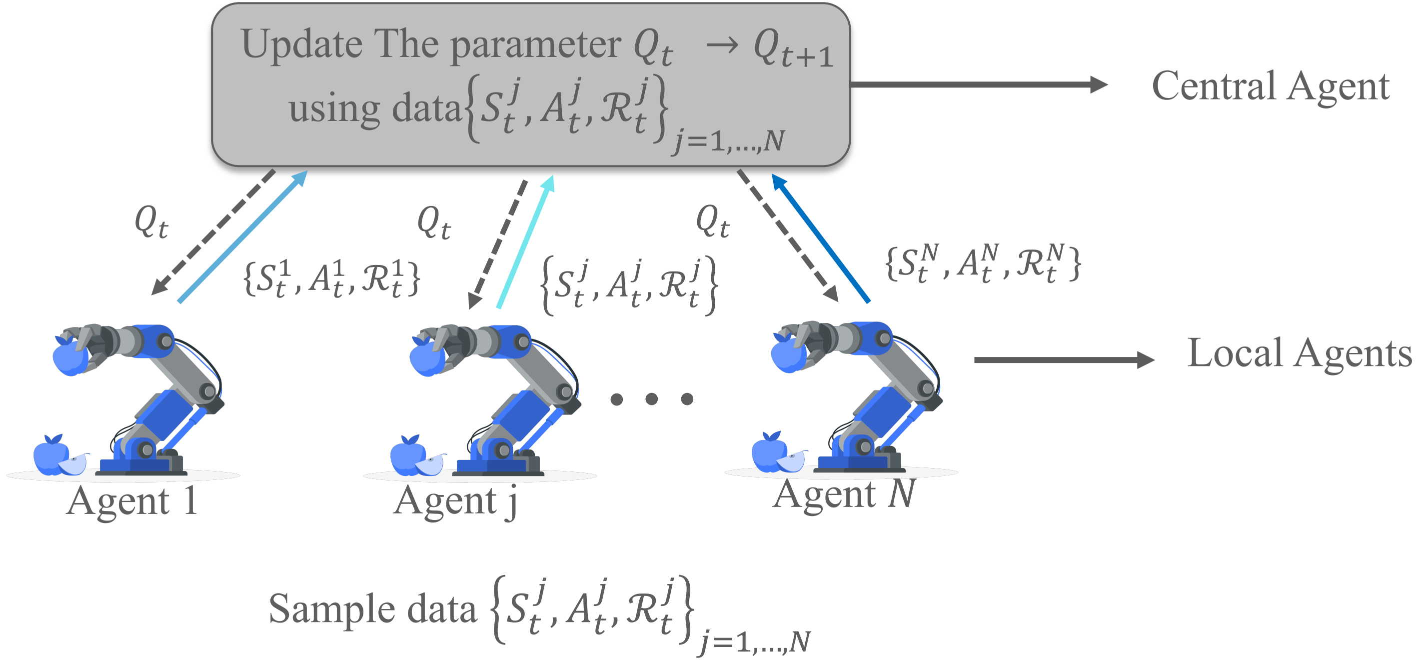
Sample data $\left\{S_t^j, A_t^j, \mathcal{R}_t^j\right\}_{j=1,\ldots,N}$

Vanilla Distributed Reinforcement Learning

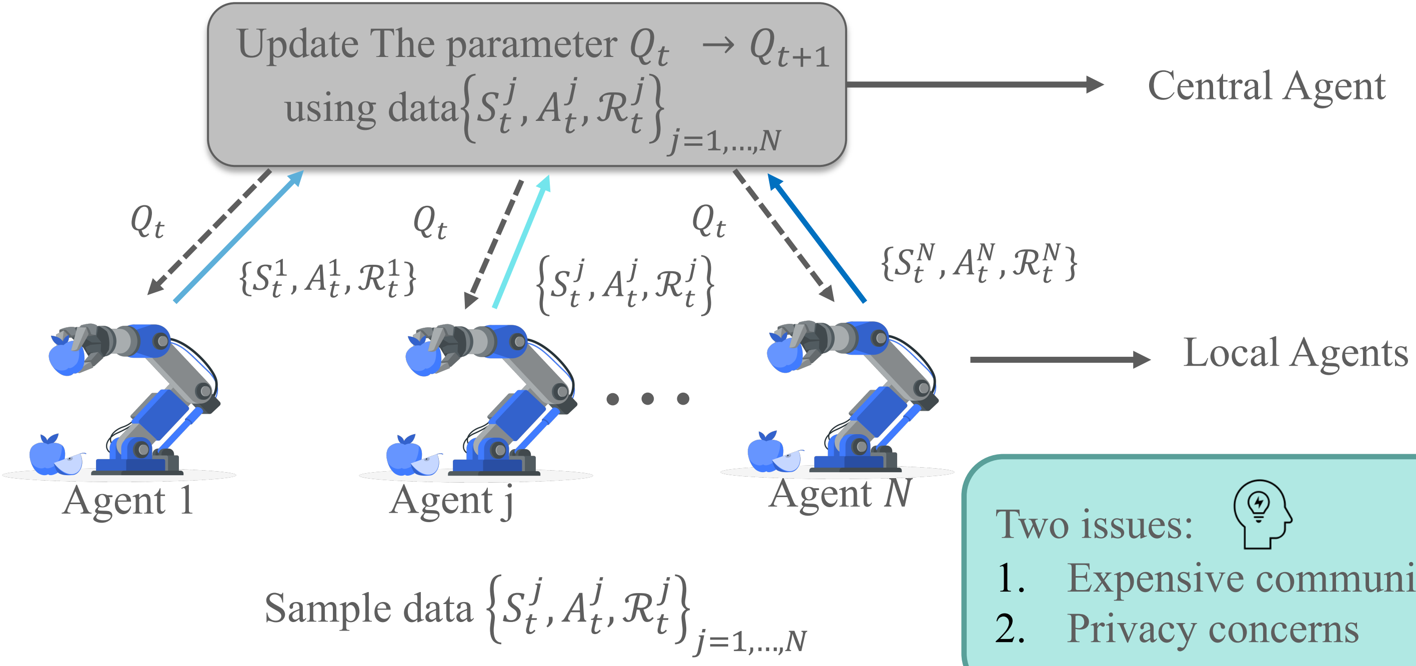


Sample data $\left\{S_t^j, A_t^j, \mathcal{R}_t^j\right\}_{j=1,\ldots,N}$

Vanilla Distributed Reinforcement Learning



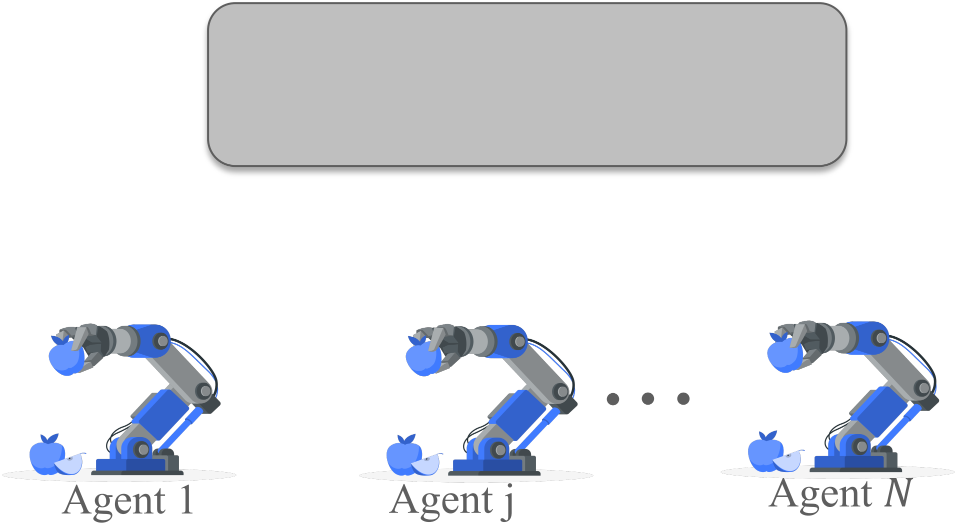
Vanilla Distributed Reinforcement Learning



