BEAR: Reinforcement Learning for Throughput Aware Borrowing in Energy Harvesting Systems

Anubhav Sachan[1], Deepak Mishra[2], and Ganesh Prasad[3]

[1]Samsung Research and Development Institute Delhi, India.

[2]University of New South Wales, Australia.

[3]National Institute of Technology Silchar, India.

OUTLINE

Introduction

INTRODUCTION

- ***Energy Harvesting (EH) systems provide greener alternative*** to self-sustaining devices functioning in a complex environment [1]
- ***Energy Borrowing strengthens these independent EH systems*** with the incorporation of a secondary (more reliable) power source [2]
- In a dynamic environment, ***EH arrival and channel conditions are random, statistically indeterminable***, hence data-driven approaches are essential
- The ***integration of concepts*** of utilizing harvested energy and temporary borrowing will play significant role in sustainability of low power wireless devices

[1] K. Tutuncuoglu and A. Yener, "*Energy harvesting networks with energy cooperation: Procrastinating policies*," IEEE Transactions on Communications, 2015.
[2] Z. Sun, L. Dan, Y. Xiao, P. Yang, and S. Li, "*Energy borrowing for energy harvesting wireless communications,*" IEEE Communications Letters, 2016.

RELATED WORKS

Reinforcement learning methods are efficiently used in a point-to-point energy harvesting communication system to

1. Learn the transmission power allocation policy to maximize the received data using SARSA algorithm [1]
2. Formulate a transmission policy to maximize the expected transmitted data with Q-Learning methods [2]
3. The actor-critic algorithm [3] is widely used in user scheduling and resource allocation and energy management in wireless EH nodes [4]

[1] A. Masadeh, Z. Wang, and A. E. Kamal, "*Reinforcement learning exploration algorithms for energy harvesting communications systems*," in Proc. IEEE International Communications Conference (ICC), 2018.
[2] P. Blasco, D. Gunduz, and M. Dohler, "*A learning theoretic approach to energy harvesting communication system optimization*," IEEE Transactions on Wireless Communications, 2013.
[3] R. S. Sutton and A. G. Barto, *Reinforcement Learning: An Introduction*. Cambridge, MA, USA: A Bradford Book, 2018.
[4] A. Masadeh, Z. Wang, and A. E. Kamal, "*An actor-critic reinforcement learning approach for energy harvesting communications systems*," in 2019 28th International Conference on Computer Communication and Networks (ICCCN), July 2019.

MOTIVATION

The discussed systems transmit data if and only if the energy source attached with the transmitter possesses a required amount of energy for transmission, *irrespective* of the channel conditions.

In a typical scenario,

Case I:

Channel Conditions ✓
Available Energy ✓

Transmission ✓

Case II:

Channel Conditions ✓
Available Energy ✗

Transmission ✗ → (with borrowing) ✓

CONTRIBUTIONS

Three-fold contribution

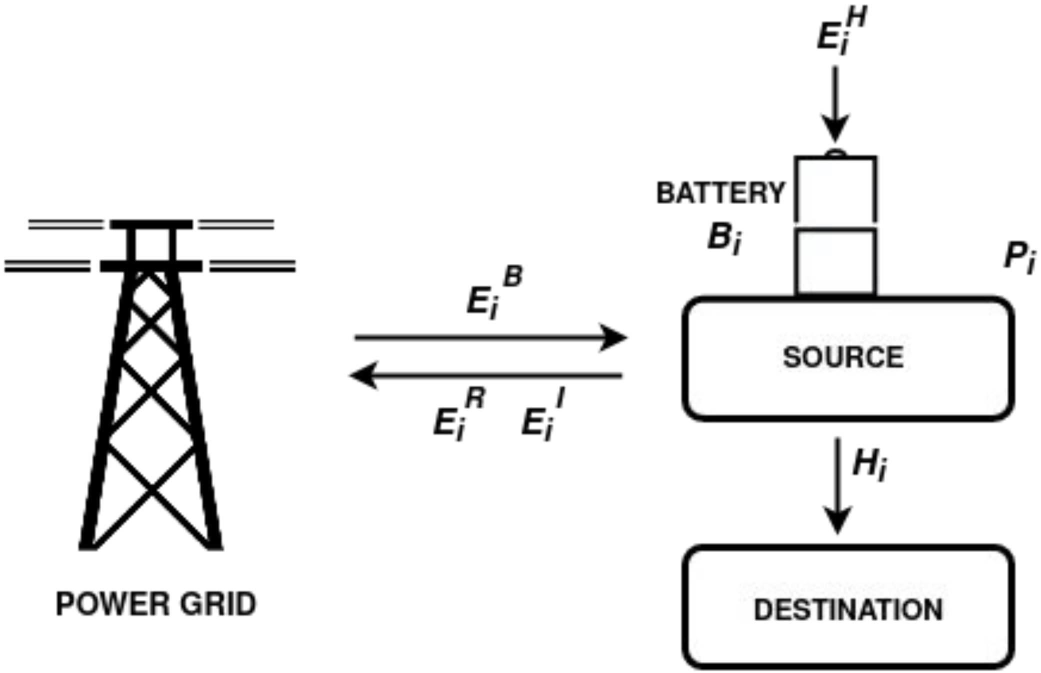
1. Introduction of **BEAR** (Borrowing Energy with Adaptive Rewards)
   a) to ***maximize the throughput over finite time slots***
   b) to establish an ***reinforcement learning benchmark*** for a borrowing-aided EH wireless communication system

