Low-Complexity Scheduling Policies for Energy Harvesting Communication Networks

Pol Blasco[1], Deniz Gündüz[2] and Mischa Dohler[1]
[1]CTTC, Barcelona, Spain
[2]Imperial College London, London, UK
Emails: {pol.blasco, mischa.dohler}@cttc.es, d.gunduz@imperial.ac.uk

***Abstract*—A time-slotted multiple access wireless system with N transmitting nodes, each equipped with an energy harvesting (EH) device and a rechargeable battery of finite capacity, is studied. The energy arrival process at each node is modeled as an independent two-state Markov process, such that a node either harvests one unit of energy, or none, at each time slot (TS). The access point (AP) schedules a subset of K nodes to transmit over K orthogonal channels at each TS. The maximum total throughput is studied for a backlogged system without the knowledge of the EH processes and nodes' battery states at the AP. The problem is identified as a partially observable Markov decision process, and the optimal policy for the general model is studied numerically. Under certain assumptions regarding the EH processes and the battery sizes, the optimal scheduling policy is characterized explicitly, and is shown to be myopic.**

I. Introduction

In energy harvesting (EH) communication systems energy arrivals are commonly modeled as Markov processes. Depending on the available information about the state of the underlying Markov processes, the problem can be modeled as either a Markov decision process (MDP) [1], or a partially observable MDP (POMDP) [2], and tools from dynamic programming (DP) can be invoked in both cases to determine the optimal scheduling policy. However, the computational complexity of DP increases with the state space (e.g., number of nodes, number of states of the EH process, etc.), and the numerical simulations do not provide much intuition about the structure of the optimal policies. It is important to characterize the behavior of the optimal scheduling policies; however, this is possible only in some special cases.

In this paper, we consider N EH nodes that transmit data to an access point (AP). The AP is in charge of scheduling the nodes to K orthogonal channels available for transmission in each time slot (TS). At each TS a node either harvests one energy unit or does not harvest any, following a Markov process. The AP is interested in finding the scheduling policy that maximizes the total throughput over T TSs assuming a backlogged system (the nodes always have data available for transmission) without the knowledge of the states of the EH processes or the nodes' batteries.

For this multiple access system with EH nodes, we prove the optimality of the myopic policy (MP) in two special cases: 1)

This work was made possible by NPRP grant No. 5 - 1047 - 2 437 from the Qatar National Research Fund (a member of The Qatar Foundation). The statements made herein are solely the responsibility of the authors. The work of Pol Blasco is also partially supported by Spanish grant TEC2010-21100/FPU AP 2009-5009.

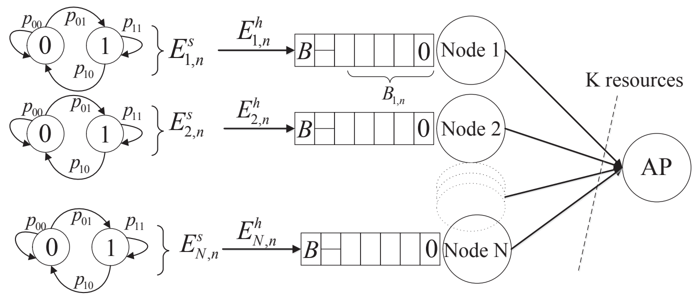


Figure 1. System model with N nodes and one AP.

the nodes are not able to harvest and transmit simultaneously, and the EH process transition probabilities are affected by the scheduling decisions, and 2) the nodes have no battery.

Optimal scheduling policies for EH communication systems are also studied in [3] and [4] for single and multi-user systems, respectively. In [3] the battery state of the node is modeled by a Markov process, and under the assumption that the battery state is known at the scheduler, the optimal transmission scheme is shown to be achieved by a threshold-based policy. In [4] EH transmitters with a unit size battery is considered, and the optimality of MP is proven under certain assumptions on the EH processes. Previously, optimality of MP was also shown for throughput optimization in a multi-user system with Markov channel states in [5].

II. System Model

The system model studied in this paper is depicted in Figure 1. We have N nodes and K resources to be allocated to theses nodes at each TS. For simplicity, we assume that $M = N/K$ is an integer, which implies that the nodes can be grouped into M sets of K nodes. The EH processes at different nodes are independent from each other, and modeled by an identical two-state Markov process with state space $\mathcal{E} = \{0, 1\}$. The transition probability of the EH process from state k to state j is p_{kj}, with $k, j \in \mathcal{E}$. We denote by $E^s_{i,n}$ and $E^h_{i,n}$ the state of the EH process associated with node i, and the amount of energy harvested by node i, respectively, in TS n. The energy harvested in TS n is stored in the battery, and is available for transmission at the beginning of TS $n+1$. Therefore, $E^h_{i,n} = 1$ when $E^s_{i,n+1} = 1$, which occurs with probability p_{11} or p_{01} if $E^s_{i,n} = 1$ or $E^s_{i,n} = 0$, respectively, and equal to zero otherwise.

