Distributed Learning: Sequential Decision Making in Resource-constrained Environments

Udari Madhushani, Naomi Ehrich Leonard
Princeton University

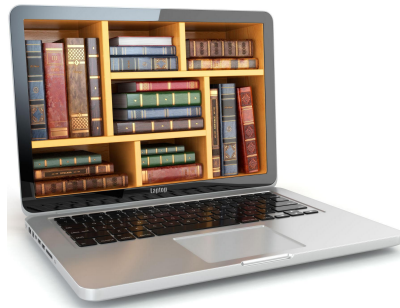April 26, 2020

Sequential Decision Making

Option 3

Option 2

Option 4

Option 1

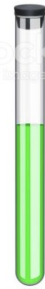Option 5

Online Educational Systems
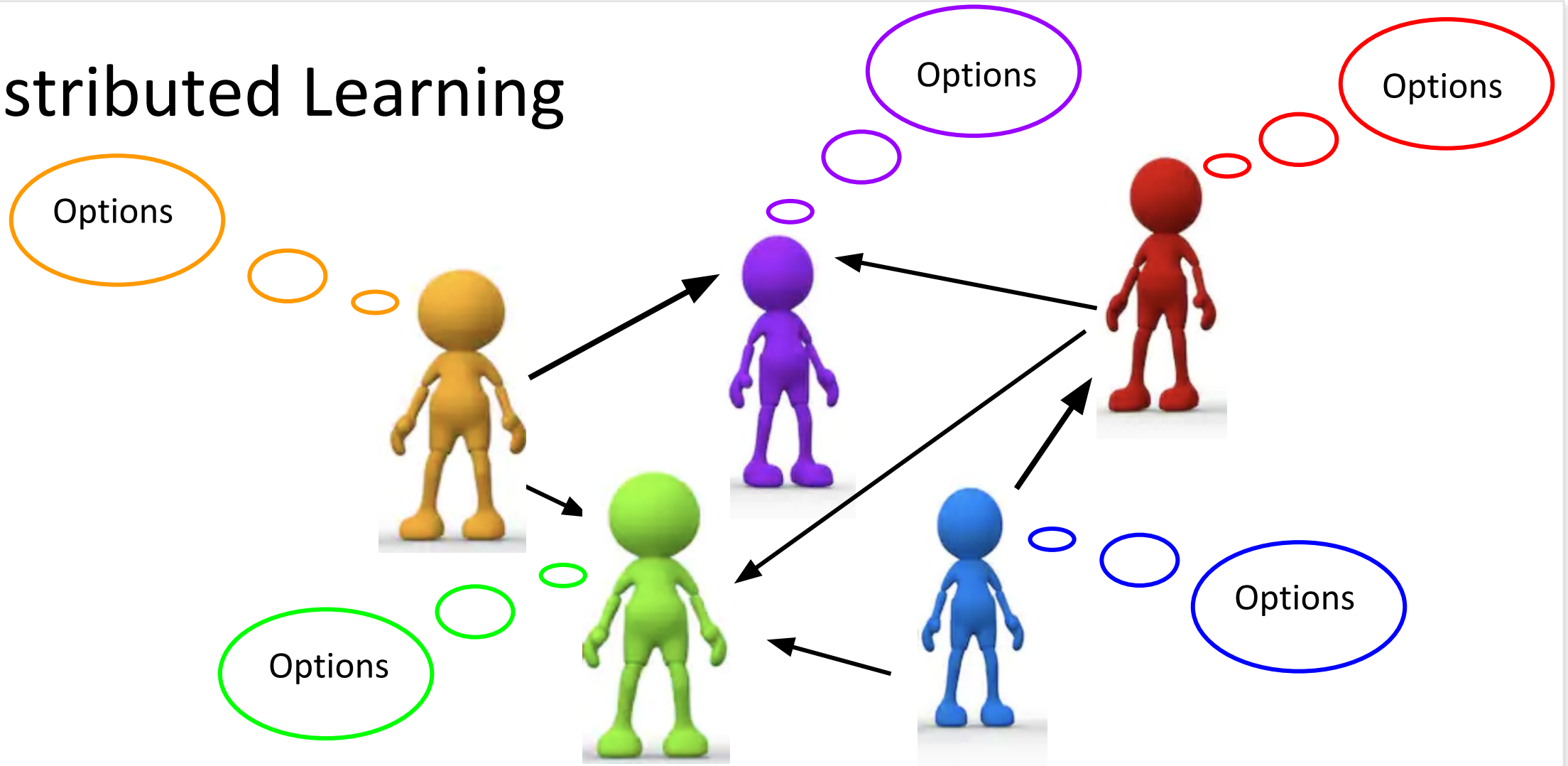
Option 1

Option 2

Option 3

Option 4

Option 5

Clinical Trials

Distributed Learning

Distributed Learning Systems

- The goal of each agent is maximizing the individual reward thereby maximizing the group reward.

- Agents communicate with their neighbors, defined according to a network graph.

- We develop a cost effective partial communication protocol that obtains same order of performance as full communication.

