Structural Properties of Optimal Transmission Policies for Delay-Sensitive Energy Harvesting Wireless Sensors

Nikhilesh Sharma*, Nicholas Mastronarde*, and Jacob Chakareski†

*Dept. Electrical Engineering, U. Buffalo, †Dept. Electrical and Computer Engineering, U. Alabama

***Abstract*—We consider an energy harvesting sensor transmitting latency-sensitive data over a fading channel. We aim to find the optimal transmission scheduling policy that minimizes the packet queuing delay given the available harvested energy. We formulate the problem as a Markov decision process (MDP) over a state-space spanned by the transmitter's buffer, battery, and channel states, and analyze the structural properties of the resulting optimal value function, which quantifies the long-run performance of the optimal scheduling policy. We show that the optimal value function (i) is non-decreasing and has increasing differences in the queue backlog; (ii) is non-increasing and has increasing differences in the battery state; and (iii) is submodular in the buffer and battery states. Our numerical results confirm these properties and demonstrate that the optimal scheduling policy outperforms a so-called greedy policy in terms of sensor outages, buffer overflows, energy efficiency, and queuing delay.**

I. Introduction

Energy-constrained wireless sensors are increasingly used for latency-sensitive applications such as real-time remote visual sensing [1], Internet of Things (IoT), body sensor networks [2], smart grid monitoring, and cyber-physical systems. However, these sensors are subject to time-varying channel conditions and generate stochastic traffic loads, which makes it very challenging for them to provide the necessary Quality of Service (QoS) to support latency-sensitive applications. This is further complicated by the introduction of wireless sensors powered by energy harvested from the environment (e.g., ambient light or RF energy [3]). Although energy harvesting sensors (EHSs) can operate autonomously in (possibly remote) areas without access to power lines and without the need to change their batteries, the stochastic nature of harvested energy sources poses new challenges in sensor power management, transmission power allocation, and transmission scheduling.

An important body of work focuses on offline computation of optimal transmission policies for EHSs [4–7]. In particular, [4] considers a multi-access channel with two EHSs and derives the optimal offline transmission power and rate allocations that maximize the sum rate, given a priori known energy and traffic arrival processes. [7] identifies Markov decision processes (MDPs [8]) as a useful tool for optimizing EHSs in unpredictable environments with only causal information about the past and present. [6] formulates both throughput-optimal and delay-optimal energy management policies as MDPs. Though these studies identify numerous techniques for calculating optimal policies, they do not provide general insights into their structures.

The work of N. Mastronarde and J. Chakareski was supported in part by the NSF under awards ECCS-1711335 and ECCS-1711592, respectively.

Another body of work focuses on characterizing the structure of optimal transmission policies for EHSs [2, 9–13]. Numerous studies have shown that optimal power allocation policies for EHSs have various water-filling structures [9–11]. Other types of structural results are derived in [12, 13]. In particular, [12] assumes that known amounts of data and energy arrive over a finite time horizon, and aims to minimize the total amount of time to transmit all data. They show that the optimal policy uses transmission rates that increase over time. [13] formulates outage-optimal power control policies for EHSs, showing that the optimal policy for the underlying MDP is threshold in the battery state for the special case of binary transmission power levels.

We study an EHS transmitting delay-sensitive data over a fading channel. We assume that it uses a fixed transmission power, can transmit at most one packet in each time slot, and experiences a variable packet loss rate depending on the channel conditions. Under these assumptions, we aim to understand the structure of optimal transmission scheduling policies that minimize the packet queuing delay given the available harvested energy. Our contributions are as follows:

- We formulate the delay-sensitive energy harvesting scheduling (DSEHS) problem as an MDP that takes into account the stochastic traffic load, harvested energy, and channel conditions experienced by the EHS.
- We show that the optimal value function, which quantifies the long-run performance of the optimal scheduling policy, (i) is non-decreasing and has increasing differences in the queue backlog; (ii) is non-increasing and has increasing differences in the battery state; and (iii) is submodular in the buffer and battery states.
- Our numerical results confirm these properties and demonstrate that the optimal scheduling policy outperforms a so-called greedy policy in terms of sensor outage, buffer overflow, energy efficiency, and queuing delay.

Our advances can facilitate *online learning* of optimal policies at lower complexity and enable efficient self-organizing operation of next generation IoT sensing systems (see, e.g., [14]). Such studies fall outside the scope of our paper.

The remainder of this paper is organized as follows. We introduce the system model in Section II, formulate the DSEHS problem in Section III, analyze the structural properties of the DSEHS problem in Section IV, present our numerical results in Section V, and conclude in Section VI.

II. Wireless Sensor Model

We consider a time-slotted single-input single-output (SISO) point-to-point wireless communication system in which an

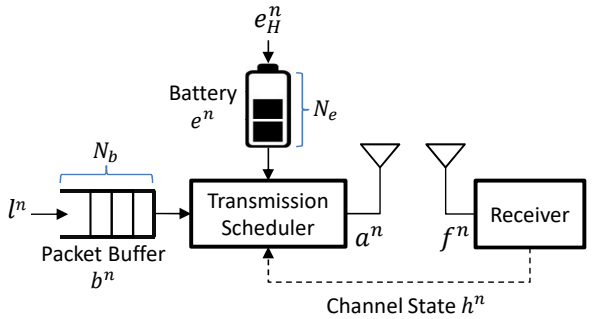


Fig. 1: System block diagram.

energy harvesting sensor transmits latency-sensitive imagery data over a fading channel. The system model is depicted in Fig. 1. The system comprises two buffers: a packet buffer with (possibly infinite) size N_b and an energy buffer (battery) with finite size N_e. We assume that time is divided into slots with length ΔT (seconds) and that the system's state in the nth time slot is denoted by $s^n \triangleq (b^n, e^n, h^n) \in \mathcal{S}$, where $b^n \in \mathcal{S}_b = \{0, 1, ..., N_b\}$ is the packet buffer state (i.e., the number of backlogged data packets), $e^n \in \mathcal{S}_e = \{0, 1, ..., N_e\}$ is the battery state (i.e., the number of energy packets in the battery), and $h^n \in \mathcal{S}_h$ is the channel fading state. At the start of the nth time slot, the transmission scheduler observes the state of the system and takes the binary scheduling action $a^n \in \mathcal{A} = \{0, 1\}$, where $a^n = 1$ indicates that it transmits the head-of-line packet in the queue and $a^n = 0$ otherwise.