The batteries have a limited capacity of B energy units, and the amount of energy in the battery (i.e. battery state) of node i

at TS n is denoted by $B_{i,n} \in [0, B]$. The joint Markov process governing the EH and battery states is depicted in Figure 2.

We consider nodes operating in the low-power regime, and hence, assume a linear rate-power function. Accordingly, when a node is scheduled, it transmits at a constant rate throughout the TS, using all the available energy in its battery. The amount of data transmitted by node i in TS n is denoted by $D_i(n)$, and given by $D_i(n) = B_{i,n}\mathbb{1}_{i\in\mathcal{K}(n)}$, where $\mathcal{K}(n)$ is the set of nodes scheduled in TS n, and $\mathbb{1}_a$ is the indicator function. Then the battery state of node i at the beginning of TS n is

$$B_{i,n} = \min(B_{i,n-1} + E^h_{i,n-1}, B)\mathbb{1}_{i\notin\mathcal{K}(n-1)} + E^h_{i,n-1}\mathbb{1}_{i\in\mathcal{K}(n-1)}. \tag{1}$$

The objective is to find the scheduling policy $\mathcal{K}(n)$ that maximizes the expected sum throughput,

$$\max_{\{\mathcal{K}(n)\}_{n=1}^{T}} \mathbb{E}\left[\sum_{n=1}^{T}\beta^{n-1}\sum_{i\in\mathcal{K}(n)} D_i(n)\right], \quad \text{s.t. (1)}, \tag{2}$$

where $0 < \beta \le 1$ is the discount factor. If $\beta < 1$, (2) models the case in which, at each TS either the AP is switched-off or its receiver is blocked with probability β.

If the AP knows $B_{i,n}$ and $E^s_{i,n}$ of each node at each TSs and the EH process transition probabilities, the system can be modeled as a finite-state MDP. However, we take a more practically relevant approach, and assume that the AP does not know either $B_{i,n}$ or $E^s_{i,n}$; in which case, the appropriate model is a POMDP [6].

A POMDP is defined via the quintuplet $\langle\mathcal{S}, \mathcal{A}, p_{a_i}(s_j, s_k), R_{a_i}(s_j, s_k), \mathcal{H}\rangle$, where $\mathcal{S}$ is the set of possible states, $\mathcal{H}$ is the set of observations, $\mathcal{A}$ is the set of actions, $p_{a_i}(s_j, s_k)$ denotes the transition probability from state s_j to state s_k when action a_i is taken, and $R_{a_i}(s_j, s_k)$ is the immediate reward yielded when in state s_j action a_i is taken and the state changes to s_k. In our model the state of the system in TS n is formed by the individual states of the nodes. Let $S_{i,n} = (E^s_{i,n}, B_{i,n}) \in \mathcal{E} \times [0, B]$ denote the state of node i in TS n, and let $S_n = (S_{1,n}, \dots, S_{N,n}) \in \mathcal{S}$ denote the state of the whole system in TS n. The set of actions is $\mathcal{A} = \{a_1, \dots, a_{|\mathcal{A}|}\}$, and each action corresponds to scheduling a set of K nodes. In particular, the action taken at TS n is denoted by $\mathcal{K}(n) \in \mathcal{A}$. In a POMDP, the state S_n is not known in TS n, and it can be shown that a sufficient statistic for optimal decision making is given by the conditional probability of a state S_n given all the past actions and observations [7].

We denote by $W_i^{k,j}(n)$ the belief that $B_{i,n} = j$ and $E^s_{i,n} = k$; that is, the probability that node i has j energy units in the battery and its EH process is in state k in TS n. Let $\mathbf{W}_i(n) = (W_i^{0,0}(n), W_i^{1,0}(n), \dots, W_i^{0,B}(n), W_i^{1,B}(n))$ be the vector of beliefs over the states of node i at TS n. The joint belief state over all states of all nodes is denoted by $\bar{\mathbf{W}}(n) = (\mathbf{W}_1(n), \dots, \mathbf{W}_N(n))$, and $\bar{\mathbf{W}}(n) \in \mathcal{W}$. Let $\bar{\mathbf{W}}(0)$ be the initial belief state, which is a probability distribution over $\mathcal{S}$. Notice that, in general, the belief state space $\mathcal{W}$ is uncountable. The objective of a POMDP is to find the optimal transmission policy $\pi(\cdot) : \mathcal{W} \to \mathcal{A}$ that maximizes the expected discounted sum reward (i.e., the total throughput in (2)). In our problem, the immediate reward function is $R_{\mathcal{K}(n)}(S_n, S_{n+1}) = \sum_{i\in\mathcal{K}(n)} D_i(n)$, and the expected sum discounted reward is equivalent to (2), where $\mathcal{K}(n) = \pi(\bar{\mathbf{W}}(n))$ is the action taken by the AP when the system is in state $\bar{\mathbf{W}}(n)$.

