Reinforcement Learning for Power Management in Energy Harvesting Sensor Nodes*

Shaswot SHRESTHAMALI, Masaaki KONDO, Hiroshi NAKAMURA

Graduate School of Information Science and Technology, The University of Tokyo

東京大学
THE UNIVERSITY OF TOKYO

INTRODUCTION

In the near future, Internet of Things (IoT) will consist of billions and trillions of nodes. Energy Harvesting Wireless Sensor Nodes (EHWSN) play a critical role in forming a **sustainable, maintenance-free network of perpetually communicating autonomous devices** for the IoT infrastructure. Energy autonomy (neutrality) of the sensor nodes needs to be ensured for perpetual operation. Here we consider a case of a solar energy harvesting sensor node.

CHALLENGES

1. Energy Neutral Operation

The node needs to be energy neutral. The **amount of energy harvested should equal the amount of energy consumed** by the node. This way we can ensure that battery will never go empty and that none of the harvested energy goes to waste. This is not a trivial issue because the energy harvested is often unreliable and unpredictable.

2. Adaptivity

The node has to be able to adapt to changes in energy harvesting profile due to

- **Diurnal Variations**
- **Seasonal Variations**
- **Climatic Variations**
- **Changes in device power consumption**
- **Battery Degradation**

3. Diversity and Scaling

When we deal trillions of different sensor nodes, customizing the power management policy for each node according to its power rating and environment will be incredibly impractical. Moreover, the policy may have to revised if their working environment changes. Instead, we need a **one-size-fits-all solution** that can work for all types of EHWSN scenarios.

REINFORCEMENT LEARNING AS A SOLUTION

Since heuristic policies will not suffice, the most practical solution is to use a **context aware perception-action cyclic approach.** Our solution is to use Reinforcement Learning (RL). The power manager learns the power management policy by interacting with the environment and memorizing the best strategy after numerous hit-and-trial interactions.

- **States**: Defined by battery level, harvested energy, weather forecast and energy neutral performance (ENP)
- **Actions**: Choose a duty cycle
- **Reward**: Received at end of day (episode) depending on net energy difference
- **Learning**: **SARSA - λ** with eligibility traces

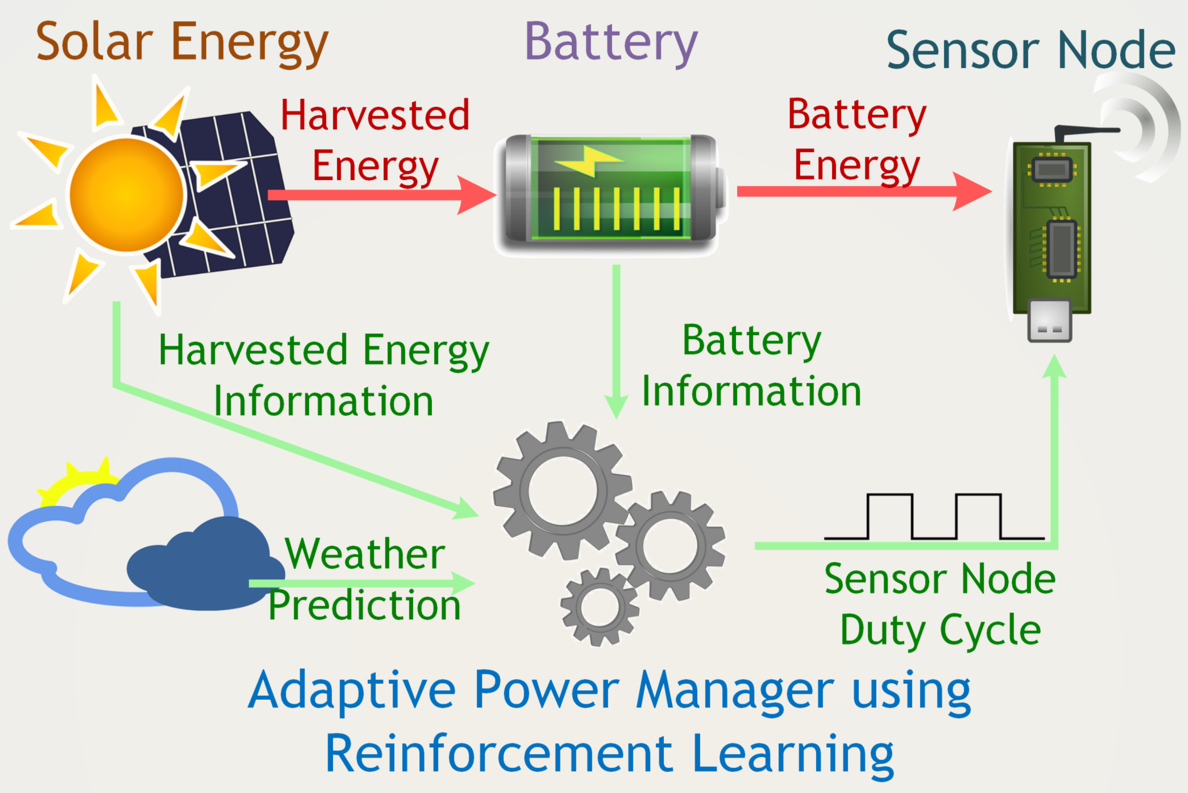
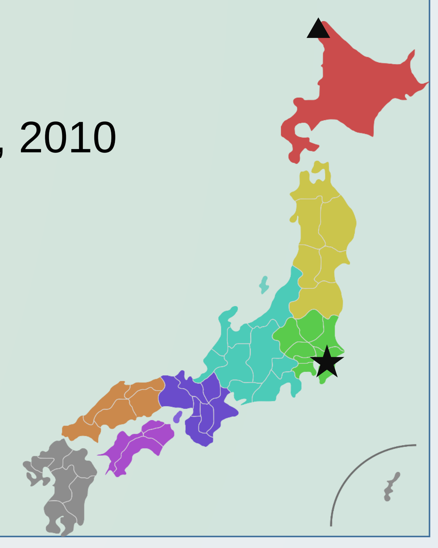