A. Channel model

We assume a block-fading channel, meaning that the channel is constant during each time slot and may change from one slot to the next. Similar to prior work [6, 9, 15–17], we assume that the channel state $h^n \in \mathcal{S}_h$ is known to the transmitter at the start of each time slot, that $\mathcal{S}_h$ denotes a finite set of N_h channel states, and that the evolution of the channel state can be modeled as a finite state Markov chain with transition probability function $P^h(h'|h)$.

B. Energy harvesting model

Similar to prior work [13, 15], we assume that battery energy is stored in the form of energy packets. Let $e_H^n \in \mathcal{E} = \{0, 1, \ldots, M_e\}$ denote the number of energy packets that are available for harvesting in the nth time slot and let $P^{e_H}(e_H)$ denote the energy packet arrival distribution. Energy packets that arrive in time slot n can be used in future time slots. Therefore, the battery state at the start of time slot $n+1$ can be found through the following recursion:

$$e^{n+1} = \min(e^n - e_{\text{TX}}(a^n) + e_H^n, N_e), \tag{1}$$

where $e_{\text{TX}}(a^n)$ denotes the number of energy packets consumed in time slot n given the scheduling action a^n. We assume that the wireless sensor uses a fixed transmission power P_{TX} (energy packets per second); therefore,

$$e_{\text{TX}}(a^n) = a^n P_{\text{TX}} \Delta T = a^n e_{\text{TX}} \text{ (energy packets)}. \tag{2}$$

For simplicity, we assume that the transmission energy e_{TX} is an integer multiple of energy packets. Note that the transmission action a^n in time slot n cannot use more energy than is available in the battery, i.e., $a^n e_{\text{TX}} \leq e^n$.

Given the current state $s = (b, e, h)$ and action a, the probability of observing battery state e' in the next slot is:

$$P^e(e'|e, a) = \mathbb{E}_{e_H}[\mathbb{I}_{\{e' = \min(e - a \cdot e_{\text{TX}} + e_H, N_e)\}}], \tag{3}$$

where $\mathbb{I}_{\{\cdot\}}$ is an indicator variable that is set to 1 when $\{\cdot\}$ is true and is set to 0 otherwise.

C. Traffic model

Let $l^n \in \mathcal{L} = \{0, 1, \ldots, M_l\}$ denote the number of data packets generated by the sensor in the nth time slot and let $P^l(l)$ denote the data packet arrival distribution. The buffer state in slot $n+1$ can be found through the following recursion:

$$b^{n+1} = \min(b^n - f^n(a^n, h^n) + l^n, N_b), \tag{4}$$

where $f^n(a^n, h^n)$ is the number of packets transmitted successfully in time slot n and $f^n(a^n, h^n) \leq a^n \leq b^n$. Note that new packet arrivals, and packets that are not successfully received, must be (re)transmitted in a future time slot. Assuming independent and identically distributed (i.i.d.) bit errors, we can characterize f^n as a binary random variable with conditional probability mass function

$$P^f(f|a, h) = \begin{cases} 1, & \text{if } f = 0 \text{ and } a = 0, \\ 0, & \text{if } f = 1 \text{ and } a = 0, \\ q(h), & \text{if } f = 0 \text{ and } a = 1, \\ 1 - q(h), & \text{if } f = 1 \text{ and } a = 1, \end{cases} \tag{5}$$

where $q(h)$ is the packet loss rate (PLR) in channel state h. Since the transmission power is fixed, $q(h^+) < q(h^-)$ if $h^+ > h^-$. We will refer to $P^f(f|a, h)$ as the *goodput distribution*.

Given the current state $s = (b, e, h)$ and action a, the probability of observing buffer state b' in the next time slot is:

$$P^b(b'|[b, h], a) = \mathbb{E}_{f,l}[\mathbb{I}_{\{b' = \min(b - f + l, N_b)\}}]. \tag{6}$$

III. The Delay-Sensitive Energy-Harvesting Scheduling (DSEHS) Problem

Let $\pi : \mathcal{S} \to \mathcal{A}$ denote a *policy* that maps states to actions. The objective of the DSEHS problem is to determine the optimal policy π^* that minimizes the average packet queuing delay given the available energy. However, this does not mean that the policy should greedily transmit packets whenever there is enough energy to do so. Instead, it may be beneficial to abstain from transmitting packets in bad channel states and wait to transmit them in good channel states to reduce retransmissions and conserve scarce harvested energy. On the other hand, the policy should not be too conservative. Instead, if the battery is (nearly) full, transmitting a packet will make room for more harvested energy, which otherwise would be lost due to the finite battery size. To balance these considerations, we formulate the DSEHS problem as an MDP [8].

We define a *buffer cost* to penalize large queue backlogs. Formally, we define the buffer cost as the sum of the *holding*

cost and the expected *overflow cost* with respect to the arrival and goodput distributions, i.e.,

$$c([b,h],a) = b + \mathbb{E}_{f,l}[\{\eta \max(b - f + l - N_b, 0)\}], \quad (7)$$

In (7), the holding cost is equal to the buffer backlog, which is proportional to the queuing delay by Little's theorem [18]. The overflow cost imposes a penalty η for each dropped packet.