2. ***Adaptive Rewards*** and ***Penalty functions*** for an efficient learning

3. Incorporation of ***borrowing with adaptive rewards*** led us to obtain 35.45% gain over an EH system without borrowing

System Model

SYSTEM MODEL



ENERGY SCHEDULING PROTOCOL[1]

- Battery Energy Level

$$B_i = \min\{(B_{i-1} + E_i^H), B_{\max}\} + E_i^B - E_i^R - t_s P_i$$

- Energy Borrowing Schedule

$$E_i^B = \begin{cases} t_s P_i - (B_{i-1} + E_i^H), & \text{if } t_s P_i > (B_{i-1} + E_i^H) \\ 0, & \text{otherwise} \end{cases}$$

[1] G. K. Reddy, D. Mishra, and L. N. Devi, "*Scheduling protocol for throughput maximization in borrowing-aided energy harvesting system,*" IEEE Networking Letters, pp. 1–1, 2020.

ENERGY SCHEDULING PROTOCOL

- Energy Return Schedule

$$E_i^R = \begin{cases} \zeta E_i^E, & \text{if } E_{i-1}^U > E_i^E \\ \zeta E_{i-1}^U, & \text{otherwise} \end{cases}$$

Unreturned Energy

$$E_i^U = \begin{cases} E_{i-1}^U + E_i^B, & \text{if } t_s P_i > (B_{i-1} + E_i^H) \\ E_{i-1}^U + E_i^I - E_i^R, & \text{if } t_s P_i \leq (B_{i-1} + E_i^H) \\ & \text{and } E_i^R \leq (E_{i-1}^U + E_i^I) \\ 0, & \text{otherwise} \end{cases}$$

Problem Definition

PROBLEM DEFINITION

- Determine the optimal values of P_i across finite time slots for sum throughput to be maximized

- During transmission,
    - if the requisite amount of energy is not available,
      the ***source can borrow energy*** from nearby power grid
      to best utilize channel condition.

- Complete pool of ***"borrowed" energy has to be returned*** along with the ***levied interest***.

PROBLEM DEFINITION

- State-Action Space[1-3]
  - State

$$S_i = (E_i^H, H_i, B_i)$$

  - Action

$$A_i = P_i$$

- Spectral Efficiency [4]

$$C_i(S_i, P_i) = \log_2\left(1 + \frac{P_i|H_i|^2}{N_P}\right)$$

[1] P. Henderson, R. Islam, P. Bachman, J. Pineau, D. Precup, and D. Meger, “Deep reinforcement learning that matters,” Proceedings of the AAAI Conference on Artificial Intelligence, Apr. 2018.
[2] E. Brunskill and R. Sarkar, “Lecture 2: Making good decisions given a model of the world,” in CS234: Reinforcement Learning – Stanford University, 2019.
[3] R. S. Sutton and A. G. Barto, *Reinforcement Learning: An Introduction*. Cambridge, MA, USA: A Bradford Book, 2018.
[4] A. Masadeh, Z. Wang, and A. E. Kamal, “*An actor-critic reinforcement learning approach for energy harvesting communications systems*,” in 2019 28th International Conference on Computer Communication and Networks (ICCCN), July 2019.