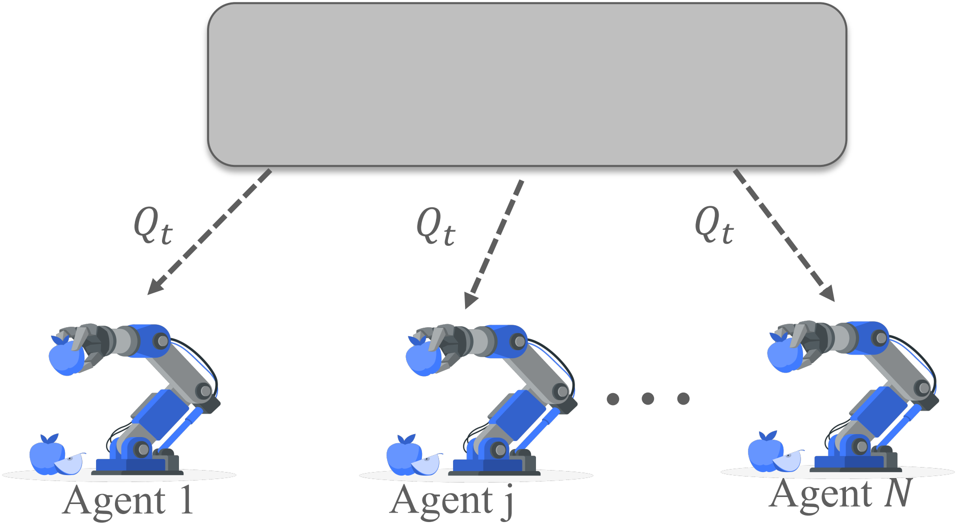
Two issues:

1. Expensive communication
2. Privacy concerns

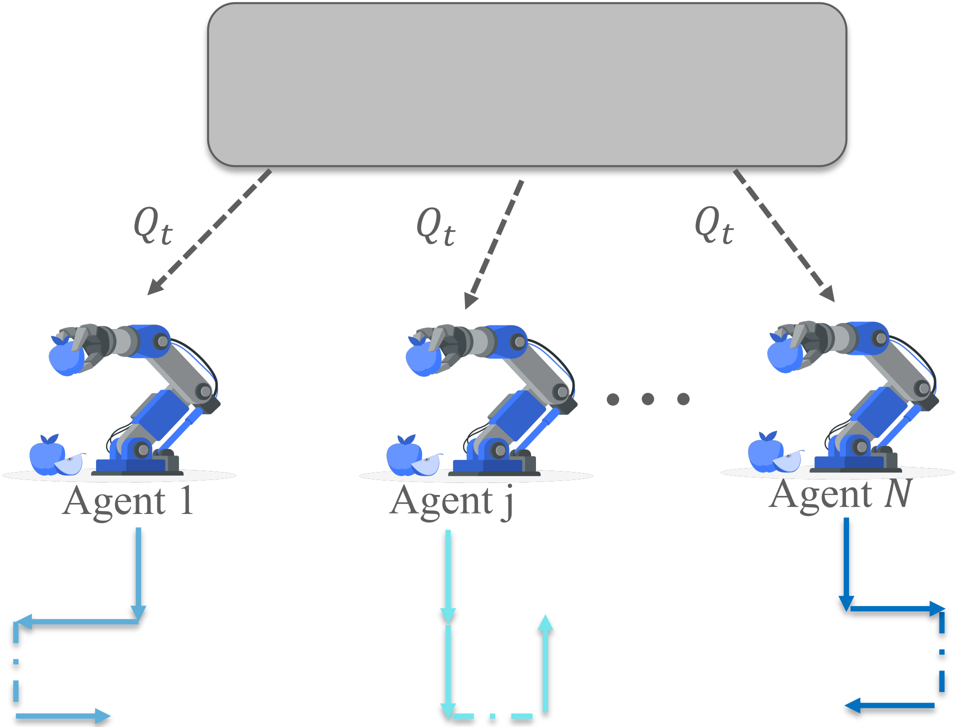
Federated Reinforcement Learning



Federated Reinforcement Learning

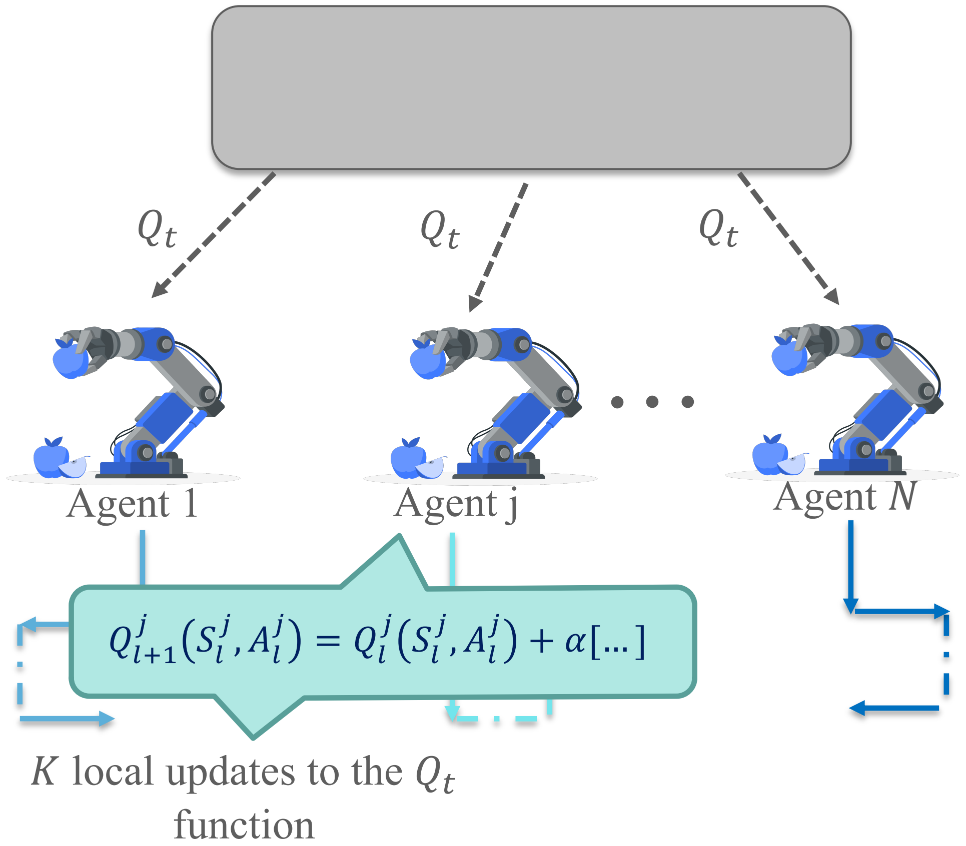


Federated Reinforcement Learning

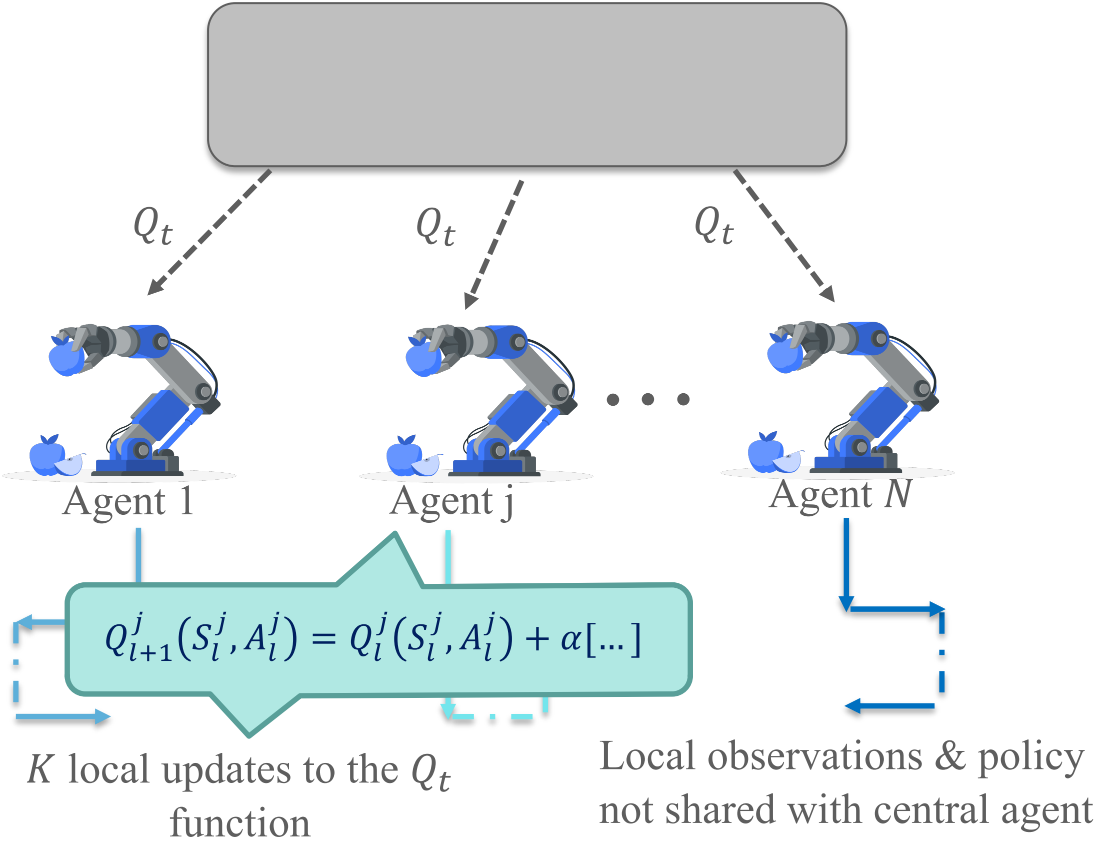


K local updates to the Q_t function

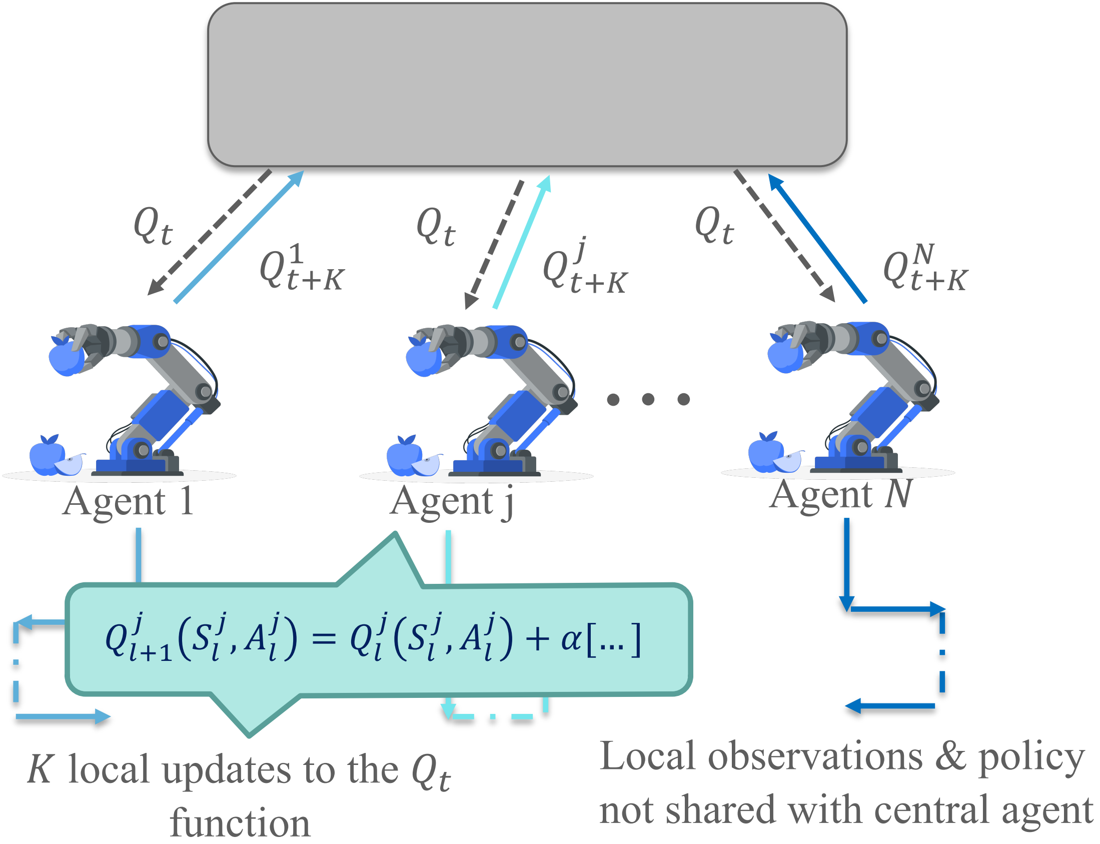
Federated Reinforcement Learning



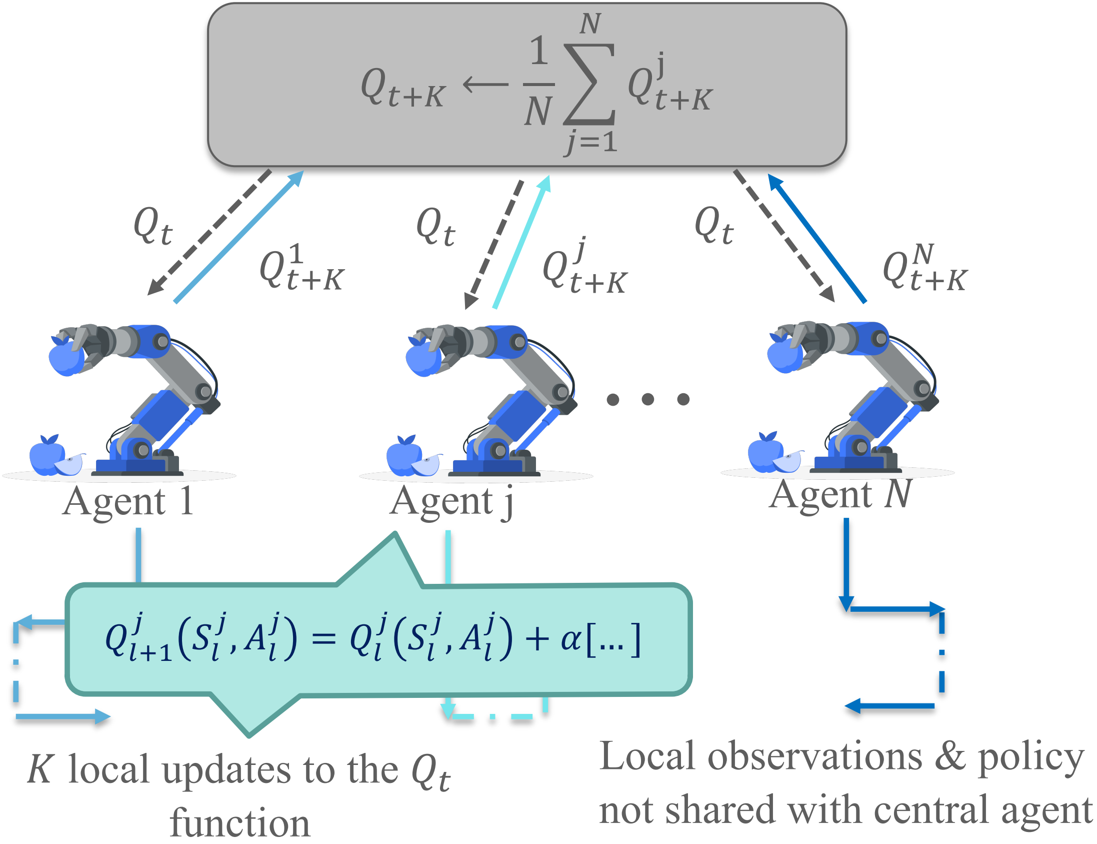
Federated Reinforcement Learning



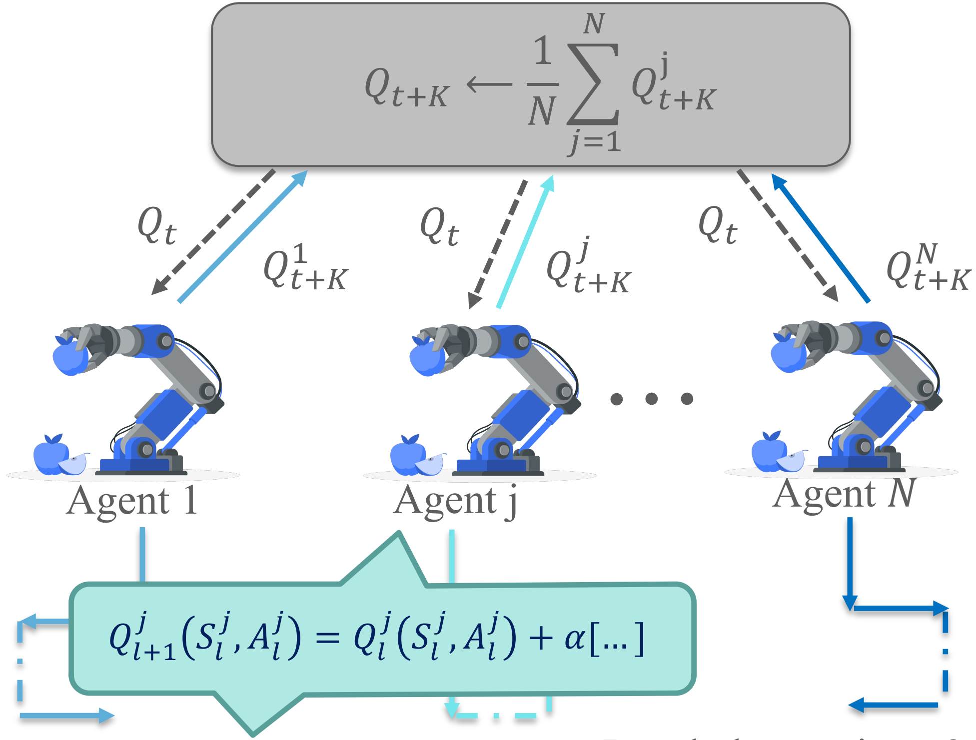
Federated Reinforcement Learning



Federated Reinforcement Learning



Federated Reinforcement Learning

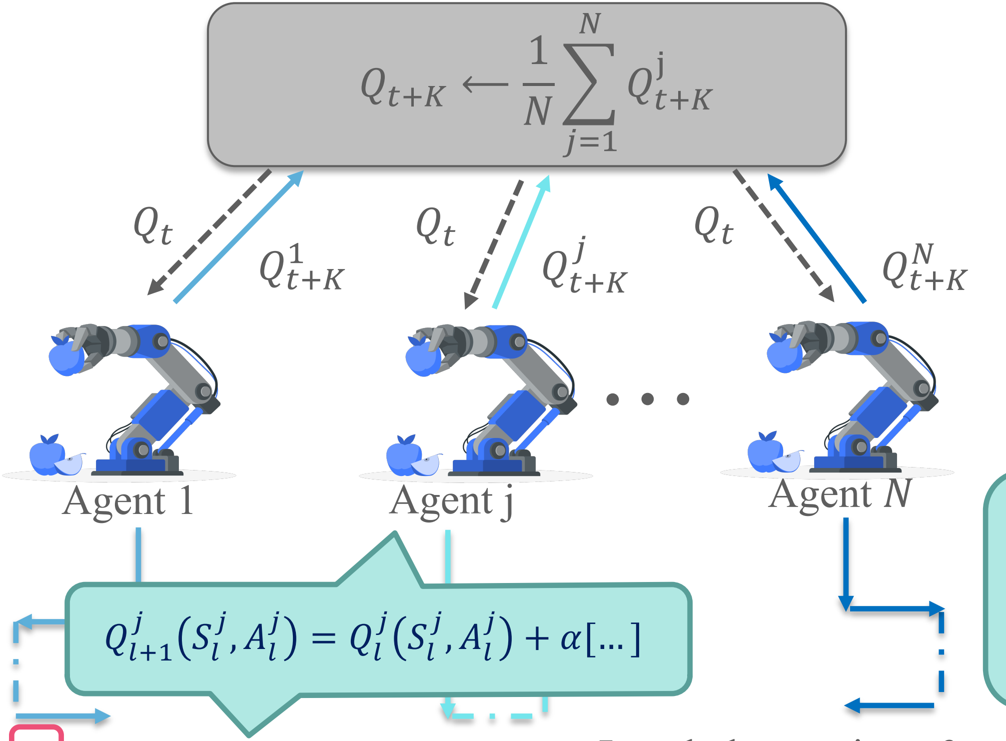


K local updates to the Q_t function

Local observations & policy not shared with central agent

Two issues:
1. Expensive communication
2. Privacy concerns

Federated Reinforcement Learning



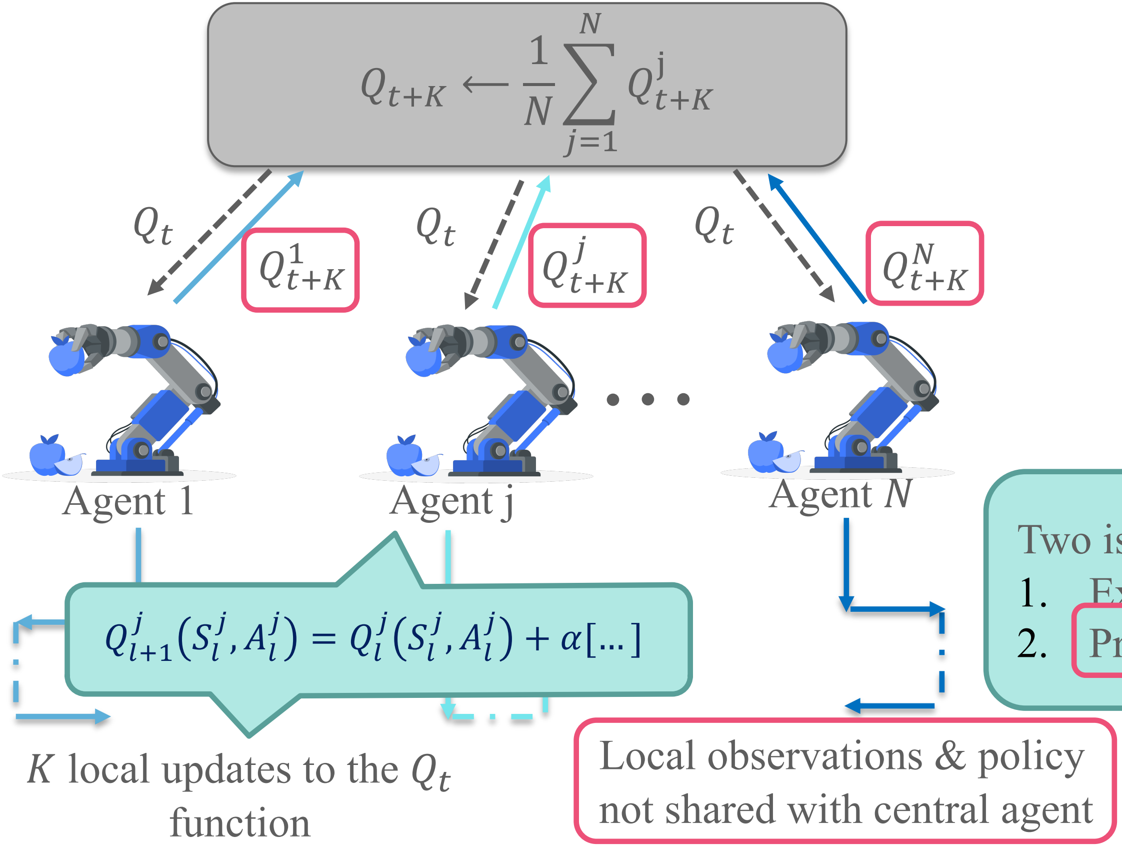
$\boxed{K}$ local updates to the Q_t function

Local observations & policy not shared with central agent