Formally, the DSEHS problem's objective is to determine the scheduling policy that solves the following optimization:

$$\underset{\pi \in \Pi}{\text{minimize}} \quad \mathbb{E}\left[\sum\nolimits_{n=0}^{\infty} (\gamma)^n c(s^n, \pi(s^n))\right], \quad (8)$$

where $\gamma \in [0,1)$ is the discount factor, Π is the set of all possible policies, and the expectation is taken over the sequence of states, which are governed by a controlled Markov chain with transition probabilities:

$$P(s'|s,a) = P^b(b'|[b,h],a)P^h(h'|h)P^e(e'|e,a). \quad (9)$$

The optimal solution to (8) satisfies the following Bellman equation, $\forall s \in \mathcal{S}$:

$$\begin{aligned} &V^*(s) \\ &= \min_{a \in \mathcal{A}(s)} \left\{ c(s,a) + \gamma \sum_{s' \in \mathcal{S}} P(s'|s,a) V^*(s') \right\}, \\ &= \min_{a \in \mathcal{A}(b,e)} \Big\{ c([b,h],a) + \gamma \mathbb{E}_{l,f,e_H,h'} \\ &\quad [V^*(\min(b - f + l, N_b), \min(e - a \cdot e_{\text{TX}} + e_H, N_e), h')] \Big\} \\ &\triangleq \min_{a \in \mathcal{A}(s)} Q^*(s,a), \end{aligned} \quad (10)$$

where $\mathcal{A}(b,e)$ is the set of feasible actions given the buffer and battery states (i.e., $\mathcal{A}(b,e) = \{0,1\}$ if $b > 0$ and $e \geq e_{TX}$, and is $\{0\}$ otherwise), $V^*(s)$ is the optimal *state-value function*, and $Q^*(s,a)$ is the optimal *action-value function*. The optimal policy $\pi^*(s)$ can be determined by taking the action in each state that minimizes the r.h.s. of (10).

A. Post-Decision State Based Dynamic Programming

We will find it useful throughout our analysis to work with so-called post-decision states (PDSs) rather than conventional states. A PDS, $\widetilde{s} \triangleq (\widetilde{b}, \widetilde{e}, \widetilde{h}) \in \mathcal{S}$, denotes a state of the system after all known dynamics have occurred, but before the unknown dynamics occur [16]. In the DSEHS problem,

$$\tilde{s}^n = (\widetilde{b}^n, \widetilde{e}^n, \widetilde{h}^n) = ([b^n - f^n], [e^n - a^n \cdot e_{\text{TX}}], h^n) \quad (11)$$

is the PDS in time slot n. The buffer's PDS $\widetilde{b}^n = b^n - f^n$ characterizes the buffer state after a packet is transmitted (if any), but before any new packets arrive; the battery's PDS $\widetilde{e}^n = e^n - a^n \cdot e_{\text{TX}}$ characterizes the battery state after an energy packet is consumed (if any), but before any new energy packets arrive; and the channel's PDS $\widetilde{h}^n = h^n$ is the same as the channel state at time n. In other words, the PDS incorporates all of the known information about the transition from state s^n to state s^{n+1} after taking action a^n. Meanwhile, the unknown dynamics in the transition from state s^n to s^{n+1}, i.e., the channel state transition from h^n to $h^{n+1} \sim P^h(\cdot|h^n)$, the data packet arrivals $l^n \sim P^l(\cdot)$, and the energy packet arrivals $e_H^n \sim P^{e_H}(\cdot)$ are not included in the PDS. Importantly, the next state can be expressed in terms of the PDS as follows:

$$\begin{aligned} s^{n+1} &= (b^{n+1}, e^{n+1}, h^{n+1}) \\ &= (\min(\widetilde{b}^n + l^n, N_b), \min(\widetilde{e}^n + e_H^n, N_e), h^{n+1}). \end{aligned} \quad (12)$$

Just as we defined a value function over the conventional states, we can define a PDS value function over the PDSs. Let $\widetilde{V}^*$ denote the optimal PDS value function. $\widetilde{V}^*$ and V^* are related by the following Bellman equations:

$$\begin{aligned} &\widetilde{V}^*(\widetilde{s}) = \eta \mathbb{E}_l[\max(\widetilde{b} + l - N_b, 0)] + \\ &\gamma \mathbb{E}_{l,e_H,h'}[V^*(\min(\widetilde{b} + l, N_b), \min(\widetilde{e} + e_H, N_e), h')] \end{aligned} \quad (13)$$

$$V^*(s) = \min_{a \in \mathcal{A}(b,e)} \left\{ b + \mathbb{E}_f[\widetilde{V}^*(b - f, e - a \cdot e_{TX}, h)] \right\} \quad (14)$$

Knowing $\widetilde{V}^*(\widetilde{s})$, $\pi^*(s)$ can be found by taking the action in each state that minimizes the r.h.s. of (14).

Algorithm 1 presents a value iteration algorithm for computing the PDS value function *offline*. Although it is too complex to be implemented on an EHS, its iterative structure facilitates the use of mathematical induction to derive structural properties of the optimal PDS value function $\widetilde{V}^*(\widetilde{s})$ (see Section IV).