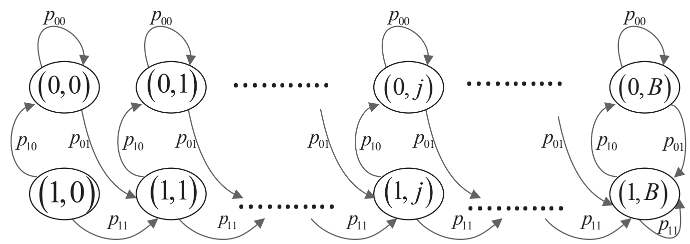


Figure 2. State transition probabilities among the node states $S_{i,n} = (E^s_{i,n}, B_{i,n})$.

In general, finding the optimal policy of a POMDP using numerical methods such as DP is PSPACE-complete (or worse) [6]. There are algorithms to find approximate optimal polices for POMDP; however, this is also hard (NP-complete or worse). In Sections III-IV we focus on two particular settings of the problem in (2), and show that, under certain circumstances regarding the EH processes and the battery sizes, there exist optimal low-complexity scheduling policies, which can be characterized explicitly.

III. Non Simultaneous Energy Harvesting and Data Transmission

Assume that nodes are not able to harvest energy and transmit simultaneously. This may account, for example, for electromagnetic harvesting systems in which the same antenna is used for harvesting and transmission. Moreover, we assume that when node i is scheduled in TS n the EH process state in TS $n+1$, $E^s_{i,n+1}$, is either 0 or 1 with probabilities p_0 and p_1, respectively, independent of the previous EH state $E^s_{i,n}$, with $p_1 \ge \frac{p_{01}}{p_{10}+p_{01}}$. This may account for the fact that, in an electromagnetic harvesting system, the RF hardware needs to be set into the harvesting mode after transmission.

If $i \in \mathcal{K}(n-1)$, the belief state for node i in TS n is

$$W_i^{k,j}(n) = \begin{cases} p_0 & \text{if } j = 0 \text{ and } k = 0, \\ 1 - p_0 & \text{if } j = 0 \text{ and } k = 1, \\ 0 & \text{otherwise.} \end{cases} \tag{3}$$

Notice that the belief state $\mathbf{W}_i(n)$ is the same for all $i \in \mathcal{K}(n-1)$, and it is denoted by $\mathbf{W}_0 = (p_0, 1-p_0, 0, \dots, 0)$. If $i \notin \mathcal{K}(n-1)$, the belief state of node i in TS n is

$$W_i^{k,j}(n) = \begin{cases} W_i^{0,j}(n-1)p_{00} + W_i^{1,j}(n-1)p_{10} & \text{if } k = 0 \text{ and } j \ne 0, \\ W_i^{1,0}(n-1)p_{10} + W_i^{0,0}(n-1)p_{00} & \text{if } k = 0 \text{ and } j = 0, \\ W_i^{1,j-1}(n-1)p_{11} + W_i^{0,j-1}(n-1)p_{01} & \text{if } k = 1 \text{ and } j \ne B, \\ W_i^{1,B-1}(n-1)p_{11} + W_i^{1,B}(n-1)p_{11} \\ \quad + W_i^{0,B-1}(n-1)p_{01} + W_i^{0,B}(n-1)p_{01} & \text{if } k = 1 \text{ and } j = B. \end{cases} \tag{4}$$

According to (3) and (4), the belief state transitions are deterministic, since for each $\mathcal{K}(n)$ and belief state $\bar{\mathbf{W}}(n)$ there is only one possible future belief state $\bar{\mathbf{W}}(n+1)$.