Figure 1: System Model

We use a simplified system model that consist of a solar panel, an ideal battery, a sensor node and a RL power management unit that specifies the duty cycle of the sensor node.

SIMULATION PARAMETERS

- **Battery**: 40,000 mWh (ideal)
- **Node power consumption**: 100 mWh to 500 mWh (5 discrete duty cycles)
- **Solar power**: 0 to 3000 mWh (Energy profiles obtained from Japan Meteorological Agency)
- **Weather Forecast**: 6 different types (Very little sun, Overcast, Partly Cloudy, Fair, Sunny, Very Sunny)
- **Training**: Solar data for Tokyo, 2010
- **Testing**: Solar data for :
  - ▲ Wakkanai, 2011
  - ★ Tokyo, 2011



EXPERIMENTAL RESULTS

We compare the performance of our proposed SARSA - λ algorithm with an optimal policy. The Optimal Policy uses non-causal information and linear optimization techniques to arrive at an optimum solution. We compare the performance of the two policies in different seasons and locations. We also show how including weather forecast information contributes to enhancing the performance of our method.

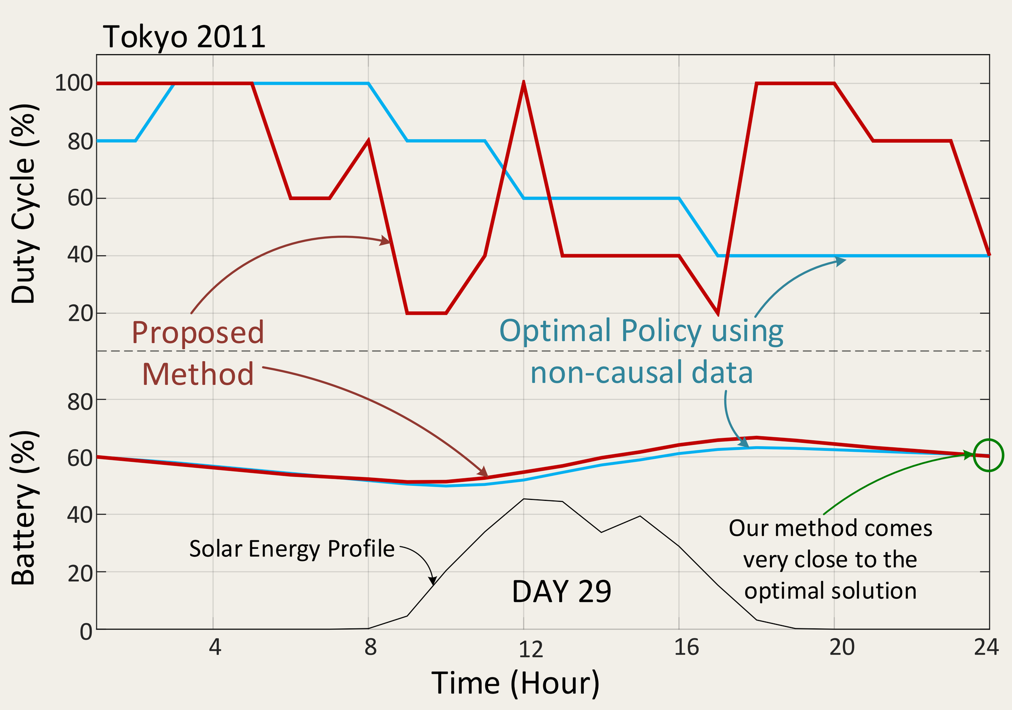


Figure 2: SARSA Method vs. Optimal Policy

In the above figure, we observe that both policies have similar battery profiles. While the two policies may differ in how they allocate the energy, at the end of the day, both of them come very close to perfect energy neutrality.

The actual net deviation from the initial battery level (60%) was only 491.875mWh and 191.875 mWh for SARSA and Optimal Policy respectively.

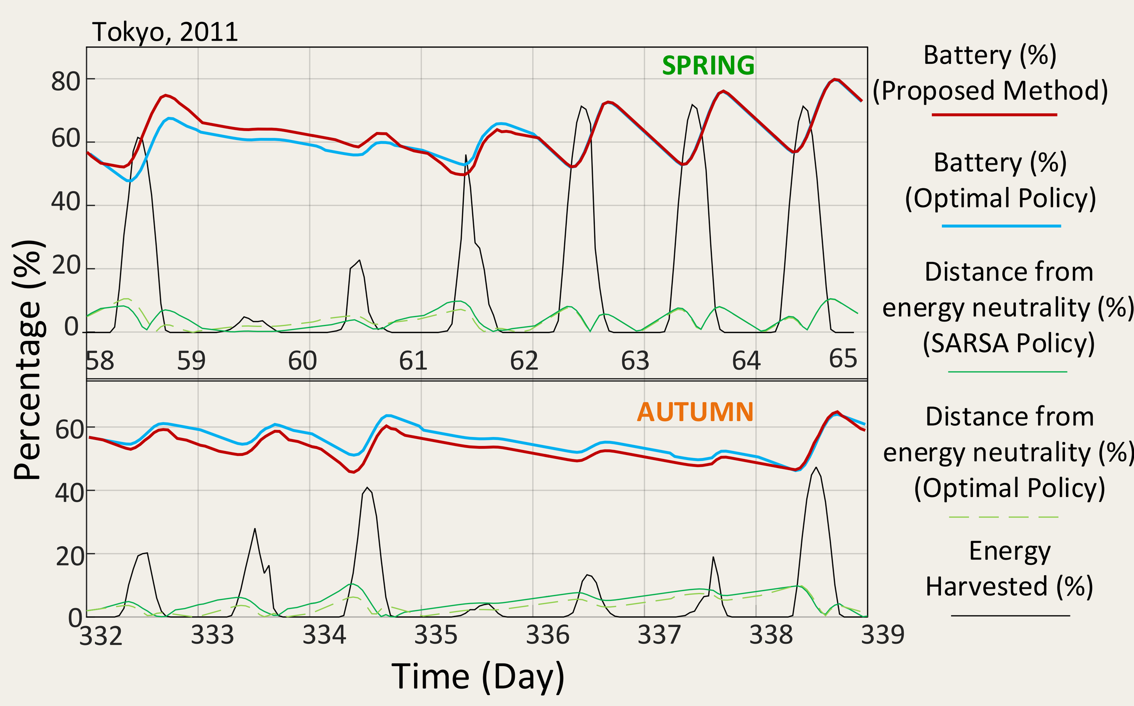


Figure 3: Adaptation to Seasonal Changes

Our method is able to adapt to both spring and autumn solar energy profiles. The resultant battery profile differs very little from the optimum policy.

This is largely due to our novel RL problem formulation (state definition and reward function). Because of our unique RL problem formulation, the power manager takes energy neutral performance into account when making decisions rather than only the battery reserve level.