Algorithm 1 Post-Decision State Value Iteration

initialize $\widetilde{V}_0(\widetilde{b}, \widetilde{e}, \widetilde{h}) = 0$ for all $(\widetilde{b}, \widetilde{e}, \widetilde{h}) \in \mathcal{S}$ and $\tau = 0$
repeat
  $\Delta \leftarrow 0$
  for $(b,e,h) \in \mathcal{S}$ **do**
    Update the value function:

$$V_\tau(b,e,h) \leftarrow \min_{a \in \mathcal{A}(b,e)} \left\{ b + \mathbb{E}_f[\widetilde{V}_\tau(b - f, e - a \cdot e_{TX}, h)] \right\} \quad (15)$$

  end for
  for $(\widetilde{b}, \widetilde{e}, \widetilde{h}) \in \mathcal{S}$ **do**
    Update the PDS value function:

$$\begin{aligned} &\widetilde{V}_{\tau+1}(\widetilde{b}, \widetilde{e}, \widetilde{h}) \leftarrow \eta \mathbb{E}_l[\max(\widetilde{b} + l - N_b, 0)] + \\ &\gamma \mathbb{E}_{l,e_H,h'}[V_\tau(\min(\widetilde{b} + l, N_b), \min(\widetilde{e} + e_H, N_e), h')] \end{aligned} \quad (16)$$

    $\Delta \leftarrow \max(\Delta, |\widetilde{V}_\tau(\widetilde{b}, \widetilde{e}, \widetilde{h}) - \widetilde{V}_{\tau+1}(\widetilde{b}, \widetilde{e}, \widetilde{h})|)$
  end for
  $\tau \leftarrow \tau + 1$
until $\Delta < \theta$ (a small positive constant)

IV. Structural Properties

In this section, we analyze the structural properties of the optimal PDS value function $\widetilde{V}^*(s)$. Understanding such properties is important because: (i) they provide insights into the optimization problem and the system being optimized;

(ii) they reveal ways in which the solution can be represented compactly, with limited memory; and (iii) they can facilitate efficient *online* computation of the optimal policy using reinforcement learning (see, e.g., [14, 16]). In this paper, we focus on point (i) above. We begin by introducing three important definitions and providing an overview of our results. Then, in Section IV-A, we analyze the properties of the cost and transition probability functions and, in Section IV-B, we analyze several key properties of the conventional value function. These properties are all needed to prove our main results, which are presented in Section IV-C.

The first useful definition is that of integer convexity.

Definition 1. *(Integer Convex [17]): An integer convex function $f(n) : \mathcal{N} \to \mathbb{R}$ on a set of integers $\mathcal{N} \in \{0, 1, \ldots, N\}$ is a function that has increasing differences in n, i.e.,*

$$f(n_1 + m) - f(n_1) \leq f(n_2 + m) - f(n_2) \tag{17}$$

for $n_1 < n_2$, $n_1, n_2, n_1 + m, n_2 + m \in \mathcal{N}$.

Our main results establish that the PDS value function has increasing differences in the PDS buffer state $\widetilde{b}$ (Proposition 1) and the PDS battery state $\widetilde{e}$ (Proposition 2).

The second useful definition is that of stochastic dominance.

Definition 2. *(Stochastic Dominance [17]): Let $\theta(x)$ be a random variable parameterized by some $x \in \mathbb{R}$. If $P(\theta(x_1) \geq a) \geq P(\theta(x_2) \geq a)$ for all $x_1 \geq x_2$ and for all $a \in \mathbb{R}$, then we say that $\theta(x)$ is first-order stochastically increasing in x. If $\theta(x)$ is first-order stochastically increasing in x, then*

$$\mathbb{E}[u(\theta(x_1))] \geq \mathbb{E}[u(\theta(x_2))] \tag{18}$$

for all non-decreasing functions $u(x)$. The reverse inequality holds for all non-increasing functions $u(x)$.

In Section IV-A, we establish that the buffer and battery state transition probabilities defined in (6) and (3), respectively, are first-order stochastically increasing in the buffer state b and the battery state e, respectively. In Section IV-B, we use these properties – combined with (18) – to show that the value function is non-decreasing in the buffer state b and is non-increasing in the battery state e. These results help us establish integer convexity of the PDS value function in Section IV-C.

Lastly, we define the concept of a submodular function.

Definition 3. *(Submodular [8]): A submodular function $f(x, y) : \mathcal{X} \times \mathcal{Y} \to \mathbb{R}$ on sets of integers $\mathcal{X} \in \{0, 1, \ldots, X\}$ and $\mathcal{Y} \in \{0, 1, \ldots, Y\}$ is a function that has decreasing differences in (x, y), i.e., for $x^+ \geq x^-$ and $y^+ \geq y^-$*

$$f(x^+, y^+) - f(x^+, y^-) \leq f(x^-, y^+) - f(x^-, y^-). \tag{19}$$

In Section IV-C, we prove that the PDS value function is submodular in $(\widetilde{b}, \widetilde{e})$ (Proposition 3).

A. Properties of the Cost and Transition Probability Functions

We now present key properties of the cost and transition probability functions that we will need for our main results. Recall that the cost does not directly depend on the battery state e, but that the action a is constrained to be in the set $\mathcal{A}(b, e)$ (i.e., a is constrained to be 0 if $e < e_{TX}$ or $b = 0$). To show this explicitly, we define an auxiliary cost function

$$d([b, e, h], a) = \begin{cases} c([b, h], a), & \text{if } b > 0 \text{ and } e \geq e_{TX} \\ c([b, h], 0), & \text{otherwise,} \end{cases} \tag{20}$$

where $c([b, h], a)$ is defined in (7). We omit the proofs of the following three lemmas due to space limitations.

Lemma 1. *The auxiliary cost $d([b, e, h], a)$ satisfies the following properties:*

1) *The auxiliary cost is non-decreasing in b.*
2) *The auxiliary cost is non-increasing in e.*

Since the auxiliary cost function satisfies Lemma 1, the cost function $c([b, h], a)$, with $a \in \mathcal{A}(b, e)$, also satisfies it.

Lemma 2. *The battery state transition probabilities are first-order stochastically increasing in the battery state e, i.e.,*

$$\sum_{e' \geq \bar{e}} P^e(e'|e + 1, a) \geq \sum_{e' \geq \bar{e}} P^e(e'|e, a), \quad 0 \leq e < N_e.$$

Lemma 3. *The buffer state transition probabilities are first-order stochastically increasing in the buffer state b, i.e.,*

$$\sum_{b' \geq \bar{b}} P^b(b'|[b + 1, h], a) \geq \sum_{b' \geq \bar{b}} P^b(b'|[b, h], a), \quad 0 \leq b < N_b.$$

Lemma 2 (Lemma 3) implies that the next battery (buffer) state has a higher probability of exceeding a threshold if the current battery (buffer) state is larger.