Notice that the belief state in TS n can be written as a function of the past actions and observations. In particular, if $i \in \mathcal{K}(n)$ and $i \notin \mathcal{K}(m), \forall m \in \{n+1, \cdots, n+l_i-1\}$, i.e., node i is scheduled once in TS n and not scheduled in the next l_i TSs, the belief state for node i in $TS\ n+l_i$ is $\mathbf{W}_i(n+l_i) = f(l_i, \mathbf{W}_0)$, where the function $f(\cdot,\cdot)$ is derived in Appendix A. We assume that in TS 0 the AP has initial belief $\bar{\mathbf{W}}(0)$, where $\mathbf{W}_i(0) = f(l_i^0, \mathbf{W}_0)$ for $i = \{1 \dots N\}$. Notice that this is a standard technique for turning a POMDP into a classic MDP by means of the belief states, which has the drawback of having an uncountable state space $\mathcal{W}$.

Now we present a transformation of the belief space $\mathcal{W}$ to a set $\mathcal{R}$, which is the expected throughput for each belief state. Let $r_i(n)$ denote the expected throughput for node i in TS n

$$r_i(n) = \sum_{j=1}^{B} j \left(f^{0,j}(l_i^n, \mathbf{W}_0) + f^{1,j}(l_i^n, \mathbf{W}_0) \right), \tag{5}$$

where l_i^n is the number of TSs since node i was last scheduled, and $f^{k,j}(l, \mathbf{W}_0)$ is the probability that the node i is in state (k,j) given that it was last scheduled l TSs before. Furthermore we denote by $\mathbf{L}^n$ the vector $(l_1^n, \dots, l_N^n)$, and by $\mathbf{r}(n)$ the vector $(r_1(n), \dots, r_N(n)) \in \mathcal{R}$. Notice that, similar to $\mathcal{W}$, $\mathcal{R}$ is also uncountable.

The throughput at each state depends only on the number of TSs since it was last scheduled. We define a function $g : \mathbb{N} \to \mathcal{R}$ such that $r_i(n) = g(l_i^n)$. We notice that, due to the special structure of (5) and (13), the function $g(l)$ is monotonically increasing in l, and is a contraction maping, such that $|g(l+1)-g(m+1)| \leq |g(l)-g(m)|$. The intuitive explanation is that if a node is not scheduled its expected throughput increases in time and converges to the expected throughput of the full battery state.

Proposition 1. *There is a one to one correspondence between the belief state $\bar{\mathbf{W}}(n)$ and the expected throughput $\mathbf{r}(n)$.*

Proof. For each possible $\mathbf{W}_i(n)$ there exist only one l_i^n such that $\mathbf{W}_i(n) = f(l_i^n, \mathbf{W}_0)$. This ensures the one to one correspondence between $\bar{\mathbf{W}}(n)$ and $\mathbf{L}^n$. Hence, due to the fact that $g(l_i^n)$ is monotonically increasing in l_i^n, there is one to one correspondence between $\mathbf{L}^n$ and $\mathbf{r}(n)$. ☐

Given Proposition 1 we can equivalently work with the expected throughput space, $\mathcal{R}$, rather than the beliefs on the original state space, $\mathcal{W}$. The immediate reward is of this POMDP is $R_{\pi(\mathbf{r}(n))}(\mathbf{r}(n), \mathbf{r}(n+1)) = \sum_{i \in \mathcal{K}(n)} r_i(n)$, and the throughput achieved by policy $\pi(\cdot) : \mathcal{R} \to \mathcal{A}$, $V_n^{\pi}(\cdot)$, is expressed through the Bellman equations

$$V_n^{\pi}(\mathbf{r}(n)) = R_{\pi(\mathbf{r}(n))}(\mathbf{r}(n), \mathbf{r}(n+1)) + \beta V_{n+1}^{\pi}(\mathbf{r}(n+1)), \tag{6}$$

where $\mathbf{r}(n+1)$ is computed using (3), (4) and (5). The optimal policy π^* accomplishes the Bellman optimality equations

$$V_n^{\pi^*}(\mathbf{r}(n)) = \max_{\mathcal{K}(n) \in \mathcal{A}} \left\{ R_{\mathcal{K}(n)}(\mathbf{r}(n), \mathbf{r}(n+1)) + \beta V_{n+1}^{\pi^*}(\mathbf{r}(n+1)) \right\}. \tag{7}$$

We next define MP for this setting, and prove its optimality.

A. Definition of a Myopic Policy (MP)

MP is a greedy policy that schedules in each TS the K nodes with the largest expected throughput so as to maximize the immediate reward $R_{\mathcal{K}(n)}(\mathbf{r}(n), \mathbf{r}(n+1))$. Let $V_n^{MP}(\cdot)$ denote the throughput of MP. We have $\pi^{MP} = \{\mathcal{K}^{MP}(1), \dots, \mathcal{K}^{MP}(T)\}$, and $\mathcal{K}^{MP}(n) = \arg\max_{a_i \in \mathcal{A}} \sum_{i \in a_i} r_i(n)$.