Sampling Rule
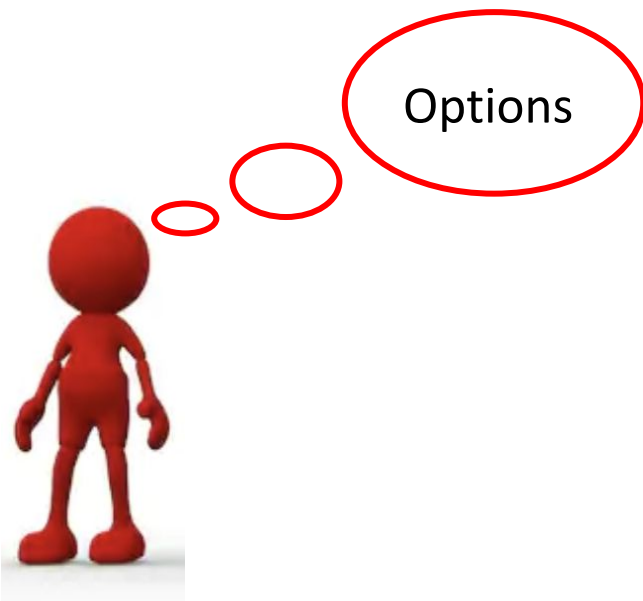
- Estimate the expected reward of options using available information.

$$N_i^k(t) : \text{Number of Observations}$$
$$\hat{\mu}_i^k(t) : \text{Average reward}$$

- An option is chosen based on an explore-exploit strategy.

$$Q_i^k(t) = \hat{\mu}_i^k(t) + C_i^k(t) : \text{Objective function}$$
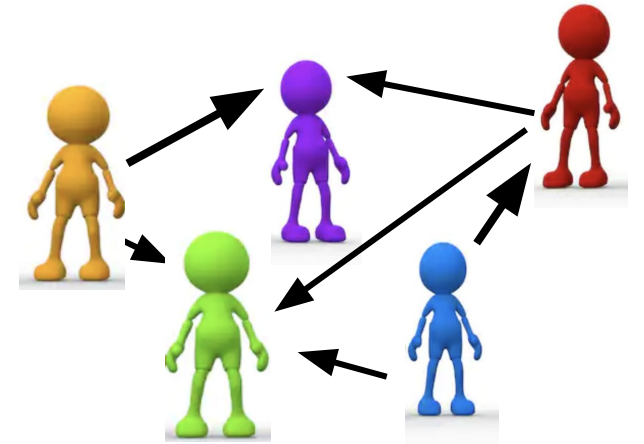
Exploit Explore

$$\varphi_{t+1}^k = arg_i \max Q_i^k(t) : \text{Sampling rule}$$

$$\text{with } C_i^k(t) = \sqrt{\frac{\log t}{N_i^k(t)}}$$

Options

Partial Communication Rules

Agents decide to communicate when they explore

$$\varphi_{t+1}^k = arg_i \max Q_i^k(t) \neq arg_i \max \hat{\mu}_i^k(t)$$

Rule 1:

Agents **observe** the reward values and actions of their neighbors

Rule 2:

Agents **broadcast** their reward values to their neighbors
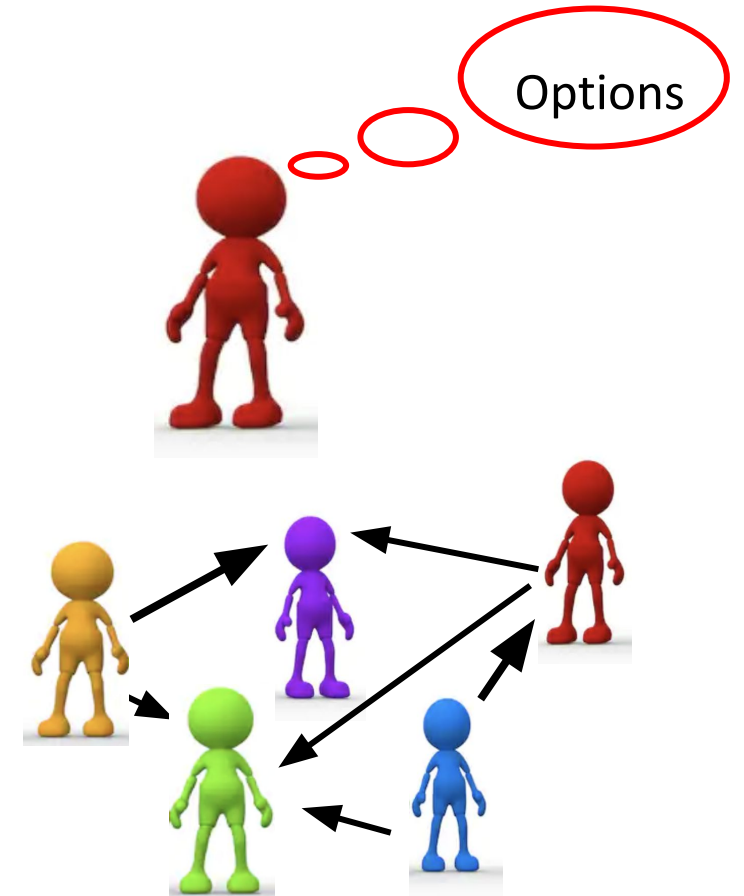
Communication cost is **logarithmic** in time