Two issues:

1. Expensive communication ✓
2. Privacy concerns

Federated Reinforcement Learning



Two issues:

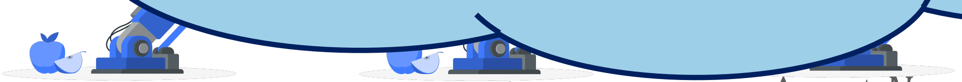
1. Expensive communication ✓
2. Privacy concerns ✓

Federated Reinforcement Learning

N

Question:

Does employing N agents result in N times faster convergence (linear speedup)?



$$Q_{l+1}^{j}(S_l^j, A_l^j) = Q_l^j(S_l^j, A_l^j) + \alpha[\ldots]$$

K local updates to the Q_t function

Local observations & policy not shared with central agent

Two issues:

1. Expensive communication ✓
2. Privacy concerns ✓

Linear Speedup in Federated Learning

- Federated Supervised Learning:

    1. Linear speedup is possible [Spiridonoff, Olshevsky, Paschalidis, NeurIPS ‘21], [….]

Linear Speedup in Federated Learning

- Federated Supervised Learning:

  1. Linear speedup is possible [Spiridonoff, Olshevsky, Paschalidis, NeurIPS ‘21], [….]
  2. Key ingredient in these results: The noise is i.i.d.