Notice that in contrast to the optimal policy in (7) MP does not take into account the expected throughput of future states.

We will show that, in this setting, MP is equivalent to the round-robin policy (RRP). RRP, with throughput $V_n^{RR}(\cdot)$, schedules the nodes in the same order as they are indexed in the vector $\mathbf{r}$.

Proposition 2. *Given any initial belief $\mathbf{L}^0$ (i.e., any initial $\mathbf{r}(0)$ or $\bar{\mathbf{W}}(0)$) MP is equivalent to RRP which operates as follows: 1) Sort the vector $\mathbf{L}^0$ in a decreasing order, to obtain $\mathbf{L}'^0 = (l_1^0, \dots, l_N^0)$ such that $l_1^0 \leq l_2^0 \leq \dots \leq l_N^0$. Re-enumerate the nodes so that node i has belief $r_i(0) = g(l_i^0)$; 2) Divide the nodes into M sets of K nodes each, so that the first set contains the first K nodes, the second set contains nodes with index from $K+1$ to $2 \cdot K$, and so on so forth. Assume that action $a_1 \in \mathcal{A}$ corresponds to scheduling the first set, $a_2 \in \mathcal{A}$ corresponds to the second set and so on so forth. 3) Schedule the sets in a round-robin fashion so that the AP actions are $\{\mathcal{K}(1) = a_1, \mathcal{K}(2) = a_2, \dots, \mathcal{K}(M) = a_M, \mathcal{K}(M+1) = a_1, \dots\}$.*

Proof. In TS n MP schedules the K nodes with the highest expected throughput values (i.e., the first set in the RRP). In TS $n+1$ the expected throughput of the nodes scheduled in TS n is $r_i(n+1) = g(1) = 0$, and $r_i(n+1) \leq r_j(n+1), \forall i \in \mathcal{K}(n)$ and $\forall j \notin \mathcal{K}(n)$. Moreover, due to the monotonicity of $g(\cdot)$, the order in $\mathbf{r}$ of the nodes that are not scheduled in TS n is preserved in TS $n+1$. Hence, in TS $n+1$ the K nodes scheduled in TS n have the smallest throughput, and the nodes with higher expected throughput are scheduled (i.e., the second set in RRP), and so on so forth. ☐

B. Optimality of MP

We give sufficient conditions for the optimality of MP in Lemma 3. Then Theorem 4 shows that, scheduling the nodes with higher expected throughput can not decrease the total throughput. Finally in Theorem 5 we show that the conditions of optimality established by Lemma 3 hold $\forall n = \{1, \dots, T\}$, and thus, MP is optimal.

Lemma 3. *Assume that MP is optimal in TSs $n+1, \dots, T$, i.e., it satisfies (7). A sufficient condition for MP to be optimal in TS n is*

$$V_n^{RR}([\mathbf{r}_a, \mathbf{r}_{\overline{a}}]) \leq V_n^{MP}([\mathbf{r}_a, \mathbf{r}_{\overline{a}}]) = V_n^{RR}(\mathbf{r}'), \tag{8}$$

where $\mathbf{r}_a$ is a vector, not necessarily ordered, of expected throughputs of any K nodes, and $\mathbf{r}_{\overline{a}}$ the vector of throughputs of all nodes not in included in $\mathbf{r}_a$ in decreasing order, and $\mathbf{r}'$ is the vector of expected throughputs in decreasing order.

Proof. To prove that a policy is optimal we need to show that (7) holds $\forall n = \{1, \dots, T\}$. Since, by assumption MP is optimal from TS $n+1$ onwards, it is only necessary to prove that MP is optimal in TS n. We need to show that scheduling any set of nodes in TS n and following MP thereafter is no better than following MP directly, where the former and the latter are the left and right-hand sides of (8), respectively. Hence, if MP is optimal from TS $n+1$ onwards, (8) is a sufficient condition for its optimality in TS n. ☐

Theorem 4. *With beliefs* $\mathbf{r}_{a_i} = (r_1, \ldots, r_j, r_k, \ldots, r_N)$ *not necessarily ordered, if* $r_j \leq r_k$*, then*

$$V_n^{RR}(r_1, \ldots, r_j, r_k, \ldots, r_N) \leq V_n^{RR}(r_1, \ldots, r_k, r_j, \ldots, r_N) \quad (9)$$

Proof. The proof is in Appendix B. ☐

Theorem 5. *MP is optimal, that is* $\pi^{MP} = \pi^*$.