OBJECTIVE: SUM THROUGHPUT MAXIMIZATION

- Reward Function

$$r_i(S_i, P_i) = \begin{cases} C_i(S_i, P_i), & \text{if } i = 0 \\ C_i(S_i, P_i) - \beta E^U_{i-1}, & \text{if } i \in (0, N] \end{cases}$$

where adaptive penalty β → Sensitivity Parameter

$$\beta = \varepsilon^{\lfloor i/l \rfloor}.$$

- Objective $\max_{P_i} \sum_{i=0}^{N} \log_2 \left(1 + \frac{P_i |H_i|^2}{N_P}\right)$

Proposed Solution

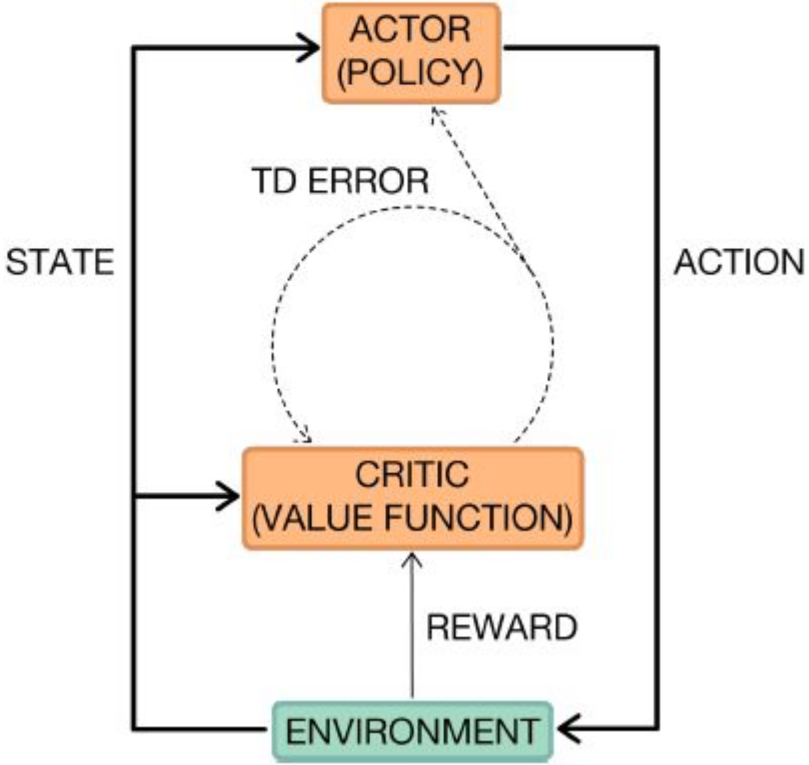
ACTOR CRITIC ALGORITHM

- Actor *(Policy)*

  “decides which action to take”

- Critic *(Action - Value function)*

  “Tells how good the action was (criticizes) and asks to improve them”



Actor Critic Algorithm

ROLE OF ACTOR

- Policy is modeled by Gaussian Distribution[1]

$$\boldsymbol{\pi}(P_i|S_i,\boldsymbol{\theta}) = \frac{1}{\sqrt{2\pi(\boldsymbol{\sigma}(S_i,\boldsymbol{\theta}_\sigma))^2}}\, e^{-\frac{(t_s P_i - \boldsymbol{\mu}(S_i,\boldsymbol{\theta}_\mu))^2}{2(\boldsymbol{\sigma}(S_i,\boldsymbol{\theta}_\sigma))^2}}$$

with parameterized mean and standard deviation as

$$\boldsymbol{\mu}(S_i,\boldsymbol{\theta}_\mu) = \max\left\{0, E^A\left(\frac{1+\tanh(\boldsymbol{\theta}_\mu^\top \boldsymbol{\phi}(S_i))}{2}\right)\right\}, \quad \sigma(S_i,\boldsymbol{\theta}_\sigma) = \exp\left(\boldsymbol{\theta}_\sigma^\top \boldsymbol{\phi}(S_i)\right)$$

[1] A. Masadeh, Z. Wang, and A. E. Kamal, "*An actor-critic reinforcement learning approach for energy harvesting communications systems*," in 2019 28th International Conference on Computer Communication and Networks (ICCCN), July 2019.

