SAPIENZA
UNIVERSITÀ DI ROMA

Autonomous Networking

Gaia Maselli
Dept. of Computer Science

Today's plan

- How to prepare your presentation

Context

- What kind of network?
- What are the main characteristics?
- What are the main challenges?

Addressed problem

- Routing?
  - the problem of selecting paths to send packets from source(s) to destination(s), while meeting QoS requirements (e.g., latency) and optimizing network resources.
- MAC?
  - The problem of controlling channel access to avoid collisions and optimize energy consumption or
- Adaptability is a key point
  - Is anything changing in the environment?

Goal

- What is the goal of the work?
  - Improve performance?
  - Make a system adaptable?
  - …

Methology

- What methodology is used?
  - Bandit
  - Q-learning
  - Other
- Present the theory

Application of RL

- Application of the RL methodology to the addressed problem
- What does the system want to learn?

RL: agent

- The learner, also called Agent, interacts with its environment and selects its actions to be applied to environment according to its current state and the reinforcement it collects from the environment

➢ Identify the agent

Who is the agent?

- The node?
- A packet?
- A central entity?

Agent collaboration

- **No collaboration**. Either there is a single agent or there are multiple agents, which make decisions only based on their local view.
- **Reactive collaboration**. When agents send feedback to each other
- **Proactive collaboration.** Agents periodically broadcast or send to selected agents information such as their Q-value

RL: action

- Actions
- The objective of the agent is to take actions in order to maximize the global discounted reward, denoted by G_t, it receives over the future.
- Identify the set of actions

RL: reward

- Reward functions
  - How is the reward defined?
  - Is the reward immediate or delayed?

Reward function

- What kind of reward?
- Reward functions can be grouped into two main classes:

1. Test-based reward function
   - They are the simplest form of reward. Reward value takes a constant value depending on test. The most common test is ``is the packet delivered to destination?''. For example, reward equals 1 when a packet is delivered to destination and 0 otherwise.
2. Linear/nonlinear reward function
   - Rewards are function of a metric or a combination of metrics (e.g., a nonlinear reward can depend on the remaining number of hops and the cost of each hop).

RL principles: optimality

- There are different models to address optimality and to define G_t
- **Finite-horizon model** in which the agent should optimize the reward for the next *h* steps
- Finite-horizon model is appropriate when the agent lifetime is known. **Infinite-horizon model** in which the agent should optimize the reward for the long-term run. A discount factor is used to "weight" future rewards. If γ=0 the agent is called `myopic' and it is only concerned by maximizing the immediate reward. As γ approaches 1, the awareness to the future rewards is stronger.

➢ How is optimality addressed?

RL: Model of the environment

- Environment
- The environment model is described by the transition matrix and the reward function
  - Model based approaches: the agent learns the environment model (computes the matrix) and improves its policy to reach optimality
  - Model-free approaches: the agent improves its policy without a priory knowledge of the environmental model (without requiring a transition probability matrix)

➢ Is learning model-free or model-based?

RL principles: exploration/exploitation

- The learner tries to improve the current solution while switching between exploration and exploitation of the solution space.
- Heuristics:
  - Greedy strategy
  - E-greedy strategy
  - Interval-based strategies
  - Probability-distribution based strategies
- The choice of exploration/exploitation technique has great impact on the speed of convergence to optimality of the learning process

convergence

- Is convergence studied/evaluated?