Proof. We will prove it using backward induction. We assume that MP is optimal in $TSs\ n+1, \ldots, T$, and prove its optimality at TS n. Since MP is optimal for TS T this will complete the proof. To prove the step induction we only need to see, from Lemma 3, that (8) holds. To show that, we use Theorem 4, which implies that if we swap two nodes that are scheduled in adjacent TSs, the throughput can be increased only if the node that is scheduled in the first TS has higher expected throughput than the node scheduled in the subsequent TS. By doing the necessary number of swaps (i.e., swapping left-wise the nodes with higher expected throughput) on $[\mathbf{r}_a, \mathbf{r}_{\overline{a}}]$ in the left-hand side of (8) we obtain $\mathbf{r}'$ in the right hand side of (8) through an array of inequalities, which guarantee that (8) holds for TS n. ☐

Theorem 6. *If the state of the EH process is not affected by the AP scheduling policy, but the AP does not observe the state of the EH process after scheduling a node, the MP is an RRP and is throughput optimal.*

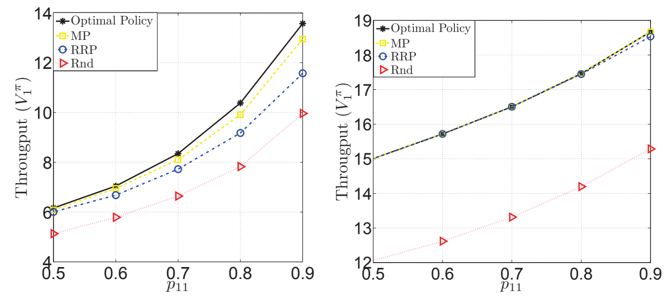
Proof. The proof follows similarly to the proof of Theorem 5. ☐

IV. Simultaneous Energy Harvesting and Data Transmission with Batteryless Nodes

In this problem setup we assume that the nodes can not store the harvested energy, and that it is lost if not used immediately. This is practically relevant for low-cost batteryless nodes with physical size constraints. Energy available for transmission in TS n, $B_{i,n}$, is equal to the energy harvested in TS $n-1$, $E^h_{i,n-1}$.

In this problem setup, the state space is the joint state of the EH processes, and is denoted by $S_n = (E^s_{1,n}, \ldots, E^s_{N,n})$. The state transition probabilities are independent of the scheduling policy $\mathcal{K}(n)$, and $P(S_n|S_{n+1}) = \prod_{i=1}^N p_{E^s_{i,n}E^s_{i,n+1}}$, where p_{ij} is the transition probability from state i to j of the EH process. The objective is to find the scheduling policy $\mathcal{K}(n)$ that maximizes the total expected throughput

$$\max_{\{\mathcal{K}(n)\}_{n=1}^T} \mathbb{E}\left[\sum_{n=1}^T \beta^{n-1} \sum_{i\in\mathcal{K}(n)} D_i(n)\right]. \quad (10)$$
$$\text{s.t. } B_{i,n} = E^h_{i,n-1}.$$



(a) $p_{00} = 0.9$ (b) $p_{00} = 0.5$

Figure 3. Throughput achieved for $p_{11} = \{0.5, \ldots, 0.9\}$.

The belief that the EH process associated with node i is in state 1 in TS n is denoted by $W_i(n)$, and is given by

$$W_i(n+1) = \begin{cases} W_i(n)p_{11} + (1-W_i(n))p_{01} & \text{with prob. } 1 \text{ if } i \notin \mathcal{K}(n), \\ p_{11} & \text{with prob. } W_i(n) \text{ if } i \in \mathcal{K}(n), \\ p_{01} & \text{with prob. } 1-W_i(n) \text{ if } i \in \mathcal{K}(n). \end{cases} \quad (11)$$

In [5], using similar arguments as in Section III, it is shown that, if the EH process transition probabilities satisfy $p_{11} \geq p_{01}$, the optimal policy that maximizes the throughput in (10) is MP. MP schedules in TS n the K nodes that are the most likely of being in state 1 (i.e., that has the highest believe $W_i(n)$). Given that $p_{11} \geq p_{01}$ and (11), MP schedules the same node until its EH state changes to state $E^s_{i,n} = 0$, then schedules the node with the highest belief, that is, the node that has not been scheduled for the highest number of TSs. In contrast with the results in Section III, MP in this setting is not RRP.