Linear Speedup in Federated Learning

- Federated Supervised Learning:

    1. Linear speedup is possible [Spiridonoff, Olshevsky, Paschalidis, NeurIPS ‘21], [….]
    2. Key ingredient in these results: The noise is i.i.d.

$$X_1, X_2, \dots, X_N \overset{\text{i.i.d.}}{\sim} F_X(\cdot) \qquad , \qquad Var(X_i) = \sigma^2$$

Linear Speedup in Federated Learning

- Federated Supervised Learning:

  1. Linear speedup is possible [Spiridonoff, Olshevsky, Paschalidis, NeurIPS ‘21], [….]
  2. Key ingredient in these results: The noise is i.i.d.

$$X_1, X_2, \ldots, X_N \overset{\text{i.i.d.}}{\sim} F_X(\cdot) \qquad , \qquad Var(X_i) = \sigma^2$$

$$\Rightarrow Var\left(\frac{\sum_{i=1}^{N} X_i}{N}\right) = \frac{\sigma^2}{N}$$

Linear Speedup in Federated Learning

- Federated Supervised Learning:

  1. Linear speedup is possible [Spiridonoff, Olshevsky, Paschalidis, NeurIPS ‘21], [….]
  2. Key ingredient in these results: The noise is i.i.d.

$$X_1, X_2, \dots, X_N \overset{\text{i.i.d.}}{\sim} F_X(\cdot) \qquad , \qquad Var(X_i) = \sigma^2$$

$$\Rightarrow Var\left(\frac{\sum_{i=1}^{N} X_i}{N}\right) = \boxed{\frac{\sigma^2}{N}}$$

This is the source of linear speedup

Linear Speedup in Federated Learning

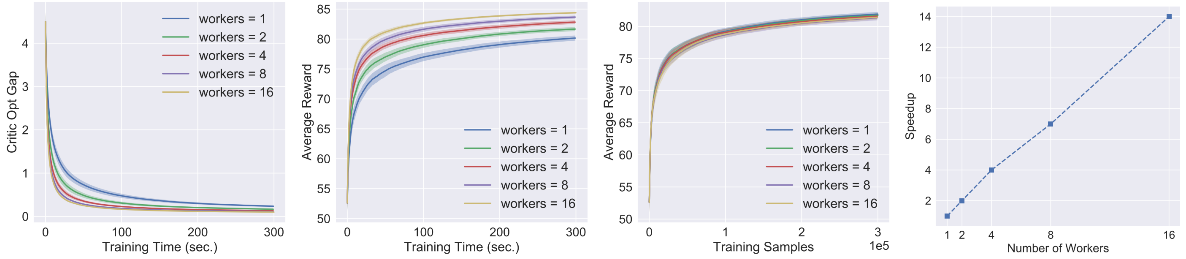
- Federated RL (TD) algorithms
  1. No linear speedup [Wai ’20] [Zeng, Doan, Romberg, ‘20]
     - In fact, they have linear penalty – but their focus is different

Linear Speedup in Federated Learning

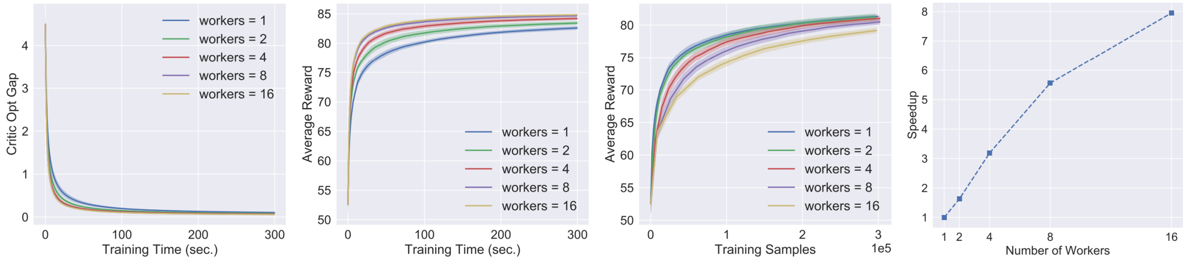
- Federated RL (TD) algorithms

  1. No linear speedup [Wai ’20] [Zeng, Doan, Romberg, ‘20]
     - ❑ In fact, they have linear penalty – but their focus is different
  2. Linear speed up under i.i.d. noise assumption [Shen, Zhang, Hong, Chen ‘20]
     - ❑ Based on experiments, conjectured that linear speedup is possible under Markov noise too

Linear Speedup in Federated Learning (A3C)[1]



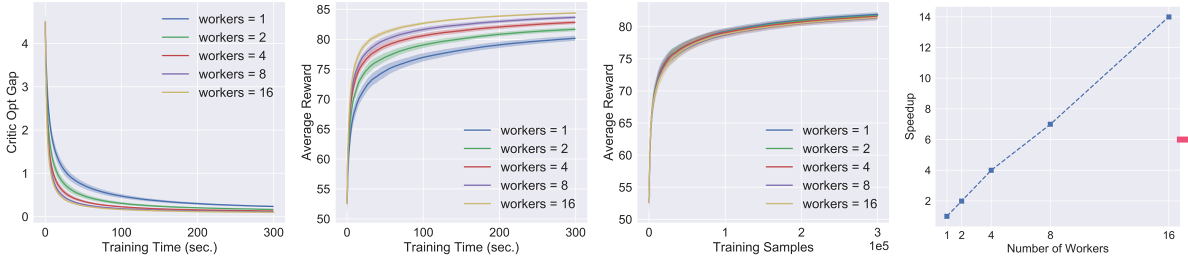
Convergence results of A3C with i.i.d. sampling in synthetic environment.



Convergence results of A3C with Markovian sampling in synthetic environment.

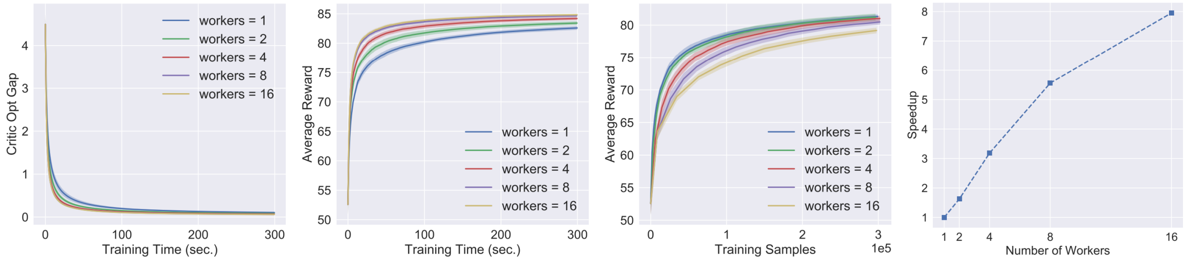
[1]Shen, Han, et al. "Asynchronous advantage actor critic: Non-asymptotic analysis and linear speedup." arXiv preprint arXiv:2012.15511 (2020).

Linear Speedup in Federated Learning (A3C)[1]



Linear speedup is established for i.i.d. noise

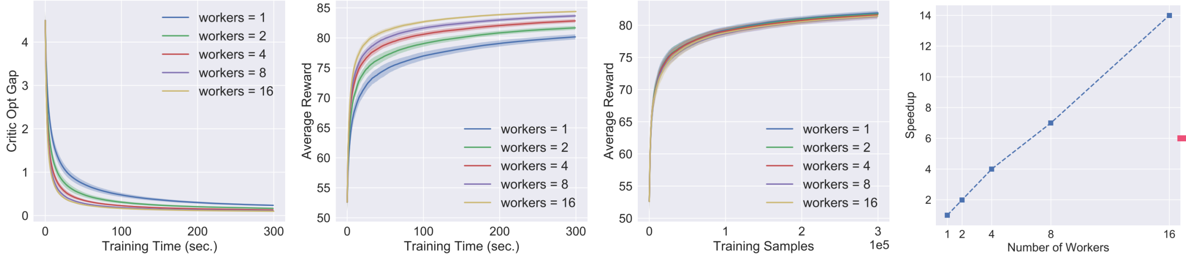
Convergence results of A3C with i.i.d. sampling in synthetic environment.



