Optimal Demand Response Using Device Based Reinforcement Learning

Zheng Wen [1]
Joint Work with Hamid Reza Maei [1] and Dan O'Neill [1]

[1]Department of Electrical Engineering
Stanford University
zhengwen@stanford.edu

June 12, 2012

Outline

Motivation

- Demand response (DR) systems dynamically adjust electrical demand in response to changing electricity prices (or other grid signals)
  - By suitably adjusting electricity prices, load can be shifted from "peak" periods to other periods
  - DR is a key component in smart grid
- DR can potentially
  - Improve operational efficiency and capital efficiency
  - Reduce harmful emissions and risk of outages
  - Better match energy demand with unforecasted changes in electrical energy generation

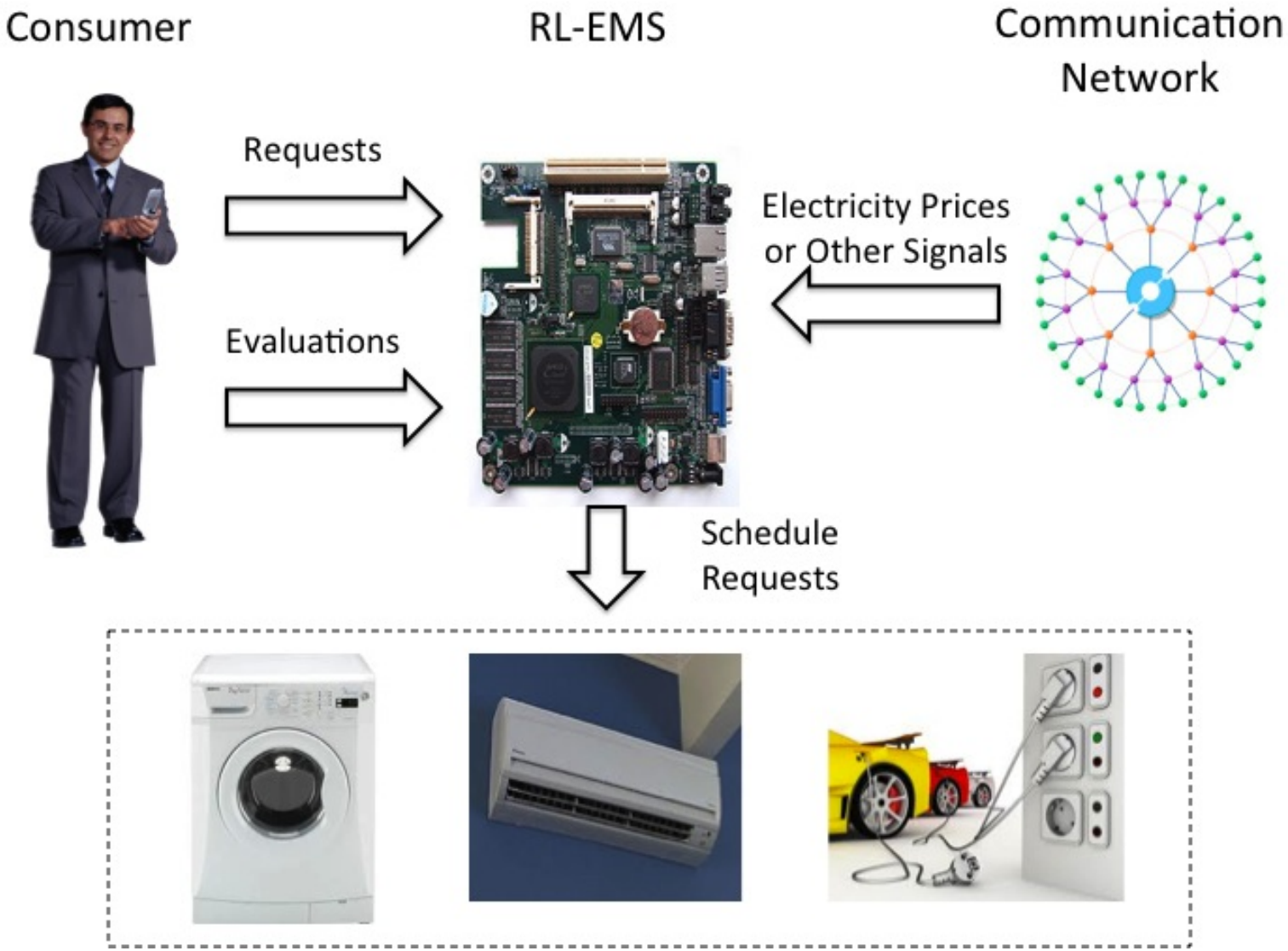
Motivation (Cont...)