B. Properties of the Conventional State Value Function

Lemma 4. *The optimal value function $V^*(b, e, h)$ is non-decreasing in the buffer state b.*

Proof. The proof is given in the appendix. □

Lemma 5. *The optimal value function $V^*(b, e, h)$ is non-increasing in the battery state e.*

Proof. We omit the proof as it is similar to Lemma 4. □

The following lemma is needed for the inductive steps in our main results (propositions 1, 2, and 3).

Lemma 6. *The following properties are propagated from the PDS value function $\widetilde{V}(\widetilde{b}, \widetilde{e}, \widetilde{h})$ to the conventional value function $V(b, e, h)$ through the Bellman equation given in* (14)*:*

1) *If $\widetilde{V}(\widetilde{b}, \widetilde{e}, \widetilde{h})$ has increasing differences $\widetilde{b}$, then $V(b, e, h)$ has increasing differences in b.*
2) *If $\widetilde{V}(\widetilde{b}, \widetilde{e}, \widetilde{h})$ has increasing differences in $\widetilde{e}$, then $V(b, e, h)$ has increasing differences in e.*
3) *If $\widetilde{V}(\widetilde{b}, \widetilde{e}, \widetilde{h})$ is submodular in $(\widetilde{b}, \widetilde{e})$, then $V(b, e, h)$ is submodular in (b, e).*

Proof. The proof is given in the appendix. □

Lemma 6 implies that the PDS value function's properties are propagated to the conventional value function during the value function update step in Algorithm 1 (see (15)).

C. Properties of the Post-Decision State Value Function

We now prove that the optimal PDS value function has increasing differences in the buffer's PDS $\widetilde{b}$ and the battery's PDS $\widetilde{e}$, and decreasing differences in $(\widetilde{b},\widetilde{e})$ (i.e., it is submodular in $(\widetilde{b},\widetilde{e})$). We then discuss the meaning of these results.

Proposition 1. *If the packet buffer has infinite size ($N_b = \infty$), then $\widetilde{V}^*(\widetilde{b},\widetilde{e},\widetilde{h})$ has increasing differences in $\widetilde{b}$, i.e.,*

$$\widetilde{V}^*(\widetilde{b},\widetilde{e},\widetilde{h}) - \widetilde{V}^*(\widetilde{b}-1,\widetilde{e},\widetilde{h}) \leq \widetilde{V}^*(\widetilde{b}+1,\widetilde{e},\widetilde{h}) - \widetilde{V}^*(\widetilde{b},\widetilde{e},\widetilde{h}). \quad (21)$$

Proof. The proof is given in the appendix. □

Proposition 2. *$\widetilde{V}^*(\widetilde{b},\widetilde{e},\widetilde{h})$ has increasing differences in $\widetilde{e}$, i.e.,*

$$\widetilde{V}^*(\widetilde{b},\widetilde{e},\widetilde{h}) - \widetilde{V}^*(\widetilde{b},\widetilde{e}-1,\widetilde{h}) \leq \widetilde{V}^*(\widetilde{b},\widetilde{e}+1,\widetilde{h}) - \widetilde{V}^*(\widetilde{b},\widetilde{e},\widetilde{h}). \quad (22)$$

Proof. We omit the proof as it is similar Proposition 1. □

Proposition 3. *$\widetilde{V}^*(\widetilde{b},\widetilde{e},\widetilde{h})$ is submodular in $(\widetilde{b},\widetilde{e})$, i.e.,*

$$\widetilde{V}^*(\widetilde{b}+1,\widetilde{e}+1,\widetilde{h}) - \widetilde{V}^*(\widetilde{b},\widetilde{e}+1,\widetilde{h}) \leq \widetilde{V}^*(\widetilde{b}+1,\widetilde{e},\widetilde{h}) - \widetilde{V}^*(\widetilde{b},\widetilde{e},\widetilde{h}). \quad (23)$$

Proof. The proof is given in the appendix. □

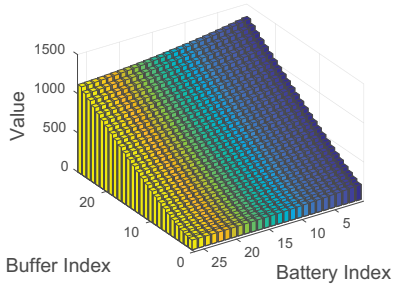
Together, Proposition 1 and Lemma 4 imply that the cost to serve an additional data packet increases with the queue backlog. Although we were only able to prove that $\widetilde{V}^*(\widetilde{b},\widetilde{e},\widetilde{h})$ has increasing differences in the buffer state for an infinite size buffer, we have not observed any cases in practice where this property does not hold for finite buffers. Together, Proposition 2 and Lemma 5 imply that the benefit of an additional energy packet decreases with the available battery energy. Finally, Proposition 3 implies that data packets and energy packets are *complementary*. That is, the cost of serving an additional data packet is smaller when more energy is available, and the benefit of having an additional energy packet is greater when more data packets need to be served.

V. Numerical Results

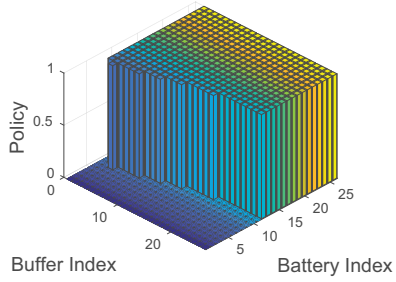
In Section V-A, we illustrate the structural properties of the optimal PDS value function. In Section V-B, we compare the optimal scheduling policy against a so-called greedy policy, which always transmits backlogged packets if there is sufficient energy (i.e., $e^n \geq e_{TX}$). The parameters used in our MATLAB-based simulator are given in Table I.