Convergence results of A3C with Markovian sampling in synthetic environment.

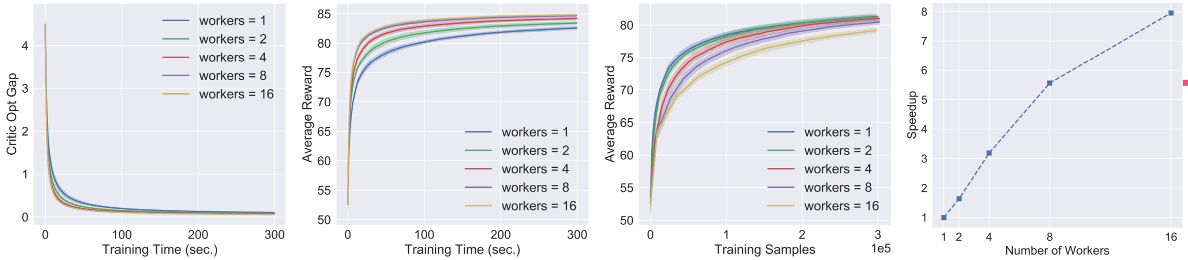
[1]Shen, Han, et al. "Asynchronous advantage actor critic: Non-asymptotic analysis and linear speedup." arXiv preprint arXiv:2012.15511 (2020).

Linear Speedup in Federated Learning (A3C)[1]



Convergence results of A3C with i.i.d. sampling in synthetic environment.

Linear speedup is established for i.i.d. noise

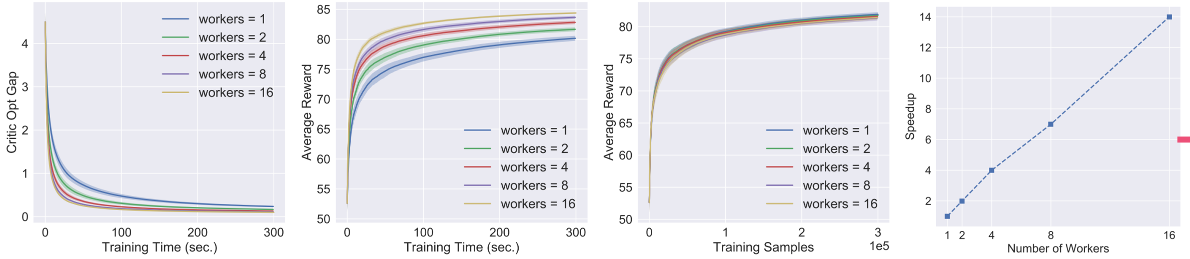


Convergence results of A3C with Markovian sampling in synthetic environment.

A3C paper do not prove a linear speedup in the Markovian setting

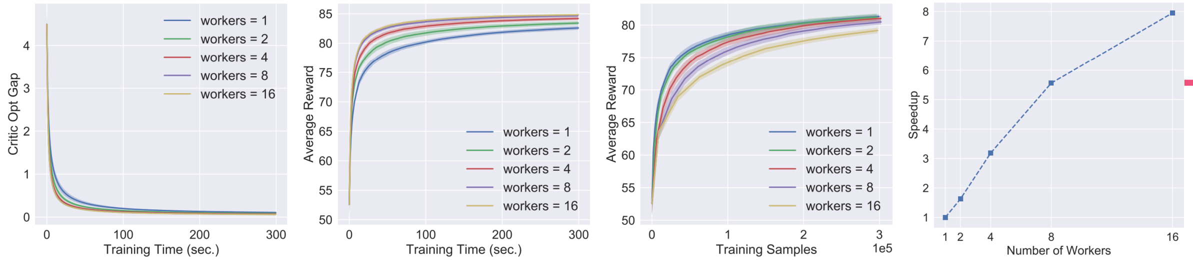
[1]Shen, Han, et al. "Asynchronous advantage actor critic: Non-asymptotic analysis and linear speedup." arXiv preprint arXiv:2012.15511 (2020).

Linear Speedup in Federated Learning (A3C)[1]



Convergence results of A3C with i.i.d. sampling in synthetic environment.

Linear speedup is established for i.i.d. noise



Convergence results of A3C with Markovian sampling in synthetic environment.

A3C paper do not prove a linear speedup in the Markovian setting

We are the first to prove this

[1]Shen, Han, et al. "Asynchronous advantage actor critic: Non-asymptotic analysis and linear speedup." arXiv preprint arXiv:2012.15511 (2020).

Federated Q-learning

Theorem: Let $\mathrm{Q}_T = \frac{1}{N}\sum_{i=1}^{N} \mathrm{Q}_T^i$,

$$\mathbb{E}[\|Q_T - Q^*\|_\infty^2] \leq \tilde{O}\left(\frac{1}{\alpha}(1 - \mathcal{C}\alpha)^T + \frac{\alpha}{N} + (K-1)\alpha^2\right).$$

Federated Q-learning

Theorem: Let $Q_T = \frac{1}{N}\sum_{i=1}^{N} Q_T^i$,

$$\mathbb{E}[\|Q_T - Q^*\|_\infty^2] \leq \tilde{O}\left(\underbrace{\frac{1}{\alpha}(1 - \mathcal{C}\alpha)^T}_{\text{Convergence Bias}} + \frac{\alpha}{N} + (K-1)\alpha^2\right).$$

Federated Q-learning

Theorem: Let $Q_T = \frac{1}{N}\sum_{i=1}^{N} Q_T^i$,

$$\mathbb{E}[\|Q_T - Q^*\|_\infty^2] \le \tilde{O}\Big(\underbrace{\tfrac{1}{\alpha}(1 - \mathcal{C}\alpha)^T}_{\text{Convergence Bias}} + \underbrace{\tfrac{\alpha}{N} + (K-1)\alpha^2}_{\text{Convergence Variance}}\Big).$$

Federated Q-learning

Theorem: Let $Q_T = \frac{1}{N}\sum_{i=1}^{N} Q_T^i$,

$$\mathbb{E}[\|Q_T - Q^*\|_\infty^2] \le \tilde{O}\left(\underbrace{\frac{1}{\alpha}(1-\mathcal{C}\alpha)^T}_{\text{Convergence Bias}} + \underbrace{\overbrace{\frac{\alpha}{N}}^{\text{Linear speedup}} + (K-1)\alpha^2}_{\text{Convergence Variance}}\right).$$

Federated Q-learning

Linear speedup

Theorem: Let $Q_T = \frac{1}{N}\sum_{i=1}^{N} Q_T^i$,

Higher order

$$\mathbb{E}[\|Q_T - Q^*\|_\infty^2] \le \tilde{O}\left(\underbrace{\frac{1}{\alpha}(1-\mathcal{C}\alpha)^T}_{\text{Convergence Bias}} + \underbrace{\frac{\alpha}{N} + \boxed{(K-1)\alpha^2}}_{\text{Convergence Variance}}\right).$$

Federated Q-learning

Linear speedup

Theorem: Let $\mathrm{Q}_T = \frac{1}{N}\sum_{i=1}^{N} \mathrm{Q}_T^i$,

Higher order

$$\mathbb{E}[\|Q_T - Q^*\|_\infty^2] \leq \tilde{\mathcal{O}}\Big(\underbrace{\tfrac{1}{\alpha}(1-\mathcal{C}\alpha)^T}_{\text{Convergence Bias}} + \underbrace{\tfrac{\alpha}{N} + \boxed{(K-1)\alpha^2}}_{\text{Convergence Variance}}\Big).$$

- If $\alpha = \mathcal{O}(\log(NT)\,/T)$ and $K = T/N$, we have $\mathbb{E}[\|Q_T - Q^\pi\|_\infty^2] \leq \epsilon$ within $T = \tilde{\mathcal{O}}\left(\frac{1}{N\epsilon}\right)$iterations.