ROLE OF CRITIC

- Responsible for action-value function, and "*criticize*" the policy chosen by the actor

- Updates itself using temporal difference error δ

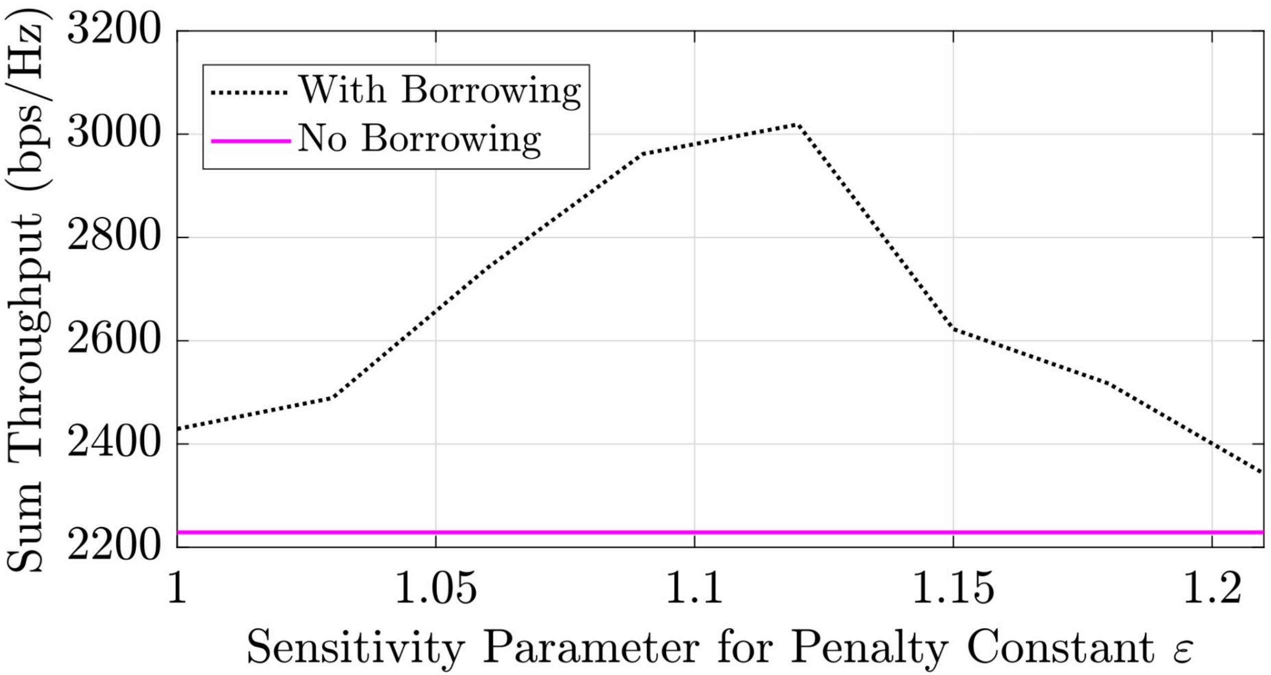
$$\delta = \{r_i + \gamma \hat{Q}(S_{i+1}, P_{i+1}, \boldsymbol{w}_t)\} - Q(S_i, P_i, \boldsymbol{w})$$

- To update critic network, we used backpropagation and stochastic gradient descent [1]