A. Structural Properties

In this section, we assume that the packet and energy arrivals are Bernoulli random variables with parameters 0.4 and 0.7, respectively. Fig. 2a and Fig. 2b show the optimal PDS value function and policy, respectively, under these assumptions. From Fig. 2a, it is clear that the optimal PDS value function (i) is non-decreasing and has increasing differences in the queue backlog (Lemma 4 and Proposition 1) and (ii) is non-increasing and has increasing differences in the battery state (Lemma 5 and Proposition 2). Fig. 3 shows that $\widetilde{V}(\widetilde{b}+1,\widetilde{e},\widetilde{h}) - \widetilde{V}(\widetilde{b},\widetilde{e},\widetilde{h})$ is non-increasing in the battery state $\widetilde{e}$, i.e., the optimal PDS value function is submodular in $(\widetilde{b},\widetilde{e})$ (Proposition 3). From Fig. 2b, we observe that the optimal policy is more conservative than the greedy policy because it does not transmit at low battery states.



(a) Optimal PDS value function



(b) Optimal policy

Fig. 2: Optimal PDS value function and policy in channel state h with PLR $q(h) = 0.8$.

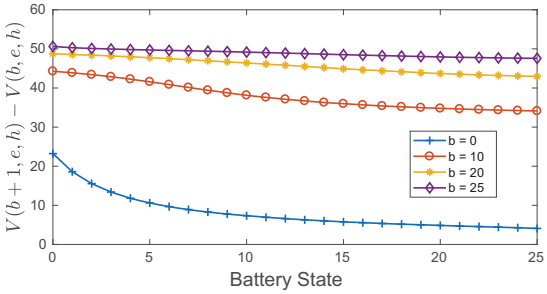


Fig. 3: Submodularity of the PDS value function in (b,e) with PLR $q(h) = 0.8$.

B. Performance Evaluation

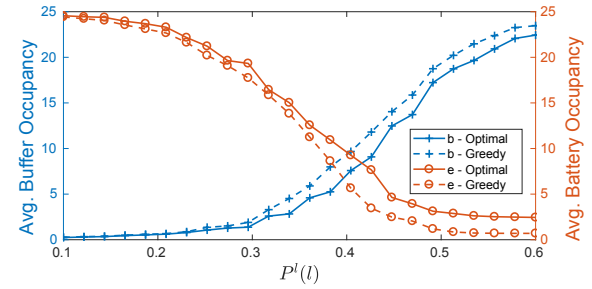
We now compare the performance of the optimal and greedy policies assuming that $P^l(l) = \text{Bernoulli}(p)$, where $p \in \{0.1, 0.122, 0.144, \ldots, 0.6\}$, $P^{e_H}(e_H) = \text{Bernoulli}(0.7)$, and $q(h) = 0.8$. Note that the optimal policies were computed offline using Algorithm 1 and then stored in a lookup table. In Fig. 4a, we show how the average queue backlog (left axis) and average battery state (right axis) vary with respect to the packet arrival rate. Each measurement is taken from a 50,000 time slot simulation of the corresponding policy. The optimal policy achieves 2.6% – 37.4% lower queue backlogs (19.1% lower

TABLE I: Simulation Parameters

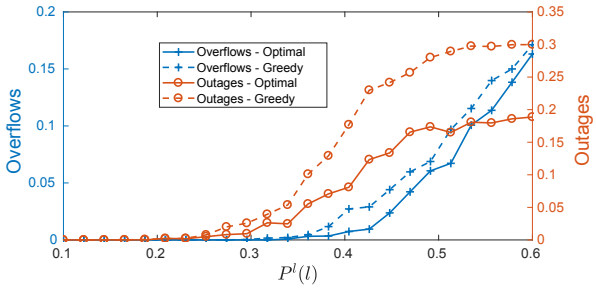
Parameter	Value	Parameter	Value
Packet Buffer Size, N_b	25	Transmission Action, $a \in \mathcal{A}$	$\{0, 1\}$
Energy Buffer Size, N_e	25	Transmission Energy, e_{TX}	1
Channel States $h \in \mathcal{H}$	$\{1, 2, ..., 7, 8\}$	Discount Factor, γ	0.98
Packet Loss Rate (PLR), $q(h)$	$\{0.8, 0.7, 0.6, ..., 0.2, 0.1\}$	Simulation Duration (time slots	50,000
Packet Arrivals (packets/time slot)	$\{0, 1\}$	Packet Arrival PMF, $P^l(l)$	Bernoulli(p) with variable p
Energy Arrivals (packets/time slot)	$\{0, 1\}$	Energy Arrival PMF, $P^{e_H}(e_H)$	Bernoulli(p) with $p = 0.7$
Overflow Penalty, η	50	Steady-State Channel Probabilities	(0.071, 0.143, 0.143, 0.143, 0.143, 0.143, 0.143, 0.071)

when averaged across all data points) and maintains 0.1% – 258.5% more battery energy (71.1% higher when averaged across all data points) than the greedy policy.

In Fig. 4b, we show how the buffer overflow (left axis) and battery outage (right axis) probabilities vary with respect to the packet arrival rate. The optimal policy achieves 37.4% – 100.0% lower outage probabilities (75.3% when averaged across all data points) and achieves 4.7% to 100.0% fewer overflows (47.62% when averaged across all data points) than the greedy policy.



(a) Average queue backlog and battery state vs. packet arrival rate



(b) Average battery outages and average overflows vs. packet arrival rate

Fig. 4: Comparison of the optimal and greedy policies.