Federated Q-learning

Linear speedup

Theorem: Let $Q_T = \frac{1}{N}\sum_{i=1}^{N} Q_T^i$,

Higher order

$$\mathbb{E}[\|Q_T - Q^*\|_\infty^2] \leq \tilde{\mathcal{O}}\Big(\underbrace{\tfrac{1}{\alpha}(1-\mathcal{C}\alpha)^T}_{\text{Convergence Bias}} + \underbrace{\tfrac{\alpha}{N} + \boxed{(K-1)\alpha^2}}_{\text{Convergence Variance}}\Big).$$

- If $\alpha = \mathcal{O}(\log(NT)/T)$ and $K = T/N$, we have $\mathbb{E}[\|Q_T - Q^\pi\|_\infty^2] \leq \epsilon$ within $T = \tilde{\mathcal{O}}\left(\frac{1}{N\epsilon}\right)$ iterations.
- Total communication cost = $\frac{T}{K} = N$

Federated Q-learning

Linear speedup

Theorem: Let $Q_T = \frac{1}{N}\sum_{i=1}^{N} Q_T^i$,

Higher order

$$\mathbb{E}[\|Q_T - Q^*\|_\infty^2] \leq \tilde{O}\Big(\underbrace{\tfrac{1}{\alpha}(1-\mathcal{C}\alpha)^T}_{\text{Convergence Bias}} + \underbrace{\tfrac{\alpha}{N} + \boxed{(K-1)\alpha^2}}_{\text{Convergence Variance}}\Big).$$

- If $\alpha = \mathcal{O}(\log(NT)/T)$ and $K = T/N$, we have $\mathbb{E}[\|Q_T - Q^\pi\|_\infty^2] \leq \epsilon$ within $T = \tilde{O}\left(\frac{1}{N\epsilon}\right)$iterations.
- Total communication cost = $\frac{T}{K} = N$

Number of agents

Federated Q-learning

Linear speedup

Higher order

Theorem: Let $Q_T = \frac{1}{N}\sum_{i=1}^{N} Q_T^i$,

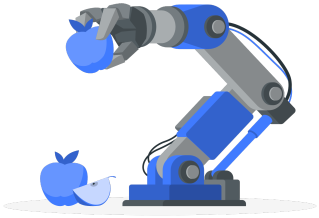
$$\mathbb{E}[\|Q_T - Q^*\|_\infty^2] \le \tilde{\mathcal{O}}\left(\underbrace{\frac{1}{\alpha}(1-\mathcal{C}\alpha)^T}_{\text{Convergence Bias}} + \underbrace{\frac{\alpha}{N} + \boxed{(K-1)\alpha^2}}_{\text{Convergence Variance}}\right).$$

- If $\alpha = \mathcal{O}(\log(NT)\,/T)$ and $K = T/N$, we have $\mathbb{E}[\|Q_T - Q^\pi\|_\infty^2] \le \epsilon$ within $T = \tilde{\mathcal{O}}\left(\frac{1}{N\epsilon}\right)$ iterations.
- Total communication cost = $\frac{T}{K} = N$

Number of agents $\rightarrow$ Constant

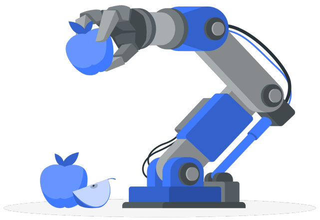
Proof sketch

- Single agent setting



Proof sketch

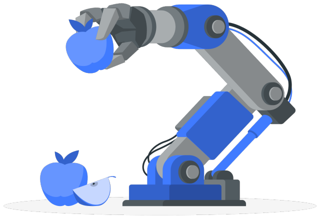
- Single agent setting
- Lyapunov type argument:



Proof sketch

- Single agent setting
- Lyapunov type argument:

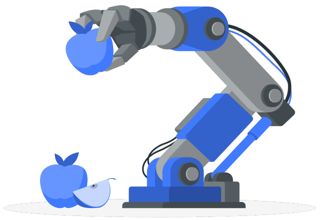
$$\mathbb{E}[\|\theta_{t+1}\|^2] \leq (1-\alpha)\mathbb{E}[\|\theta_t\|^2] + \sigma^2\alpha^2$$



Proof sketch

- Single agent setting
- Lyapunov type argument:

$$\mathbb{E}[\|\theta_{t+1}\|^2] \leq (1-\alpha)\mathbb{E}[\|\theta_t\|^2] + \overbrace{\sigma^2\alpha^2}^{\text{Correspond to variance}}$$

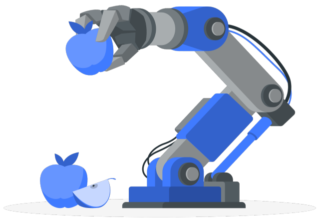


Proof sketch

- Single agent setting
- Lyapunov type argument:

$$\mathbb{E}[\|\theta_{t+1}\|^2] \leq (1-\alpha)\mathbb{E}[\|\theta_t\|^2] + \overbrace{\sigma^2\alpha^2}^{\text{Correspond to variance}}$$

$$\mathbb{E}[\|\theta_T\|^2] \leq (1-\alpha)^T\|\theta_0\|^2 + \alpha$$



Proof sketch

- Single agent setting
- Lyapunov type argument:

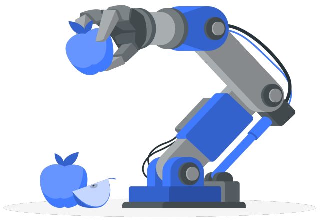
$$\mathbb{E}[\|\theta_{t+1}\|^2] \leq (1-\alpha)\mathbb{E}[\|\theta_t\|^2] + \overbrace{\sigma^2\alpha^2}^{\text{Correspond to variance}}$$

$$\downarrow$$

$$\mathbb{E}[\|\theta_T\|^2] \leq (1-\alpha)^T\|\theta_0\|^2 + \alpha$$

$$\downarrow$$

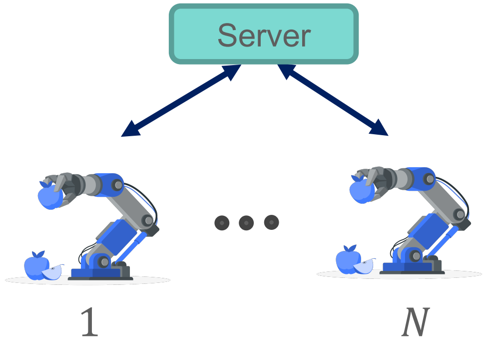
$\tilde{O}(1/\epsilon)$ sample complexity



Proof sketch

- Multiple agents, favorable recursion

$$\mathbb{E}[\|\theta_{t+1}\|^2] \leq (1-\alpha)\mathbb{E}[\|\theta_t\|^2] + \alpha^2/N$$

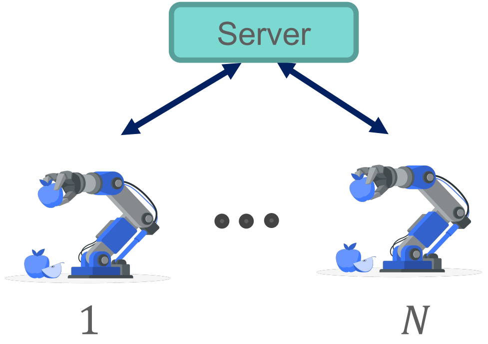


Proof sketch

- Multiple agents, favorable recursion

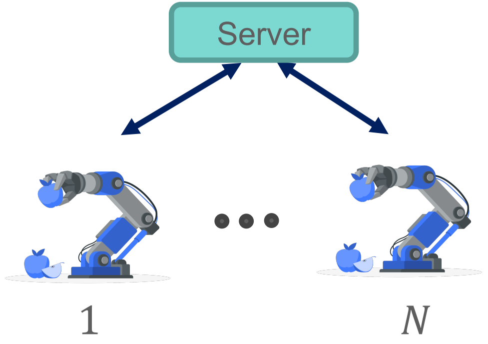
$$\mathbb{E}[\|\theta_{t+1}\|^2] \leq (1-\alpha)\mathbb{E}[\|\theta_t\|^2] + \alpha^2/N$$