- DR has been extensively investigated for larger energy users
- Residential and small building DR offers similar potential benefits
- However, "decision fatigue" prevents residential consumers to respond to real-time electricity price (see O'Neill et al. 2010)
  - We can't stay at home every day, watching the real-time electricity price and deciding when to use each device
- Fully-automated Energy Management Systems (EMS) are a necessary prerequisite to DR in residential and small building settings

Energy Management Systems

- EMS makes the DR decisions for the consumer
- Critical to a successful DR EMS approach is learning the consequences of shifting energy consumption on consumer satisfaction, cost and future energy behavior
- O'Neill et al. 2010 has proposed a residential EMS algorithm based on reinforcement learning (RL), called CAES algorithm
- In this project, we extend the work of O'Neill et al. 2010 and propose a new RL-based energy management algorithm called RL-EMS algorithm

RL-EMS System



What is New?

- Compared with CAES, RL-EMS is new in the following aspects:
  1. RL-EMS is allowed to perform speculative jobs
  2. A consumer request's *target time* can be different from its *request time*
  3. An unsatisfied consumer request can be cancelled by the consumer
  4. RL-EMS learns the dis-satisfaction of the consumer on completed jobs and cancelled delays
- New Insight: A device-centric view of the problem

Outline

RL-EMS and Consumer Requests

- Our proposed RL-EMS performs the following functions:
  1. It receives requests from the consumer, and then schedules when to fulfill the received requests (requested jobs)
  2. If a device is idle, RL-EMS could speculatively power on that device (speculative job)
- Each consumer request is a four-tuple $J = (\text{requested device } n, \text{requested time } \tau_r, \text{target time } \tau_g, \text{priority } g)$
  - Jobs are standardized and can be completed in one time step
  - Require $\tau_r \leq \tau_g \leq \tau_r + W(n)$, where $W(n)$ is a known device-dependent time window
  - At each time, there is at most one unsatisfied request for each device

Consumer Preference and Dis-utiltiy Function

- In economics, *utility function* is used to model the preference of consumers
- In this project, we work with the negative of utility function, called *dis-utility function*, which models the dissatisfaction of the consumer
- Assumption 1: the dis-utility of the consumer at time t = the sum of his evaluations on jobs completed/cancelled at time t
  - Dis-utility is additive over jobs
  - At each time there is at most one job completed/cancelled at a device, hence, dis-utility is also additive over devices

Performance Metric

- We assume that the "instantaneous cost" at time t is

$$\text{electricity bill paid at time } t + \text{dis-utility at time } t$$

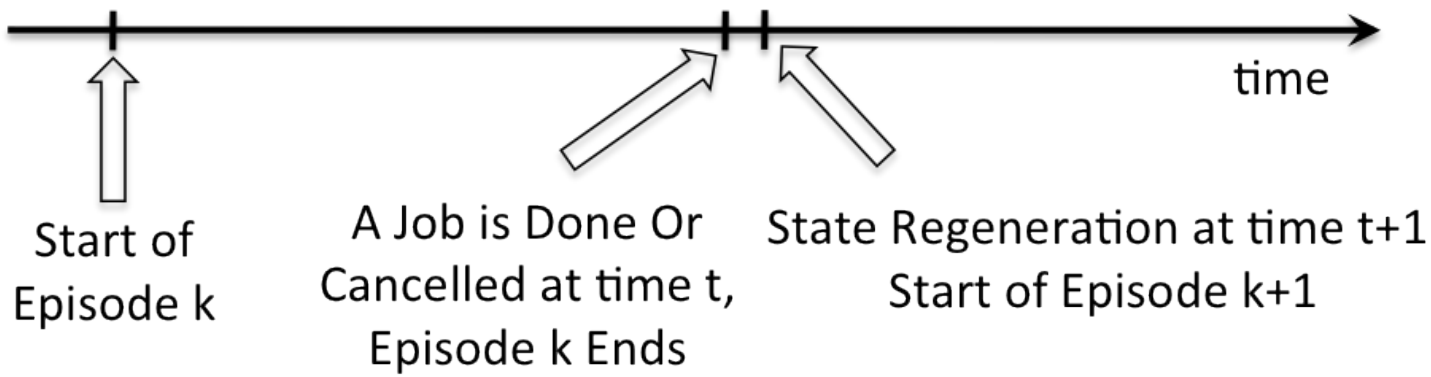
- RL-EMS aims to minimize the expected infinite-horizon discounted cost
- The instantaneous cost is additive over devices