V. Numerical Results

For the general model described in Section II, with nodes' states as given in Figure 2, we compare the throughput of several scheduling policies, for $N = 3$, $T = 200$, $\beta = 0.9$, and $B = 2$. The optimal policy is obtained approximately through the policy iteration algorithm [7] and assuming that after not scheduling a node for more than ten TSs its state does not change. This assumption allows us to transform the infinite state space $\mathcal{W}$ in a finite space. We compare the optimal policy with RRP, which is optimal for the setting described in Section III, and with MP, which is optimal for the setting described in Sections III and IV. We also plot the throughput of a random scheduling policy (Rnd) that schedules the nodes in a random order. From Figure 3 we observe that if the EH state has low correlation across TSs (i.e., Figure 3(b)) the throughput obtained with RRP and MP are similar to the optimal one. On the contrary, if the EH state has high correlation across TSs (i.e., Figure 3(a)) the throughput of RRP and MP is lower compared to the optimal. This is due to the fact that, contrary to the optimal policy, MP and RRP do not consider the impact of their actions on the future rewards. But when the state transitions have low correlation it is difficult to reliably predict the impact of the actions on the future rewards.

VI. Conclusions

We have studied the scheduling problem in an EH multiple access system, in which the harvested energy at each node is modeled as a Markov process. The AP schedules a subset of the nodes at each TS with the goal of maximizing the total throughput without the knowledge of the states of the EH processes or the nodes' batteries. We have modeled the system as a POMDP, and showed the optimality of MP in two settings. In particular, when the nodes cannot harvest energy and transmit simultaneously and the EH process is affected by the scheduling policy, and when the nodes have no battery. We also have provided numerical results for the general setting.

Appendix A

We will first compute the belief $W_i^{1,j}(n+l_i)$ given that $W_i^{1,0}(n) = 1$. Then we will use this result to compute $W_i^{k,j}(n+l_i)$ for any k, j and initial state $\mathbf{W}_0$.

To harvest one unit of energy a transition from state $(1,j)$ to state $(1,j+1)$ must occur. There are two possible ways that this can happen: i) the state changes to $(0,j)$ with probability p_{10}, remains in this state for v TSs with probability ${p_{00}}^v$, and finally changes to $(1,j+1)$ with probability p_{01}, harvesting one unit of energy; and ii) the state changes directly to $(1,j+1)$ with probability p_{11}, and one unit of energy is harvested. Let N_A and N_B denote the number of times the events in i) and ii) occur between TS n and TS $n+l_i$, respectively. Notice that the amount of energy harvested is equal to the total number of i) and ii) events, hence $j = N_A + N_B$. Denote by m the total number of TSs that the EH process state remains in state 0 in the N_A events, we call this p_{00}-event, and $m = l_i - N_B - 2N_A$. The number of combinations in which the events i) and ii) can occur is $\binom{j}{N_B}$. The number of combinations that m p_{00}-events can occurs in N_A events is the weak composition of N_A natural numbers that sum up to m, that is $N_m = \binom{m+N_A-1}{N_A-1}$. The belief $W_i^{1,j}(n+l_i)$ given that $W_i^{1,0}(n) = 1$ is denoted by $h(j,l_i,B)$, and is given by

$$h(j,l_i,B)=\sum_{j'=\min\{j,B\}}^{H_{max}}\left\{\sum_{\substack{N_B=\\ \max\{0,2j'-l_i\}}}^{N_{B,max}} p_{11}^{N_B}(p_{10}p_{01})^{N_A}{p_{00}}^m N_m\binom{j}{N_B}\right\}, \quad (12)$$

where $H_{max} = \{l_i \text{ or } j\}$ if $j = B$ or $j < B$, respectively, and $N_{B,max} = \{j' \text{ or } j'-1\}$ if $j' = l_i$ or $l_i > j'$, respectively. Using similar arguments as in (12), the belief state $W_i^{k,j}(n+l_i)$, given the initial belief state $\mathbf{W}_0$, is

$$W_i^{k,j}(n+l_i) = f^{k,j}(l_i,\mathbf{W}_0) =$$
$$\begin{cases} p_1 h(j,l_i,B)+ \\ p_0\displaystyle\sum_{l=1}^{l_i-j+1}\Big(p_{01}{p_{00}}^{l-1}h(j-1,l_i-l,B-1)\Big) & \text{if } k=1, \\ p_1\displaystyle\sum_{l=1}^{l_i-j+1} p_{10}{p_{00}}^{l-1}h(j,l_i-l,B)+ \\ p_0\displaystyle\sum_{l=2}^{l_i-j+1} p_{10}p_{01}{p_{00}}^{l-2}h(j-1,l_i-l,B-1)(l-1) & \text{if } k=0, \end{cases} \quad (13)$$

where the super-script in $f^{k,j}$ refers to the components (k,j).