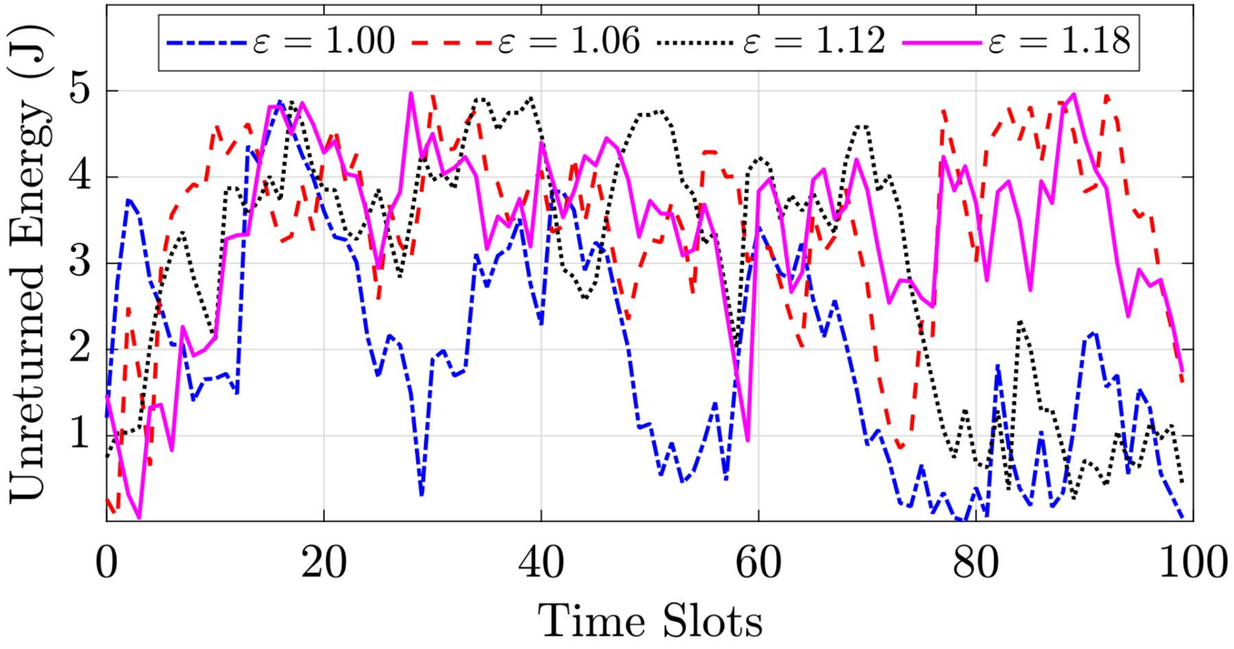
[1] L. Bottou, "*Online learning and stochastic approximations,*" Online learning in Neural Networks, vol. 17, no. 9, p. 142

