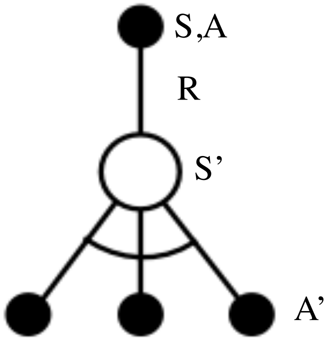
Autonomous Networking

Gaia Maselli
Dept. of Computer Science

Today’s plan

- Q-learning based Link Layer

Q-Learning Control Algorithm



$$Q(S,A) \leftarrow Q(S,A) + \alpha \left(R + \gamma \max_{a'} Q(S',a') - Q(S,A) \right)$$

- Q-learning control converges to the optimal action-value function, $Q(s, a) \rightarrow q_*(s, a)$

Q-learning algorithm

Q-learning (off-policy TD control) for estimating $\pi \approx \pi_*$

Algorithm parameters: step size $\alpha \in (0, 1]$, small $\varepsilon > 0$
Initialize $Q(s, a)$, for all $s \in \mathcal{S}^+, a \in \mathcal{A}(s)$, arbitrarily except that $Q(terminal, \cdot) = 0$

Loop for each episode:
    Initialize S
    Loop for each step of episode:
        Choose A from S using policy derived from Q (e.g., ε-greedy)
        Take action A, observe R, S'
        $Q(S, A) \leftarrow Q(S, A) + \alpha\big[R + \gamma \max_a Q(S', a) - Q(S, A)\big]$
        $S \leftarrow S'$
    until S is terminal

An Adaptive Link Layer for Heterogeneous Multi-Radio Mobile Sensor Networks

Wireless sensor networks

- Energy efficiency is a main goal
- Radios
  - Xtend and the XE1205 radios are designed for low-bitrate long-range communication over distances of **a mile or more**
  - 802.11 and CC2420 radios enable high and low bandwidth communication, respectively, over short ranges of **hundreds of feet or less**.
- Critical choice
  - a long-range radio enables nodes to communicate over long distances but at the expense of expending more power
  - a shorter range radio that is more power-efficient but forego communication over longer distances
- It is possible to choose a long range radio and use lower power settings for short range communication, but doing so is far less efficient than using a short range radio for communicating over shorter distances.
- Using a radio at its maximum range is never desirable, as packet loss rates increase with distance

Multi-radio choice

- Two complementary radios with heterogeneous range characteristics to enable mobile sensor nodes the ability to achieve a significantly greater range diversity at a lower total energy cost when compared to a single radio
- The key idea is to operate each radio over a range where it is more energy efficient and to switch to the other radio whenever a mobile node moves from one radio's effective range to another.
- Proposal: A novel reinforcement learning-based link layer algorithm that continually learns channel characteristics and dynamically decides when to switch between radios depending on
  - Current interference
  - Current communication needs

Try to give a solution

- A sensor node with two radios
- Q-learning to adaptively switch from one radio to the other based on channel conditions
- Goal: minimize energy consumption

- Think about:
  - Where is the agent implemented?
  - What are the states?
  - What is the reward?

Q-Learning based Link Layer

- The algorithm needs to continually monitor and “learn” channel characteristics for the two radios and determine which one provides the lowest energy communication channel
- a reinforcement-learning based algorithm that enables adaptation across radios with different power/range trade-offs
- The algorithm learns the characteristics of radio channels through exploration and continually adapts to use the more efficient one
- It uses simple link layer statistics to dynamically choose a radio

Model

- First case: each radio is set to a single power level
- Q-Learning uses a ***two-state model*** (one for each radio) where the action taken by the agent is either to stay with the same radio or switch to the alternate radio
- The agent will switch radios if conditions deteriorate on the current radio (or is disconnected), or if conditions improve on the alternate radio
- The two state model may be expanded to an ***n-state model***, where each state represents a radio at a particular transmit power level, each representing a particular range/power trade- off
  - Example: four states would be required for two radios, each with two transmit power level options

How can we estimate channel conditions?