$$\mathbb{E}[\|\theta_T\|^2] \leq (1-\alpha)^T\|\theta_0\|^2 + \alpha/N$$



Proof sketch

- Multiple agents, favorable recursion



$$\mathbb{E}[\|\theta_{t+1}\|^2] \leq (1-\alpha)\mathbb{E}[\|\theta_t\|^2] + \alpha^2/N$$

$$\downarrow$$

$$\mathbb{E}[\|\theta_T\|^2] \leq (1-\alpha)^T\|\theta_0\|^2 + \alpha/N$$

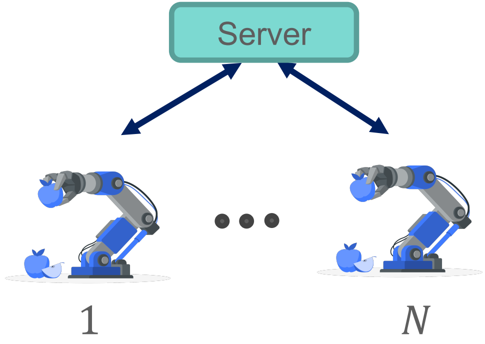
$$\downarrow$$

$\tilde{O}(1/N\epsilon)$ iteration complexity, linear speedup

Proof sketch

- Multiple agents, favorable recursion

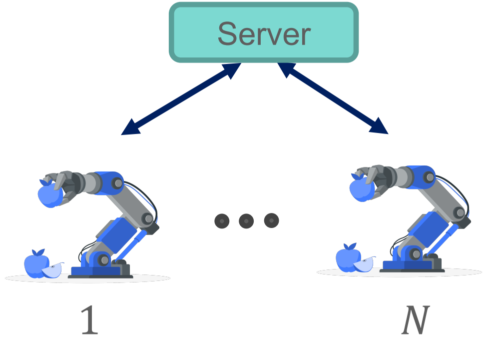
$$\mathbb{E}[\|\theta_{t+1}\|^2] \leq (1-\alpha)\mathbb{E}[\|\theta_t\|^2] + \alpha^2/N$$



- However, we get

$$\mathbb{E}[\|\theta_{t+1}\|^2] \leq (1-\alpha)E[\|\theta_t\|^2] + \frac{\alpha^2}{N} + \alpha^3 + \Omega_t$$

Proof sketch



- Multiple agents, favorable recursion

$$\mathbb{E}[\|\theta_{t+1}\|^2] \leq (1-\alpha)\mathbb{E}[\|\theta_t\|^2] + \alpha^2 / N$$

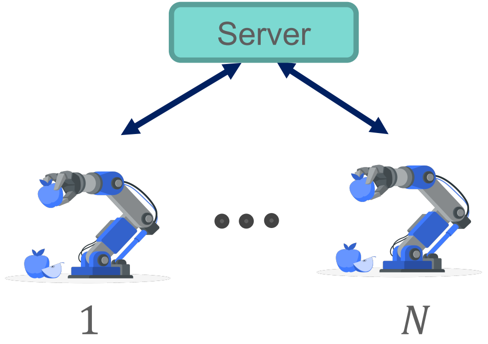
- However, we get

$$\mathbb{E}[\|\theta_{t+1}\|^2] \leq (1-\alpha)E[\|\theta_t\|^2] + \frac{\alpha^2}{N} + \underbrace{\alpha^3}_{\text{Higher order terms}} + \Omega_t$$

Proof sketch

- Multiple agents, favorable recursion

$$\mathbb{E}[\|\theta_{t+1}\|^2] \leq (1-\alpha)\mathbb{E}[\|\theta_t\|^2] + \alpha^2 / N$$

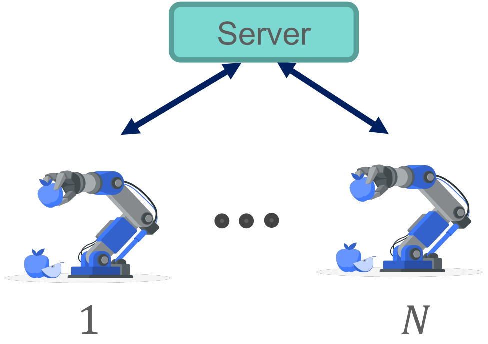


- However, we get

$$\mathbb{E}[\|\theta_{t+1}\|^2] \leq (1-\alpha)E[\|\theta_t\|^2] + \frac{\alpha^2}{N} + \underbrace{\alpha^3}_{\text{Higher order terms}} + \Omega_t$$

Not important

Proof sketch



- Multiple agents, favorable recursion

$$\mathbb{E}[\|\theta_{t+1}\|^2] \leq (1-\alpha)\mathbb{E}[\|\theta_t\|^2] + \alpha^2/N$$

- However, we get

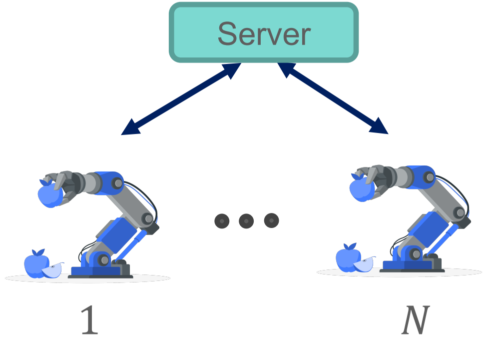
$$\mathbb{E}[\|\theta_{t+1}\|^2] \leq (1-\alpha)E[\|\theta_t\|^2] + \frac{\alpha^2}{N} + \underbrace{\alpha^3}_{\text{Higher order terms}} + \overbrace{\Omega_t}^{\text{Due to local updates}}$$

Not important

Proof sketch

- Multiple agents, favorable recursion

$$\mathbb{E}[\|\theta_{t+1}\|^2] \le (1-\alpha)\mathbb{E}[\|\theta_t\|^2] + \alpha^2/N$$



- However, we get

$$\mathbb{E}[\|\theta_{t+1}\|^2] \le (1-\alpha)E[\|\theta_t\|^2] + \frac{\alpha^2}{N} + \underbrace{\alpha^3}_{\text{Higher order terms}} + \overbrace{\Omega_t}^{\text{Due to local updates}}$$

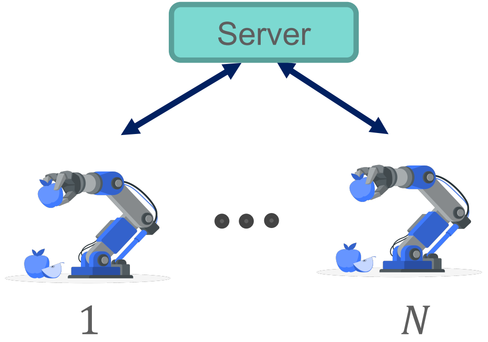
Handled by a special weighting

Not important

Proof sketch

- Multiple agents, favorable recursion

$$\mathbb{E}[\|\theta_{t+1}\|^2] \leq (1-\alpha)\mathbb{E}[\|\theta_t\|^2] + \alpha^2 / N$$



- However, we get

$$\mathbb{E}[\|\theta_{t+1}\|^2] \leq (1-\alpha)E[\|\theta_t\|^2] + \frac{\alpha^2}{N} + \underbrace{\alpha^3}_{\text{Higher order terms}} + \Omega_t$$

Due to local updates

Handled by a special weighting

Not important

$\mathcal{O}(1/N\epsilon)$ iteration complexity, linear speedup

Other results

1. Federated Temporal Difference with Linear Function Approximation, on-policy data

Other results

1. Federated Temporal Difference with Linear Function Approximation, on-policy data
2. Federated Temporal Difference, off-policy data

Other results

1. Federated Temporal Difference with Linear Function Approximation, on-policy data
2. Federated Temporal Difference, off-policy data
3. Federated stochastic approximation with Markovian noise

Other results

1. Federated Temporal Difference with Linear Function Approximation, on-policy data
2. Federated Temporal Difference, off-policy data
3. Federated stochastic approximation with Markovian noise

Linear speedup + Constant communication cost

Other results

1. Federated Temporal Difference with Linear Function Approximation, on-policy data
2. Federated Temporal Difference, off-policy data
3. Federated stochastic approximation with Markovian noise

Linear speedup + Constant communication cost

THANK YOU FOR YOUR ATTENTION!

ICML
International Conference On Machine Learning