VI. Conclusion

We formulated the DSEHS problem as an MDP and analyzed its structural properties. Our analysis does not assume specific data and energy arrival distributions, save for they are i.i.d., and does not require assumptions on the channel transition probabilities, save that they are Markovian. This makes our structural results broadly applicable. We demonstrate that the optimal scheduling policy achieves fewer battery outages, fewer packet overflows, and a better energy-delay trade-off than a greedy policy. As future work, we plan to leverage the structural properties of the DSEHS problem to develop low-complexity reinforcement learning algorithms that can find the optimal scheduling policy with no a priori knowledge of the packet arrival, energy harvesting, and channel dynamics.

Appendix

A. Proof of Lemma 4

The proof follows by induction. Since value iteration converges for any initialization, select $V_0([b, e, h])$ to be non-decreasing in b. Assume that $V_t([b, e, h])$ is non-decreasing in b. We prove that $V_{t+1}([b, e, h])$ is also non-decreasing in b. By definition

$$\begin{aligned} &V_{t+1}([b, e, h]) \\ &= \min_{a \in \mathcal{A}(b,e)} \Big\{ c([b, h], a) + \gamma \mathbb{E}_{b',e',h'}[V_t([b', e', h'])] \Big\}. \\ &= \min_{a \in \mathcal{A}(b,e)} Q_{t+1}([b, e, h], a) \end{aligned}$$

In Lemma 1.1, we established that the cost function $c([b, h], a)$ is non-decreasing in b. Additionally, since $V_t([b, e, h])$ is non-decreasing in b by the induction hypothesis, and $P^b(b'|[b, h], a)$ is stochastically increasing in b by Lemma 3, the expected future value is non-decreasing in b. It follows that $Q_{t+1}([b, e, h], a)$ is also non-decreasing in b.

Let a^* be the optimal action in state $(b + 1, e, h)$. We have

$$\begin{aligned} V_{t+1}([b + 1, e, h]) &= Q_{t+1}([b + 1, e, h], a^*) \\ &\geq Q_{t+1}([b, e, h], a^*) \\ &\geq \min_{a \in \mathcal{A}} Q_{t+1}([b, e, h], a) \\ &= V_{t+1}([b, e, h]) \end{aligned}$$

where the first inequality follows from the fact that $Q_{t+1}([b, e, h], a)$ is non-decreasing in b and the second inequality follows from optimality. Thus, the optimal value function V^* is non-decreasing in the buffer state b.

B. Proof of Lemma 6

We may express the value function defined in (14) as

$$\begin{aligned} V(b, e, h) = &\min_{a \in \mathcal{A}(b,e)} \Big\{ b + \mathbb{E}_f[\widetilde{V}^*(b - f, e - a \cdot e_{TX}, h)] \Big\} \\ &= b + (1 - a^*)\widetilde{V}([b, e, h]) + \\ &\quad a^* q(h) \widetilde{V}([b, e - e_{TX}, h]) + \\ &\quad a^*(1 - q(h))\widetilde{V}([b - 1, e - e_{TX}, h]), \end{aligned}$$

where $a^* \in \{0,1\}$ is the optimal action in state (b,e,h) and $q(h) \in [0,1]$ is the packet loss rate. If the PDS value function $\widetilde{V}(\widetilde{b},\widetilde{e},\widetilde{h})$ (i) has increasing differences in $\widetilde{b}$, (ii) has increasing differences in $\widetilde{e}$, or (iii) has decreasing differences in $(\widetilde{b},\widetilde{e})$, then the results follow from the fact that a non-negative weighted sum of functions with increasing (decreasing) differences has increasing (decreasing) differences.

C. Proof of Proposition 1

Consider the value iteration algorithm, which converges for any initial condition. Initialize the PDS value function $\widetilde{V}_0(\widetilde{b},\widetilde{e},\widetilde{h})$ to satisfy (21). Assume that (21) holds for $\widetilde{V}_t(\widetilde{b},\widetilde{e},\widetilde{h})$, for some $t>0$. We aim to show that (21) holds for $\widetilde{V}_{t+1}(\widetilde{b},\widetilde{e},\widetilde{h})$. Recall from (13) that the PDS value function can be expressed as a function of the conventional value function. The first term on the r.h.s. of (13) has increasing differences in $\widetilde{b}$. Thus, we only need to show that the the second term on the r.h.s. of (13) has increasing differences in $\widetilde{b}$. This is implied if the following condition holds:

$$\begin{aligned} &V_t([\widetilde{b}+l]^{N_b},e',h') - V_t([\widetilde{b}-1+l]^{N_b},e',h') \\ &\quad \le V_t([\widetilde{b}+1+l]^{N_b}),e',h') - V_t([\widetilde{b}+l]^{N_b},e',h'), \quad (24) \end{aligned}$$

where $[x]^N = \min(x,N)$ and $e' = \min(\widetilde{e}+e_H, N_e)$. If we let $N_b = \infty$, then (24) reduces to

$$\begin{aligned} &V_t(\widetilde{b}+l,e',h') - V_t(\widetilde{b}-1+l,e',h') \\ &\qquad \le V_t(\widetilde{b}+1+l,e',h') - V_t(\widetilde{b}+l,e',h'), \end{aligned}$$

which holds by Lemma 6.1. That concludes the proof.