Device Based MDP Model

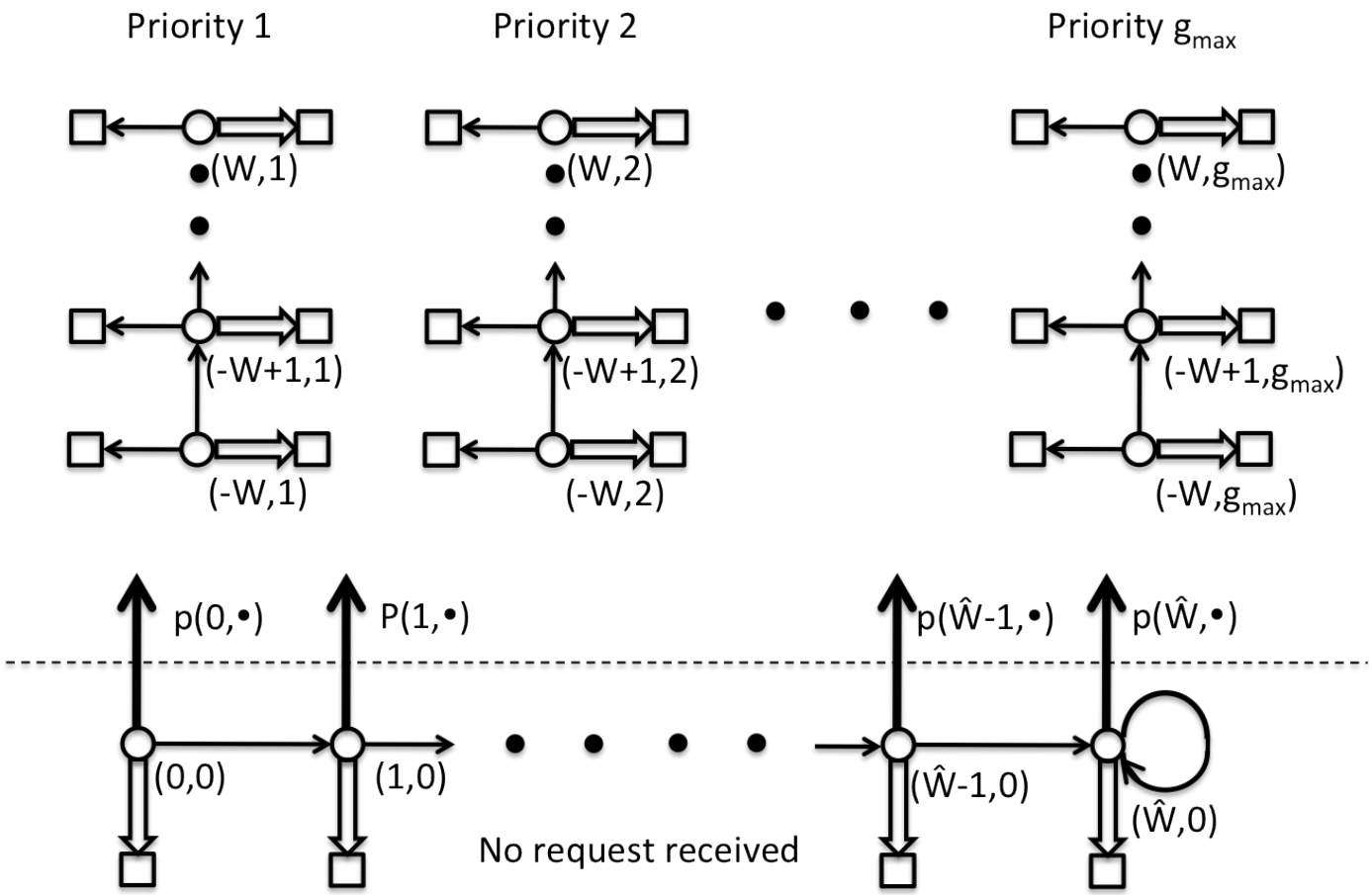
- Assumption 2: Both the electricity price and the consumer requests to the RL-EMS follow exogenous Markov chains. Furthermore, we assume that
  - Electricity price process is independent of the consumer requests process
  - Consumer requests to different devices are independent
- Under Assumption 1 and 2, RL-EMS aims to solve an infinite-horizon discounted MDP, and this MDP decomposes over devices
- Hence, we can derive an optimal scheduling policy for a single device by solving the associated device-based MDP