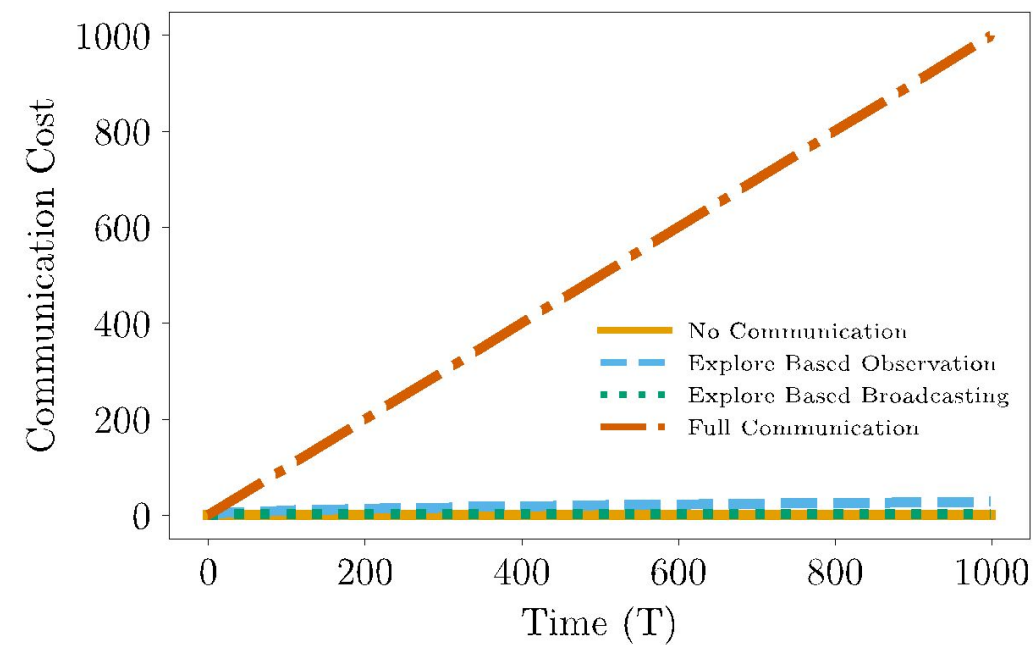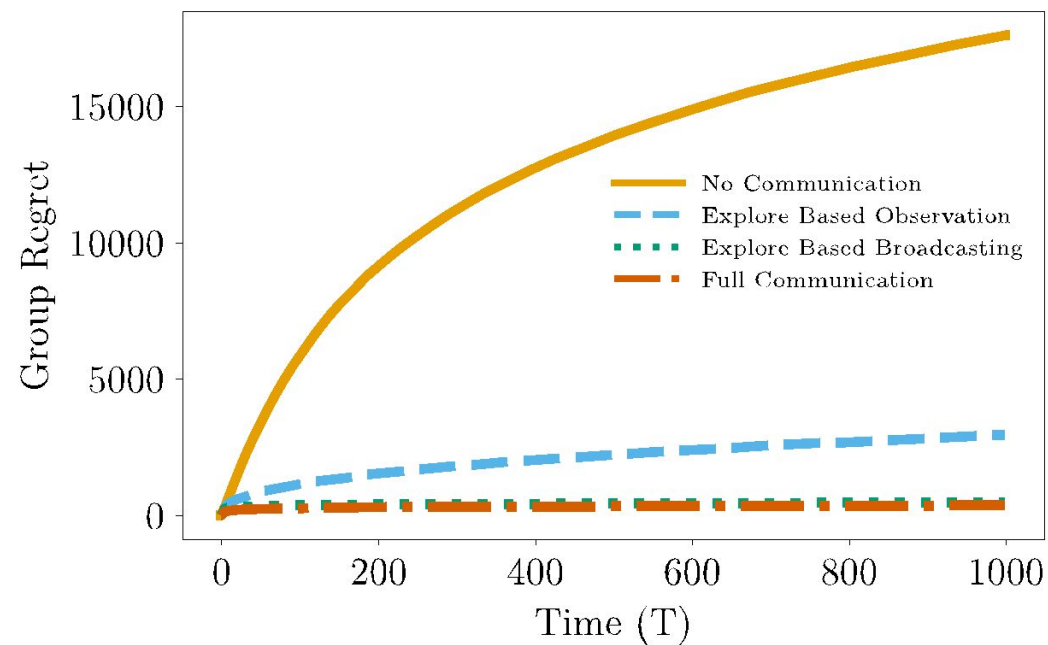# Performance Analysis

Options

Sampling Regret :

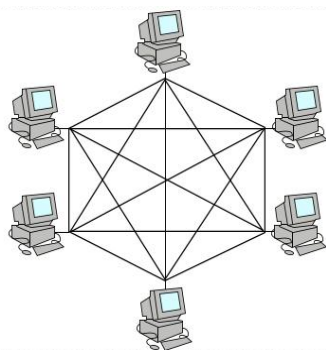Accumulated expected loss due to sampling suboptimal options.

