Autonomous Networking

Gaia Maselli
Dept. of Computer Science

Today's plan

- Q-learning based MAC for sensor networks
- Practical exercises

ALOHA protocol

- Contention based protocol
- Time is slotted
- Each node randomly transmits in a slot
- Framed slotted aloha groups slots into frames
- Is it possible an intelligent transmission strategy to avoid as much as possible collisions?
- **Goal: Can nodes find unique transmission slots in a distributed manner ?**

ALOHA and Q-learning: ALOHA-Q

- ALOHA-Q divides time into repeating frames where a certain number of slots are included in each frame for data transmission
- Each slot is initiated with a Q-value to represent the willingness of this slot for reservation, which is initialised to 0 on start-up
- Upon a transmission, the Q-value of corresponding slot is updated, using the Q-learning update rule

$$Q_{t+1}(i, s) = Q_t(i, s) + \alpha\big(R - Q_t(i, s)\big)$$

- where *i* indicates the present node, *s* is the slot identifier, *R* is the current reward and α is the learning rate

ALOHA and Q-learning: ALOHA-Q

- Upon a successful transmission, R takes a value of r = +1 which constitutes a reward
- Upon a failed transmission, R takes a punishment value of p = -1
- Nodes always select the slots with maximum Q-values
- Nodes are restricted to access only one slot per frame for their generated packets and they can use multiple slots in a frame for relaying the received packets

Example

- Updating the Q-values for 10 slots per frame
- A node is allowed to send a maximum of 3 packets in each frame.

Q values

Success

Failure

0	0	0	0	0	0	0	0	0	0	Frame 1
0.1	0	0	0	0.1	0	0	0	0	0	Frame 2
0.19	0	0	0	-0.01	0	0	0	0	-0.1	Frame 3
0.271	0	-0.1	0	-0.01	0	0.1	0	0	-0.1	Frame 4
0.343	0	-0.1	0	-0.01	0	0.19	0	0.1	-0.1	Frame 5

....

0.999	0	-0.1	0	-0.01	0	0.999	0	0.999	-0.1	Frame N

Aloha-Q with decreasing-Є greedy method: Aloha-Q-DEPS

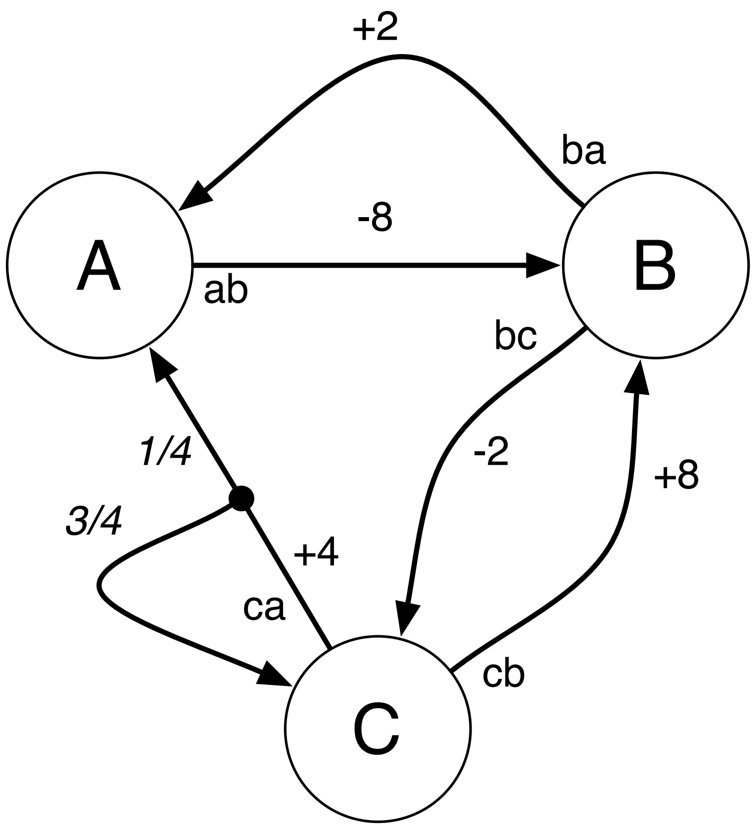
- A decreasing-Є method is developed to allow nodes to explore more until they achieve a certain level of exploration

$$Є = \begin{cases} 1 - Q_{value} & \text{before convergence} \\ 1 - Q_{convergence} & \text{after convergence} \end{cases}$$