Performance Evaluation

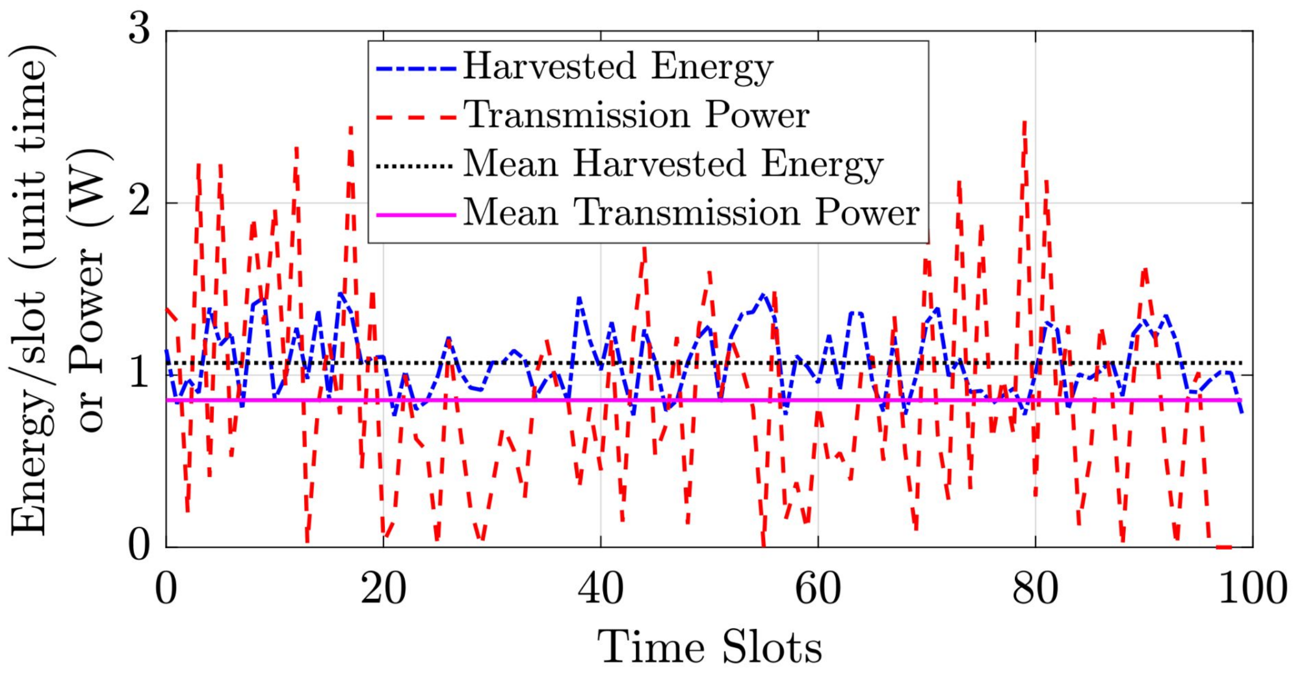
SUM THROUGHPUT



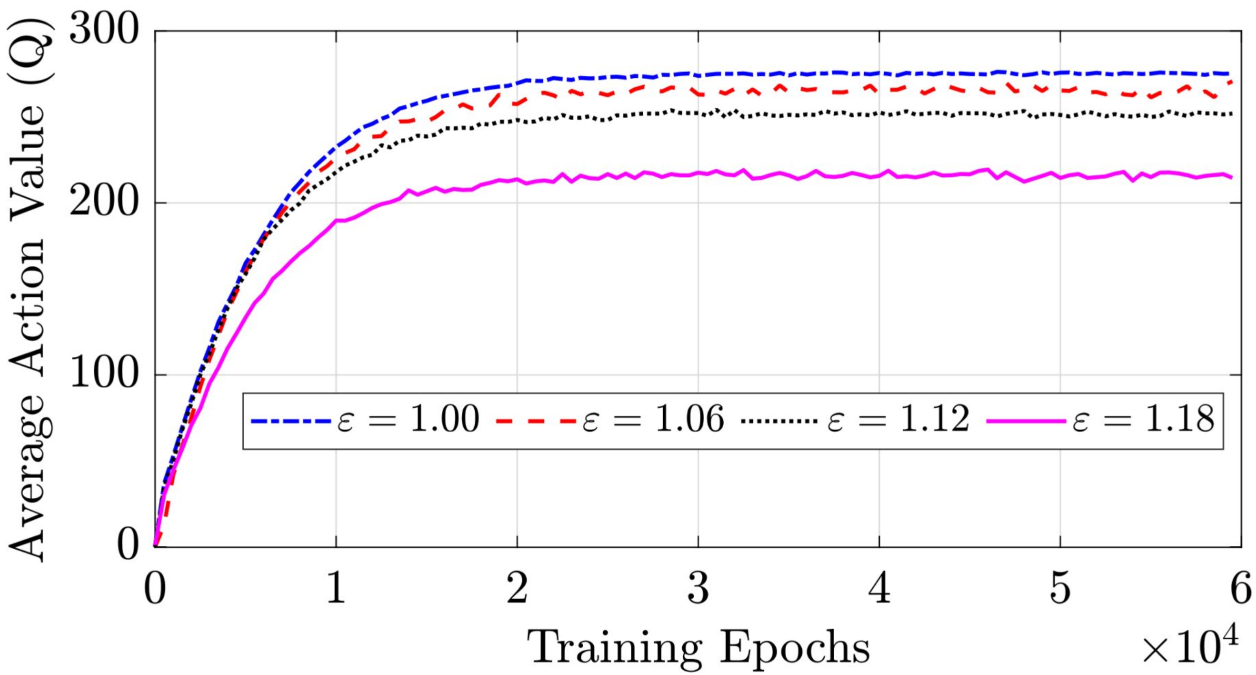
UNRETURNED ENERGY



ENERGY AT EACH TIME SLOT



AVERAGE Q-VALUE



CONCLUSION

- An *enhanced EH wireless communication system* with borrowing, ***formulates a power allocation policy to optimize the harvested energy and sum throughput jointly.***
- Equipped with the concept of ***adaptive rewards.***
- The ***sensitivity parameter*** ε shows that there exists an optimal value of ε for throughput maximization.
- The significant ***gain of 35.45%*** in sum throughput over a non-borrowing energy harvesting system.

Questions?

Reach out: hi@anubhavsachan.com
twitter.com/anubhav4sachan