Probability Transition Model

- Electricity price follows an exogenous Markov chain
- The timeline for a device can be divided into "episodes"
  - Whenever a device completes a job, or an unsatisfied request is cancelled, the current episode terminates
  - At the next time step, the device "regenerates" its state according to a fixed distribution



Probability Transition Model (Cont...)



- $|\mathcal{S}| = P_{max}\left[(2W+1)g_{max} + \hat{W} + 1\right]$

DP Solution and Motivation for RL

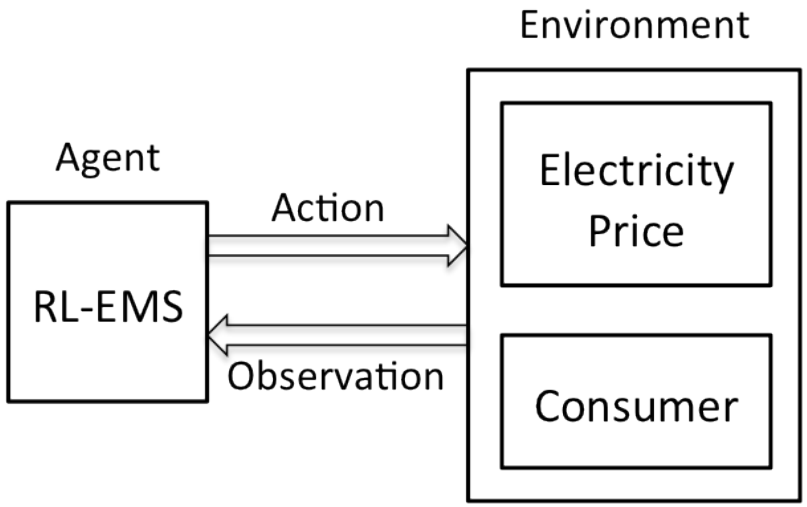
- If RL-EMS knows
  1. The transition model of the device based MDP
  2. The dis-utility function of the consumer

  then the optimal scheduling strategy can be derived based on finite-horizon dynamic programming (DP)
- In fact, it is an optimal stopping problem in each episode
- However, in practice, RL-EMS needs to learn the transition model and dis-utility function while interacting with the consumer and the electricity price
  - RL is the approach in this case

Outline

RL-EMS Algorithm

- In the classical RL literature, the *environment* consists of an unknown MDP, and an *agent* learns how to make decisions while interacting with the environment
- In this case, the "agent" is the RL-EMS and the "environment" includes both the electricity price and the consumer



- In this project, we use $Q(\lambda)$ algorithm, which combines classical Q-learning with (1) Eligibility traces and (2) Importance sampling

$Q(\lambda)$ Algorithm

Initialize Q_0 arbitrarily, set eligibility parameter $\lambda \in [0, 1]$.
Repeat for each episode:
Choose a small constant step-size $\beta > 0$ for each episode.
Initialize eligibility trace vector $e_{t-1} = 0$.
Take $a(t)$ from $x(t)$ according to μ_b (e.g. ϵ-softmin policy), and arrive at $x(t+1)$.
for each time step in an episode **do**
Observe sample, $(x(t), a(t), x(t+1), \Phi_t)$ at time step t, where Φ_t is the instantaneous cost.
$\delta_t \stackrel{\text{def}}{=} \Phi\left(x(t), a(t), x(t+1)\right) + \alpha \min_{a'} Q_t\left(x(t+1), a'\right) - Q_t\left(x(t), a(t)\right)$.
If $a(t) \in \operatorname{argmin}_a Q_t(x(t), a)$, then $\rho_t \leftarrow \frac{1}{\mu_b(a(t)|x(t))}$; otherwise $\rho_t \leftarrow 0$.
$e_t = \psi_t + \rho_t \alpha \lambda e_{t-1}$, where ψ_t is a binary vector whose only nonzero element is $(x(t), a(t))$.
$Q_{t+1} \leftarrow Q_t + \beta \delta_t e_t$.
end for