- In ALOHA-Q, the term convergence in a slot occurs when the Q-value of this slot approaches to 1
- $Q_{convergence} = 0.9$

Exercise

- Consider the MDP with discount factor $\gamma=0.5$
- A, B, C are states
- *ab, bc, ba, ca, cb,* represent actions
- Signed integers represents rewards
- Fractions represent transition probabilities
- Define the state-value function



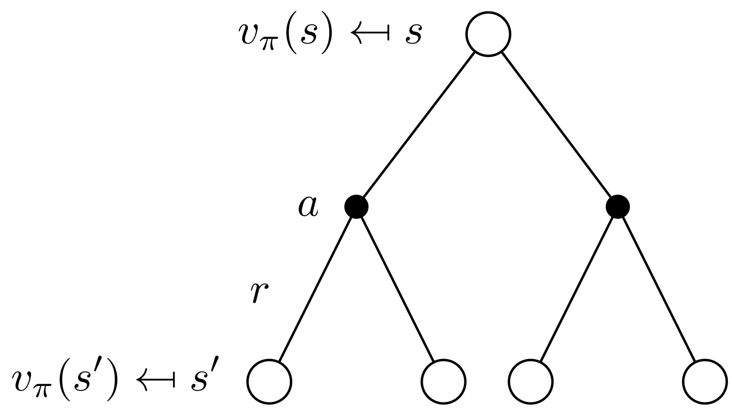
Question 1

- Define the state-value function Vπ(s) for a discounted MDP

$$v_\pi(s) = \mathbb{E}_\pi\left[R_{t+1} + \gamma v_\pi(S_{t+1}) \mid S_t = s\right]$$

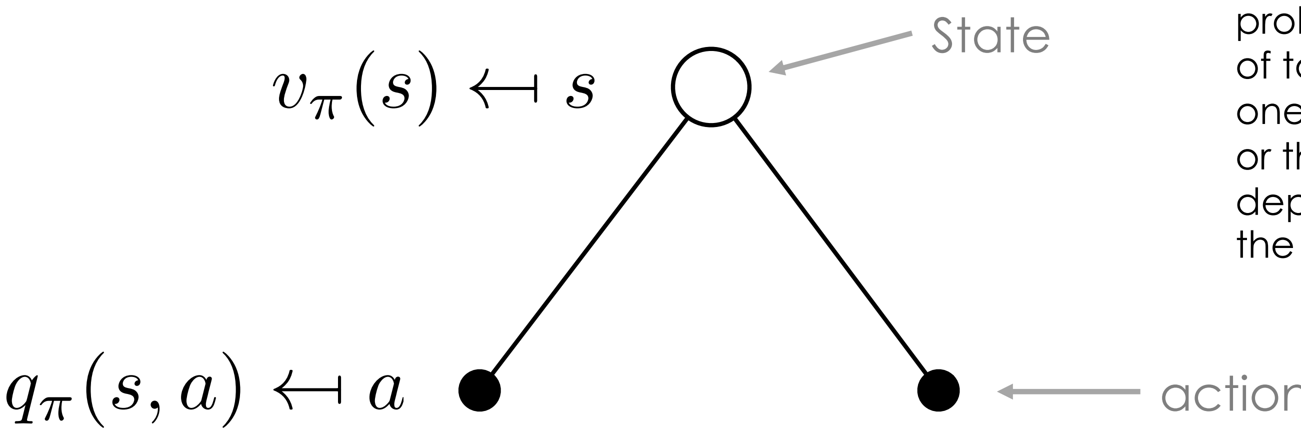
Question 2

- Write down the Bellman exectation equation for state-value functions



$$v_\pi(s) = \sum_{a \in \mathcal{A}} \pi(a|s) \left(\mathcal{R}_s^a + \gamma \sum_{s' \in \mathcal{S}} \mathcal{P}_{ss'}^a v_\pi(s') \right)$$