Appendix B

If r_j and r_k belong to the same set of nodes, the inequality holds with equality because both nodes are scheduled in the same TSs. To prove the case in which r_j and r_k belong to different sets we first need to compute the throughput of RRP policy, and then show that inequality (9) holds. The contribution of each node is independent of the contribution of others nodes, so the total throughput can be computed from the individual throughputs of the nodes. We consider the throughput of node j which belongs to the set a_m, with $m \leq M$, that is, it is scheduled at TSs $m, M+m, 2M+m, \ldots$. The throughput in TS m is $r_j(m) = \beta^m g(l_j^n)$, and the throughput achieved in TS k, for $k \in \{M+m, 2M+m, \ldots\}$, is $r_j(k) = \beta^{m+k}g(M)$. Hence, if node j is scheduled T_j times the throughput is

$$\begin{aligned} S_D(T_j,l_j^0) &= \sum_{e=1}^{T_j}\beta^{m+(e-1)M}r_j(m+(e-1)M) \\ &= \beta^m g(l_j^m) + \beta^m g(M)\tfrac{1-(\beta^M)^{T_j-2}}{1-(\beta^M)}. \end{aligned} \quad (14)$$

Now we can compare both sides of (9). First we have to notice that all nodes but nodes j and k are scheduled in the same TSs in both sides of (9). Hence, we only have to focus on the throughput of nodes j and k since the other nodes contribute equally. Notice that, nodes j and k are scheduled in TS m and in TS $m+1$, respectively; in the left-hand side of (9). In the right-hand side of (9) nodes j and k are scheduled for the first time in TS $m+1$ and TS m, respectively. We denote the throughput achieved by node i in the left-hand side and the right-hand side of (9) by $S_D^{i,L}$ and $S_D^{i,R}$, respectively; for $i = \{j,k\}$. Considering only the contributions of nodes j and k, (9) is simplified to $S_D^{j,L} + S_D^{k,L} \leq S_D^{j,R} + S_D^{k,R}$, and using (14) , we obtain

$$\begin{aligned} &\beta^m g(l_j^m) + \beta^m g(M)\frac{1-(\beta^M)^{T_j^L-2}}{1-(\beta^M)} + \beta^{m+1}g(l_k^{m+1}) \\ &+\beta^{m+1}g(M)\frac{1-(\beta^M)^{T_k^L-2}}{1-(\beta^M)} \leq \beta^m g(l_k^m) + \beta^m g(M)\frac{1-(\beta^M)^{T_k^R-2}}{1-(\beta^M)} \\ &+\beta^{m+1}g(l_j^{m+1}) + \beta^{m+1}g(M)\frac{1-(\beta^M)^{T_j^R-2}}{1-(\beta^M)}, \end{aligned} \quad (15)$$

where T_i^L and T_i^R are the number of times node i is scheduled in the left and right-hand sides of (9), respectively; for $i = \{j,k\}$. It is easy to se that $T_j^L = T_k^R$ and that $T_k^L = T_j^R$. Then

$$g(l_j^m) + \beta g(l_k^{m+1}) \leq g(l_k^m) + \beta g(l_j^{m+1}) \quad (16)$$

which holds due to the properties of the function $g(\cdot)$.

References

[1] Z. Wang, A. Tajer, and X. Wang, "Communication of energy harvesting tags," *IEEE Trans. Commun.*, vol. 60, no. 4, pp. 1159–1166, Apr. 2012.

[2] H. Li, N. Jaggi, and B. Sikdar, "Relay scheduling for cooperative communications in sensor networks with energy harvesting," *IEEE Trans. Wireless Commun.*, vol. 10, no. 9, pp. 2918–2928, Sep. 2011.

[3] J. Lei, R. Yates, and L. Greenstein, "A generic model for optimizing single-hop transmission policy of replenishable sensors," *IEEE Trans. Wireless Commun.*, vol. 8, no. 4, pp. 547–551, Apr. 2009.

[4] F. Iannello, O. Simeone, and U. Spagnolini, "Optimality of myopic scheduling and Whittle indexability for energy harvesting sensors," in *46th Annual Conference on Information Sciences and Systems (CISS)*, Princeton, NJ, USA, Mar. 2012, pp. 1–6.

[5] A. H. S. Ahmad and M. Liu, "Multi-channel opportunistic access: A case of restless bandits with multiple plays," in *Communication, Control, and Computing. Allerton 2009. 47th Annual Allerton*, Monticello, IL, Sep. 2009, pp. 1361–1368.

[6] K. P. Murphy, "A survey of POMDP solution techniques," U.C. Berkeley, Tech. Rep., 2000.

[7] D. P. Bertsekas, *Dynamic Programming and Optimal Control*. Athenea Scientific, 2005.