D. Proof of Proposition 3

Consider the value iteration algorithm, which converges for any initial condition. Initialize the PDS value function $\widetilde{V}_0(\widetilde{b},\widetilde{e},\widetilde{h})$ to satisfy (23). Assume that (23) holds for $\widetilde{V}_t(\widetilde{b},\widetilde{e},\widetilde{h})$, for some $t>0$. We aim to show that (23) holds for $\widetilde{V}_{t+1}(\widetilde{b},\widetilde{e},\widetilde{h})$. Recall from (13) that the PDS value function can be expressed as a function of the conventional value function. The first term on the r.h.s. of (13) is submodular in $(\widetilde{b},\widetilde{e})$. Thus, we only need to show that the expected future value (i.e., the second term on the r.h.s. of (13)) is submodular in $(\widetilde{b},\widetilde{e})$. This is implied by the following condition

$$\begin{aligned} &V_t([b''+1]^{N_b},[e''+1]^{N_e},h') - V_t([b'']^{N_b},[e''+1]^{N_e},h') \\ &\le V_t([b''+1]^{N_b},[e'']^{N_e},h') - V_t([b'']^{N_b},[e'']^{N_e},h'), \quad (25) \end{aligned}$$

where we use $[x]^N \triangleq \min(x,N)$, $b'' \triangleq \widetilde{b}+l$ and $e'' \triangleq \widetilde{e}+e_H$ to keep the equations compact. To verify that (25) holds, we consider the following two cases.

Case 1 ($b''+1 \le N_b$): Assuming that $b''+1 \le N_b$, we may rewrite (25) as follows:

$$\begin{aligned} &V_t(b''+1,[e''+1]^{N_e},h') - V_t(b'',[e''+1]^{N_e},h') \\ &\qquad \le V_t(b''+1,[e'']^{N_e},h') - V_t(b'',[e'']^{N_e},h'). \end{aligned}$$

If $e''+1 \le N_e$, then the condition holds by Lemma 6.3 and, if $e''+1 > N_e$, then both sides are equal; thus, Case 1 holds.

Case 2 ($b'' \ge N_b$): Assuming that $b'' \ge N_b$, we may rewrite (25) as follows:

$$\begin{aligned} &V_t(N_b,[e''+1]^{N_e},h') - V_t(N_b,[e''+1]^{N_e}),h') \\ &\qquad \le V_t(N_b,[e'']^{N_e},h') - V_t(N_b,[e'']^{N_e},h'), \end{aligned}$$

where both sides are equal to 0; thus, Case 2 holds. This concludes the proof.

REFERENCES

[1] J. Chakareski, "Uplink scheduling of visual sensors: When view popularity matters," *IEEE Trans. Commun.*, vol. 2, no. 63, pp. 510–519, Feb. 2015.

[2] A. Seyedi and B. Sikdar, "Energy efficient transmission strategies for body sensor networks with energy harvesting," *IEEE Trans. Commun.*, vol. 58, no. 7, pp. 2116–2126, 2010.

[3] R. J. Vullers, R. Van Schaijk, H. J. Visser, J. Penders, and C. Van Hoof, "Energy harvesting for autonomous wireless sensor networks," *IEEE Solid State Circuits Mag.*, vol. 2, no. 2, pp. 29–38, 2010.

[4] B. Gurakan and S. Ulukus, "Energy harvesting multiple access channel with data arrivals," in *IEEE GLOBECOM*, 2015.

[5] X. Lu, P. Wang, D. Niyato, and E. Hossain, "Dynamic spectrum access in cognitive radio networks with rf energy harvesting," *Wireless Commun.*, vol. 21, no. 3, pp. 102–110, 2014.

[6] V. Sharma, U. Mukherji, V. Joseph, and S. Gupta, "Optimal energy management policies for energy harvesting sensor nodes," *IEEE Trans. Wireless Commun.*, vol. 9, no. 4, 2010.

[7] D. Gunduz, K. Stamatiou, N. Michelusi, and M. Zorzi, "Designing intelligent energy harvesting communication systems," *IEEE Communications Magazine*, vol. 52, no. 1, pp. 210–216, 2014.

[8] M. L. Puterman, *Markov decision processes: discrete stochastic dynamic programming*. John Wiley & Sons, 2014.

[9] O. Ozel, K. Tutuncuoglu, J. Yang, S. Ulukus, and A. Yener, "Transmission with energy harvesting nodes in fading wireless channels: Optimal policies," *IEEE J. Sel. Areas Commun.*, vol. 29, no. 8, pp. 1732–1743, 2011.

[10] C. Ho and R. Zhang, "Optimal energy allocation for wireless communications with energy harvesting constraints," *IEEE Trans. Signal Process.*, vol. 60, no. 9, pp. 4808–4818, 2012.

[11] J. Yang and S. Ulukus, "Optimal packet scheduling in a multiple access channel with energy harvesting transmitters," *Journal of Communications and Networks*, vol. 14, no. 2, pp. 140–150, 2012.

[12] ——, "Optimal packet scheduling in an energy harvesting communication system," *IEEE Trans. Commun.*, vol. 60, no. 1, pp. 220–230, 2012.

[13] A. Aprem, C. R. Murthy, and N. B. Mehta, "Transmit power control policies for energy harvesting sensors with retransmissions," *IEEE J. Sel. Topics Signal Process.*, vol. 7, no. 5, pp. 895–906, 2013.

[14] N. Toorchi, J. Chakareski, and N. Mastronarde, "Fast and low-complexity reinforcement learning for delay-sensitive energy harvesting wireless visual sensing systems," in *IEEE ICIP*, 2016, pp. 1804–1808.

[15] D. Zordan, T. Melodia, and M. Rossi, "On the design of temporal compression strategies for energy harvesting sensor networks," *IEEE Trans. Wireless Commun.*, vol. 15, no. 2, pp. 1336–1352, Feb 2016.

[16] N. Mastronarde and M. van der Schaar, "Fast reinforcement learning for energy-efficient wireless communication," *IEEE Trans. Signal Process.*, vol. 59, no. 12, pp. 6262–6266, 2011.

[17] D. V. Djonin and V. Krishnamurthy, "Mimo transmission control in fading channels- a constrained markov decision process formulation with monotone randomized policies," *IEEE Trans. Signal Process.*, vol. 55, no. 10, pp. 5069–5083, 2007.

[18] D. P. Bertsekas, R. G. Gallager, and P. Humblet, *Data networks*. Prentice-hall Englewood Cliffs, NJ, 1987, vol. 2.