Outline

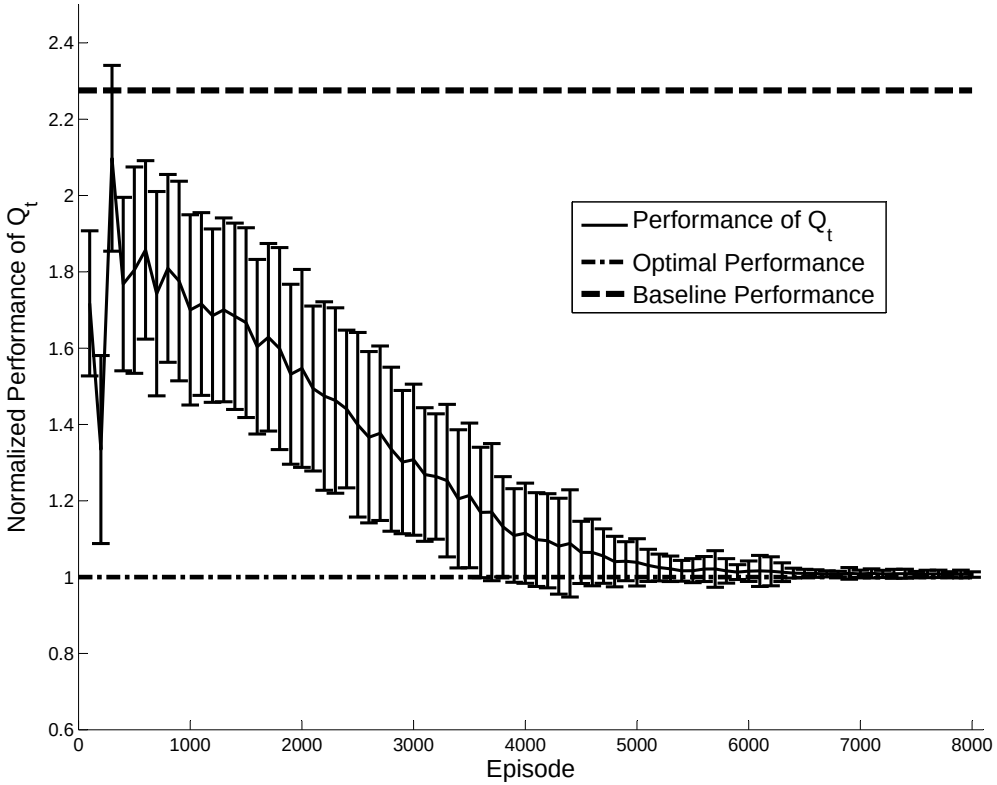
Simulation Result

- A toy example
  - $P_{max} = 4$, $g_{max} = 2$, $W = 4$, $\hat{W} = 5$ and $|\mathcal{S}| = 96$
- Assume that at the beginning of each episode, the "device portion" of the regenerated state is fixed (i.e. $(0, 0)$)
- For each scheduling policy μ, its performance is

$$V(\mu) = \mathbb{E}_{P \sim \pi_P^*} \left[\min_{a \in \mathcal{A}} Q_\mu \left([P, 0, 0]^T, a \right) \right],$$

  and we normalize the performance with respect to the optimal performance.
- We run the proposed RL-EMS algorithm for $8,000$ episodes, and repeat the simulation for 100 times
- Baseline: the default policy without DR

Simulation Result (Cont...)



- Reduce the consumer's cost by 56%

Outline

Conclusion

- We present a RL approach (RL-EMS Algorithm) to DR for residential and small commercial buildings
- RL-EMS does not require the prior knowledge of models on electricity price, consumer behavior and consumer dis-utility
- RL-EMS learns to make optimal decisions by
  - rescheduling the requested jobs
  - anticipating the future uses of devices and doing speculative jobs
- Future Work: RL algorithms with better sample efficiency