RL-BLH: Learning-Based Battery Control for Cost Savings and Privacy Preservation for Smart Meters

Jinkyu Koo, Xiaojun Lin, and Saurabh Bagchi
Purdue University, West Lafayette
{kooj, linx, sbagchi}@purdue.edu
Introduction
Smart meters

• Smart meters
  ▪ Report fine-grained profiles of energy usage

• Many benefits to utility companies
  ▪ Management cost down, demand prediction, time-of-use pricing, and so on

• Customers also beneficial

• But also threaten user privacy!

from https://stopsmartmeters.org
Privacy issues (1/2)

The Privacy Problem -- The Smart Meter Keeps Track.

Electricity Usage Encodes Information about Human Behavior

Electricity Usage → Load Profile → Appliance Profile → Human Behavior

Smart Metering → Inference

Real Power (kWhrs)

Time of Day (DST)

Family Sleeping
Family Arises for bathing, meal/coffee preparation
Home Unoccupied
Active Family at Home: meal preparation, watching TV, etc.
Family Sleeping

from https://smartgridawareness.org

low-frequency variation

behavior profiling
Privacy issues (2/2)

from https://smartgridawareness.org

high-frequency variation

Types of appliances in use
Delaying the era of smart grids

“I want my old meter back, paying $5 fee each month for employees to read the meter.”

from CBS 5 News in Phoenix, Arizona

Several lawsuits ongoing to stop installing smart meters

from https://stopsmartmeters.org
Battery-based load hiding (BLH)

- A battery between the smart meter and appliances
- What the smart meter reports
  - How we charge the battery
- Appliances use energy stored in the battery
  - Decouples meter readings from actual usage profile
- Has some limitations

$x_n$: usage profile
$y_n$: meter reading
Typical ways to control the battery

- **Flattening high-frequency components [1,2]**
  - Effective in hiding load signatures
  - Does not change much the shape of usage profile envelop

- **Discrete-state Markov decision process (MDP) [3]**
  - Can hide both low- and high-frequency components
  - Required to know the probability distribution of usage profile
  - Quantization: performance vs. complexity

---

Our contributions

• **Hides both low- and high-frequency variations of usage profile in practical setup**
  - No quantization of energy usage
  - No knowledge about probability distribution of usage

• **Cost savings by exploiting Time-of-Use (TOU) pricing**
  - Charge the battery when price is low and use the stored energy when the price is high
  - Reinforcement learning based optimal decisions on how much to charge

• **Speedup learning**
  - Synthetic data generation in run-time
  - Reuse of data in early phases
Solution approach
System model

**Meter reading**

\[ 0 \leq y_n \leq x_M \]

reported to utilities

\( x_n \): usage profile
\( y_n \): meter reading

\( y_n \): a continuous variable

\( 0 \leq x_n \leq x_M \)

physical limit

a continuous variable

**Battery level**

\[ b_n = b_{n-1} + y_{n-1} - x_{n-1} \]

\[ 0 \leq b_n \leq b_M \]

capacity

\( b_n \): battery level

consumed by appliances

rechargeable battery

same limit

usage profile

rechargeable battery

Meter reading

Usage profile

Battery level
Privacy protection

• Changing $y_n$ in every $n$ was shown to be not good.
  ▪ Causes significant correlation between $x_{n-1}$ and $y_n$


• We shape the meter readings as rectangular pulses.
  ▪ Change the values of $y_n$ only once every $n_D$ measurement intervals
  ▪ Like high-frequency flattening, this reduces correlation between $x_n$ and $y_n$ for $n_D$ intervals

• The pulse magnitude changes for cost savings
  ▪ Hides low-frequency variation as well, since the magnitude is determined mainly based on the current battery level, not the shape of usage profile
Cost savings (1/2)

- **How to achieve cost savings?**
  - Charge a battery when price is low, and use the stored energy when price is high
- **Cost savings of a day, denoted by $S$:**

$$S = \sum_{n=1}^{n_M} r_n x_n - \sum_{n=1}^{n_M} r_n y_n$$

$$= \sum_{n=1}^{n_M} r_n (x_n - y_n)$$

- what you pay w/o RL-BLH
- what you pay w/ RL-BLH

- **rate (price)**

- **maximum cost savings** = \((r_H - r_L)b_M\)

  e.g., \(r_L = 7.04\) cent per kWh and \(r_H = 21.09\) cent per kWh
Cost savings (2/2)

- **Cost savings for the** $k$-th **decision interval**
  \[ S_k(a) = \sum_{n=(k-1)n_D+1}^{kn_D} r_n(x_n - a) \]
  \[ S = \sum_{n=1}^{n_M} r_n(x_n - y_n) \]

- **The maximum cost savings of a day**
  \[ \max E \left( \sum_{k=1}^{k_M} S_k(a) \right) = \max_a Q^*(1, B_1, a) \]

- **Bellman equations**
  \[ Q^*(k, B_k, a) = \int_{-x_Mn_D}^{x_Mn_D} P_k(z) \left( S_k(a) + \max_{a'} Q^*(k + 1, B_k + z, a') \right) dz \]

The maximum cost savings we can achieve with $a$ from $k$ to $k_M$

Probability that the change in the battery level is $z$ from $k$ to $k + 1$

Immediate return with $a$ at $k$

The maximum we can achieve from $k+1$ to $k_M$
Reinforcement learning
Reinforcement learning to maximize cost savings