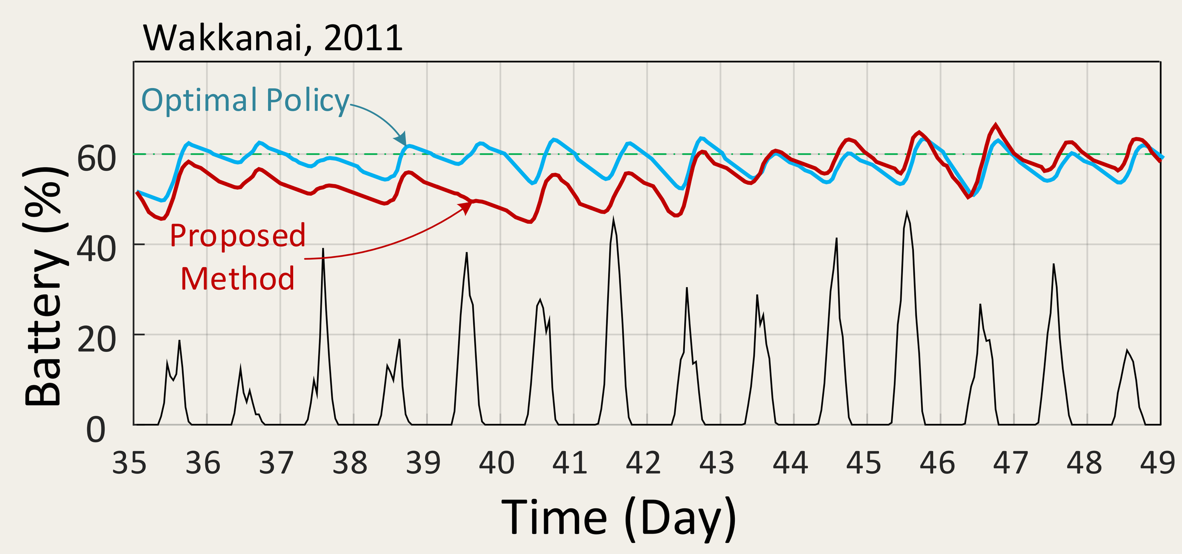


Figure 4: Adaptation to a new location

Our proposed method adapts well in Wakkanai, a location that has a drastically different climate from the one it was trained in.

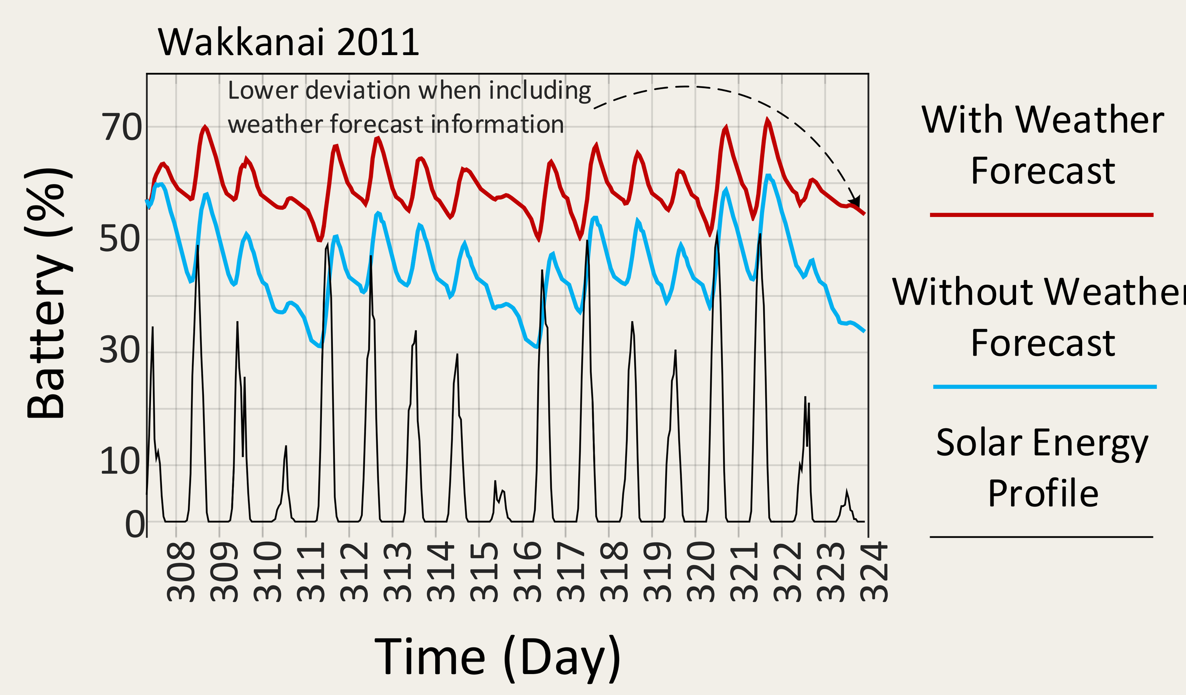


Figure 5: Inclusion of weather forecast enhances performance

Furthermore, we show that inclusion of weather forecast information results in better results.

*This work was partially supported by JSPS KAKENHI Grant Number 16K12405.