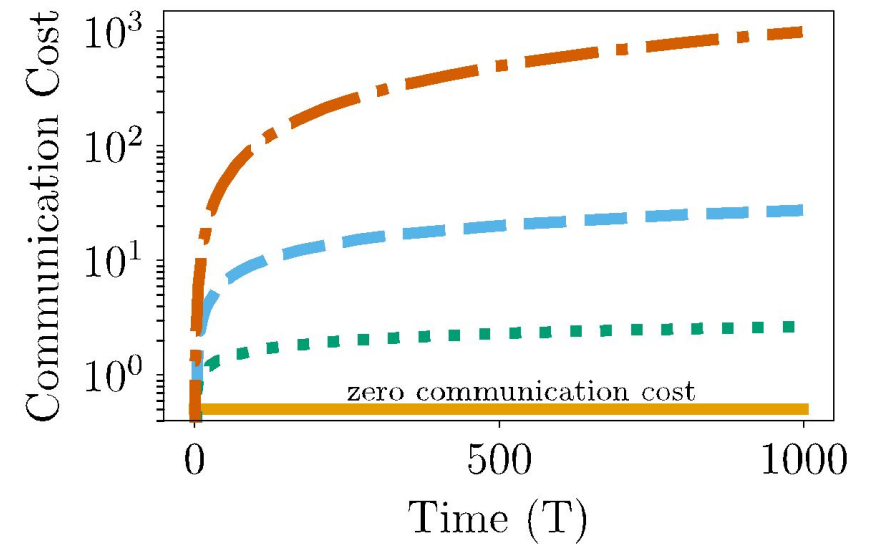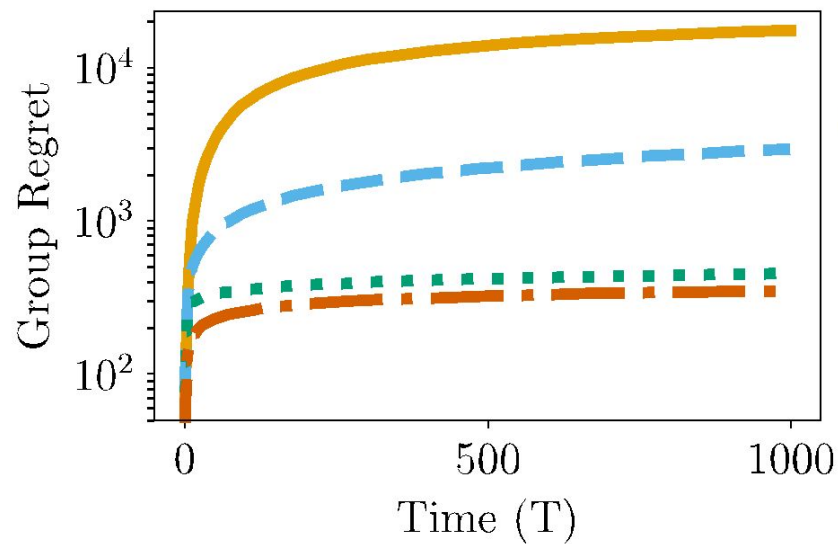Communication Cost :

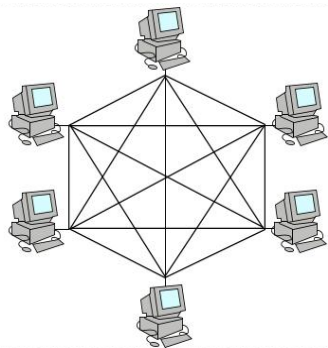Number of times agents communicated.

Complete Graph

Explore based communication provide significant improvement in performance for a small communication regret

Complete Graph

Explore based observation provides same order of performance as full communication while utilizing a incurring a significantly small communication cost

Conclusions

- We propose high performance communication protocols that can be utilized in resource-constrained environments.

- Same order of performance as full communication can be obtained by using the proposed partial communication rule

- Communication cost of under the partial communication rule is logarithmic in time. Communication cost under full communication is linear in time.

Thank You