- $Q^*(k, B_k, a)$ estimated by a running average

\[
Q^*(k, B_k, a) = \int_{-X_M}^{X_M} P_k(z) \left( S_k(a) + \max_{a'} Q^*(k + 1, B_k + z, a') \right) dz
\]

\[
Q^*(k, B_k, a) = E(\cdot) \approx \frac{1}{N} \sum_{i=1}^{N} \text{sample}_i
\]

\[
Q(k, B_k, a) \leftarrow (1 - \alpha)Q(k, B_k, a) + \alpha \left( S_k(a) + \max_{a'} Q(k + 1, B_{k+1}, a') \right)
\]

Q learning

$Q(k, B_k, a)$ can be rewritten as:

\[
Q(k, B_k, a) \leftarrow Q(k, B_k, a) + \alpha \left( S_k(a) + \max_{a'} Q(k + 1, B_{k+1}, a') - Q(k, B_k, a) \right)
\]

$\Delta Q(k, B_k, a)$

$Q(k, B_k, a)$ converges when $\Delta Q(k, B_k, a)$ goes to zero
Q approximation

- The number of possibilities for state \((k, B_k, a)\) is infinite

  \[Q(k, B_k, a) \leftarrow Q(k, B_k, a) + \alpha \left( S_k(a) + \max_{a'} Q(k + 1, B_{k+1}, a') - Q(k, B_k, a) \right)\]

  Explicitly representing \(Q(k, B_k, a)\) for all possible states is infeasible.

- Approximate \(Q(k, B_k, a)\) by a linear combination of representative features

  \[Q(k, B_k, a) = \sum_{i=0}^{5} w_i^{(a)} f_i(k, B_k)\]

<table>
<thead>
<tr>
<th>(i)</th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>(f_i(k, B_k))</td>
<td>1</td>
<td>(\bar{k})</td>
<td>(\bar{b})</td>
<td>(\bar{k}\bar{b})</td>
<td>(\bar{k}^2)</td>
<td>(\bar{b}^2)</td>
</tr>
</tbody>
</table>

\(\bar{k} = k/k_M\) \hspace{1cm} \(\bar{b} = B_k/b_M\)
• **We minimize** $E(\Delta Q(k, B_k, a)^2)$

\[
Q(k, B_k, a) \leftarrow Q(k, B_k, a) + \alpha \left( S_k(a) + \max_{a'} Q(k + 1, B_{k+1}, a') - Q(k, B_k, a) \right)
\]

• **With the stochastic gradient descent, the weights can be learned by:**

\[
w_i^{(a)} \leftarrow w_i^{(a)} + \alpha \Delta Q(k, B_k, a) f_i(k, B_k)
\]

learning rate
Means to expedite learning

- **Generating synthetic data on the fly**
  - Convergence to the optimal decision policy takes time, which is proportional to the time to collect enough number of training samples
  - Can reduce the time to convergence by feeding artificially generated data
  - Every $d_G$ days, we generate $t_G$ days of artificial usage profiles
    - $x_n$ is sampled according to its statistical characteristic that is coarsely learned

- **Reuse of data**
  - Initial values of weights $w_i^{(a)}$ are random
  - In early phase, data is not fully utilized
  - Until the first $d_R$ days, we store the usage profile of each day, and re-train the system $t_R$ times using the profiles
Experiments
Evaluation metrics

- **Mutual information (MI)**
  The smaller the better
  \[
  MI = \frac{1}{n_M - 1} \sum_{n=1}^{n_M-1} \frac{H(X_n) - H(X_n|Y_n)}{H(X_n)}
  \]
  
  uncertainty reduction by observing \(Y_n\)

- **Pearson correlation coefficient (CC)**
  The smaller the better
  \[
  CC = \frac{\sum_{n=1}^{n_M} (x_n - \bar{x}) \sum_{n=1}^{n_M} (y_n - \bar{y})}{\sqrt{\sum_{n=1}^{n_M} (x_n - \bar{x})^2 \sum_{n=1}^{n_M} (y_n - \bar{y})^2}}
  \]

  High-frequency variation: Load signatures
  Low-frequency shape: Behavioral patterns

- **Saving ratio (SR)**
  The higher the better
  \[
  SR = E \left( \frac{\sum_{n=1}^{n_M} r_n (x_n - y_n)}{\sum_{n=1}^{n_M} r_n x_n} \right)
  \]

  Cost savings

\[X_n = (x_n, x_{n+1})\]
\[Y_n = (y_n, y_{n+1})\]
Comparison with a prior scheme (1/2)

(a) RL-BLH \( (n_D = 10) \)

(b) Low-pass (high-frequency flattening)

Comparison with a prior scheme (2/2)

(a) Correlation coefficient

(b) Mutual information

(c) Saving ratio
Effects of heuristics for speedup

\[ \sim |\Delta Q(k, B_k, a)| \]
Concluding remarks

- **RL-BLH hides both low- and high-frequency signals in energy usage**
  - Protection to high-frequency information comparable to the low-pass filtering
  - Protection to low-frequency information superior to the low-pass filtering

- **Cost savings by exploiting Time-of-Use (TOU) pricing**
  - ~15% cost savings with 5kWh battery in a typical home
    ✓ Cost saving is proportional to the battery capacity
  - Provides an economical benefit in addition to privacy protection
  - Caters to cost-conscious as well as privacy-conscious users

- **Speedup learning**
  - Significantly reduces the learning time
  - Makes the solution practical