- If there is interference →
  - Transmission will fail frequently (no acks)
  - Frequent backoff
- Every protocol at the MAC layer retries to transmit a packet for a number of times if it does not receive an acknowledgment
- The more the node retransmit, the more the energy consumed

Reward

- The reward is modelled as an estimate of the amount of energy associated with the channel metrics collected for a given packet
- The amount of energy to transmit a given packet is a function of
  - packet size
  - static radio parameters such as receive/transmit power
  - number of retransmission attempts
  - number of congestion backoffs
- Energy is a cost, rather than a reward, so its value is negative

Reward

- If a packet is *successfully* transmitted, after *i retransmissions*

Energetic cost of
i retransmissions

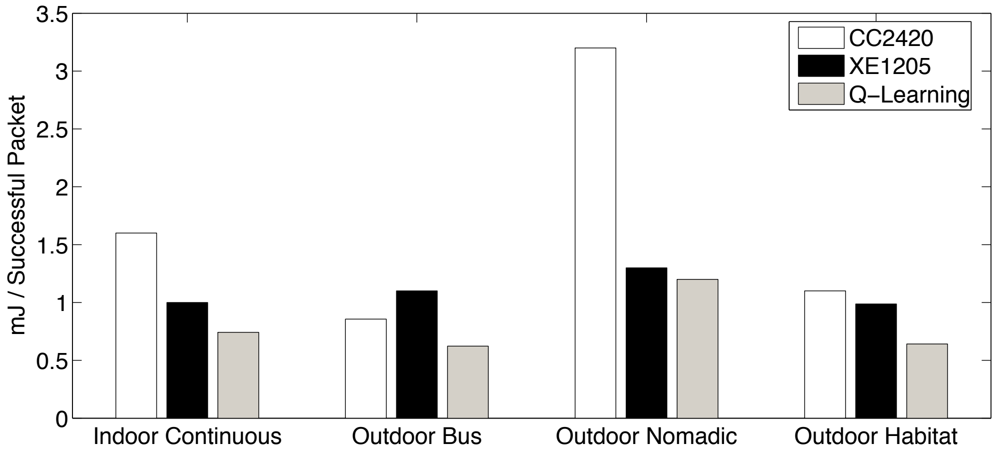
$$\begin{aligned} r[i] \quad = \quad & -\boxed{(i \cdot (PacketSize \cdot ByteTime \cdot TxPower + \\ & AckTimeOut \cdot RxPower)} \\ & + RxPower \cdot (AckRTT) \\ & + PacketSize \cdot ByteTime \cdot TxPower) \end{aligned}$$

- If a packet is *not* transmitted after a pre-defined maximum number of retries
  - a large negative reward is given to encourage the algorithm to switch to a higher power state sooner, thereby limiting the number of lost packets

Policy

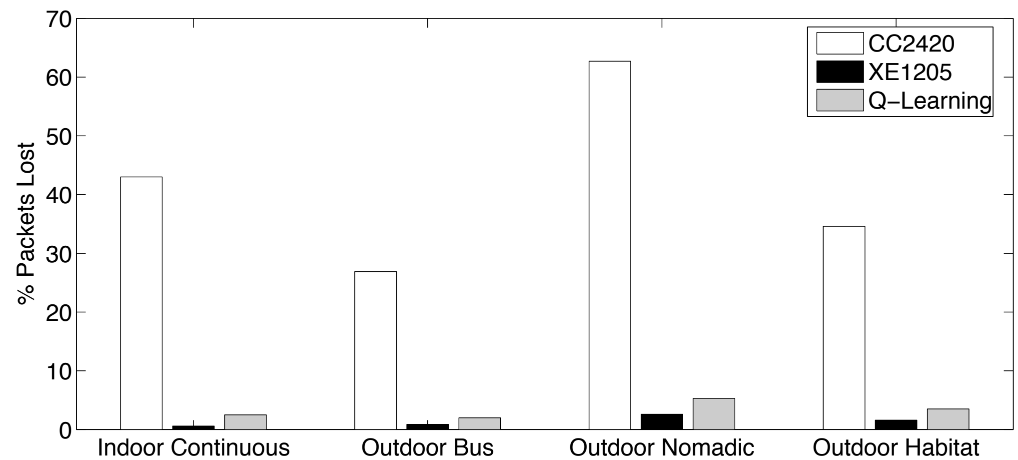
- choose an action a from the set of actions with the maximal Q value or the minimal expected energy consumption for a packet transmission
- P=(1-ε) → choose action with highest Q value
- P=ε → choose a random action

Results



Energy consumed per successful packet for each dataset and strategy

Percent Packets lost for the two radio interfaces and Q-Learning implementation



Readings

- J. Gummeson, D. Ganesan, M. D. Corner and P. Shenoy, "An adaptive link layer for heterogeneous multi-radio mobile sensor networks," in *IEEE Journal on Selected Areas in Communications*, vol. 28, no. 7, pp. 1094-1104, September 2010