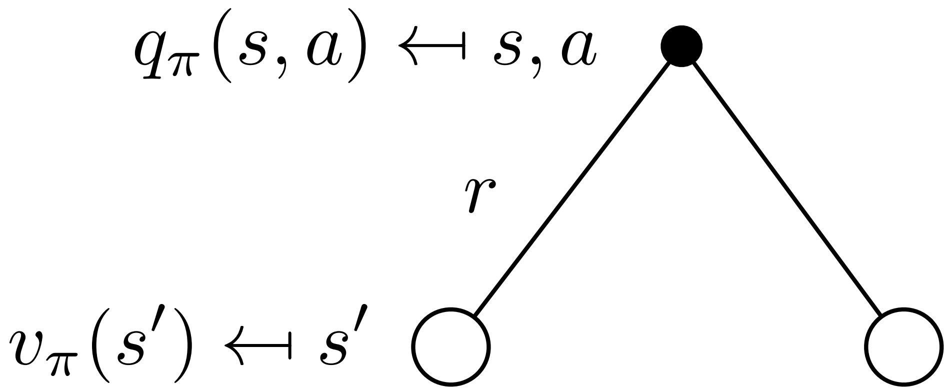
Bellman Expectation Equation for V^{π}



The probability of taking one action or the other depends on the policy

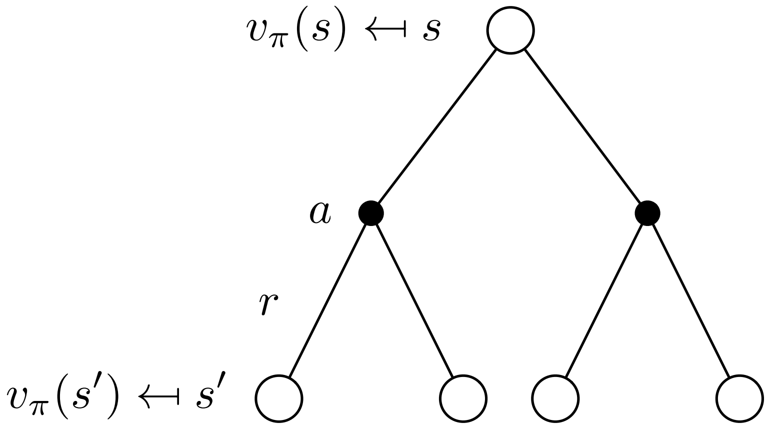
$$v_\pi(s) = \sum_{a \in \mathcal{A}} \pi(a|s) q_\pi(s,a)$$

Bellman Expectation Equation for Q^{π}



$$q_\pi(s,a) = \mathcal{R}_s^a + \gamma \sum_{s' \in \mathcal{S}} \mathcal{P}_{ss'}^a v_\pi(s')$$

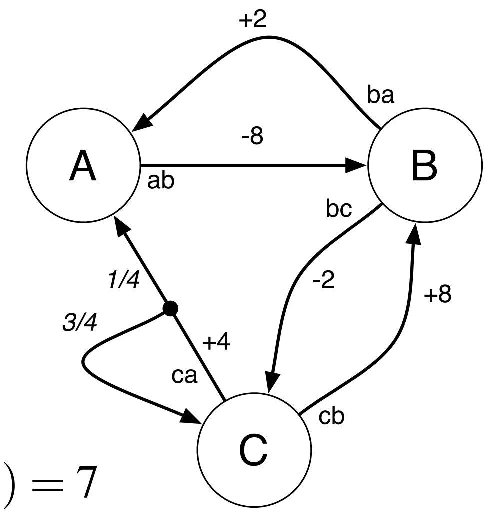
Bellman Expectation Equation for v_π (2)



$$v_\pi(s) = \sum_{a \in \mathcal{A}} \pi(a|s) \left(\mathcal{R}_s^a + \gamma \sum_{s' \in \mathcal{S}} \mathcal{P}_{ss'}^a v_\pi(s') \right)$$

Question 3

- Consider the uniform random policy $\pi_1(s,a)$ that takes all actions from state *s* with equal probability. Starting with an initial value function of $V_1(A) = V_1(B) = V_1(C) = 2$ apply one synchronous iteration of iterative policy evaluation to compute a new value function $V_2(s)$
- $V_2(A)$=?, $V_2(B)$=?, $V_2(C)$=?



$$V_2(A) = -8 + 0.5V_1(B) = -7$$

$$V_2(B) = 0.5(2 + 0.5V_1(A)) + 0.5(-2 + 0.5V_1(C)) = 1$$

$$V_2(C) = 0.5(8 + 0.5V_1(B)) + 0.5(4 + 0.5(1/4V_1(A) + 3/4V_1(C))) = 7$$