Cooperative Relay Scheduling under Partial State Information in Energy Harvesting Sensor Networks

Huijiang Li†, Neeraj Jaggi‡ and Biplab Sikdar†
†Department of ECSE, Rensselaer Polytechnic Institute, Troy, NY 12180
‡ Department of EECS, Wichita State University, Wichita, KS 67260

***Abstract*— Sensors equipped with energy harvesting and cooperative communication capabilities are a viable solution to the power limitations of Wireless Sensor Networks (WSNs) associated with current battery technology. However, the optimal scheduling of transmissions in such networks is challenging due to the requirement of complete state information of the relay nodes. This paper addresses the problem of transmission scheduling in such networks when only *partial* state information about the relays is available at the source. We formulate the scheduling problem as a Partially Observable Markov Decision Process (POMDP), and show that it can be decomposed into an equivalent Markov Decision Process (MDP) problem. Simulation results are used to show the performance of the scheduler.**

I. Introduction

Although current battery technology is incapable of facilitating sensor networks with a sufficiently long life for many applications, energy harvesting or energy scavenging has become a promising and feasible approach to address the energy supply problem [3], [2]. However, to improve the performance of energy harvesting WSNs to acceptable levels, harvesting-aware communication policies and protocols need to be developed. This paper addresses the problem of scheduling transmissions in sensor networks with energy harvesting capability, where nodes may use cooperative communications strategies. The cooperative communication considered in this paper is the simple discrete memoryless three-terminal relay network developed in the landmark papers by van der Meulen [5], and Cover and El Gamal [6].

We consider a time-slotted source-relay-destination system, where a sensor (the source) has the option to have another sensor (the relay) help to transmit its data to the destination. All sensor nodes under consideration are equipped with energy harvesting capability. From an energy efficiency perspective, the source may achieve the same bit error rate (BER) for a lower transmission power if it uses a relay, as compared to a direct transmission. However, this increases the power consumed by the relay and as a result, the relay sensor may not have energy to report its own data in the future. At any given instant of time, the problem of interest is to determine how the source decides whether to transmit data on its own or cooperatively with the relay in order to maximize the long term ratio of the data that is successfully delivered, to the total data that is generated.

In order to optimally determine if the relay should be used or not at a given time, in addition to its own state information (e.g. current battery level), the source also needs to know the state information at the relay. It is reasonable to assume that when a relay transmits or relays data, the headers of the packets include the state information. However, in periods without data, conveying the state information of the relay in real time represents a significant overhead. Thus the source may have to base its decision on stale state information. The focus of this paper is to determine the optimal decision based on *partial information* about the system.

The rest of the paper is organized as follows. Section II describes the system model. A POMDP formulation of the transmission scheduling problem is presented in Section III. Finally, simulation results are presented in Section IV and Section V concludes the paper.

II. System Model

We consider a WSN where each sensor can be categorized as either a source or a relay sensor. Every source sensor has a designated relay counterpart. In this paper, we assume that decode-and-forward relaying is adopted [4]. A source sensor has two transmission modes: the *direct mode* in which the sensor transmits the packet directly to the destination and consumes δ_1^s units of energy and the *relay mode* (which consumes δ_2^s units of energy) in which the packet is transmitted by the source and relayed by the relay sensor. A relay sensor also has two transmission modes: *own-traffic mode* and *relay mode*. In own-traffic mode, the relay sensor transmits its own packet to the destination consuming δ_1^r units of energy while in relay mode the relay sensor's own traffic is discarded and δ_2^r units of energy is consumed to relay another sensor node's packet. We have $\delta_1 > \delta_2$ where we have dropped the superscript to indicate that the relation holds for both source and relay sensors. We assume that the sensors are working in real-time monitoring scenarios. Thus no retransmissions are attempted for packets with errors. Also, if a packet is not transmitted in the slot in which it arrives, it is dropped to allow more recent data to be transmitted. Finally, a sensor is considered available for operation if it has enough energy to transmit or relay a packet.

A discrete time model is assumed where time is slotted in intervals of unit length. Each slot is long enough so that a source node and a relay node can either cooperatively transmit one data packet for the source, or both can transmit one of their own packets. At most one data packet is generated at a node in a slot. Each sensor has a rechargeable battery and an energy harvesting device. The energy generation process at each sensor is modeled by a correlated, two-state process with parameters (q_{on}, q_{off}). In the *on* state (i.e. when ambient conditions are conducive to energy harvesting), the sensor generates energy at a constant rate of c units in a time slot. In the *off* state, no energy is generated. If the sensor harvested energy in the current slot, it harvests energy in the next slot with probability q_{on} and no energy is harvested with probability $1 - q_{on}$. On the other hand, if no energy was harvested in the current slot, no energy is harvested in the next slot with probability q_{off}, and energy is harvested with probability $1-q_{off}$, and we assume $0.5 < q_{on}, q_{off} < 1$. We assume that the energy generated during a recharge event is available at the end of the slot.

The data packets at the sensors are also generated according to a correlated, two-state process with parameters (p_{on}, p_{off}) with $0.5 < p_{on}, p_{off} < 1$, where in the *on* state an event (i.e. data packet) is generated in each slot, and no events are generated in the *off* state. We assume that an event is generated and detected at the beginning of a time slot. The average duration of a period of continuous events, $E[N]$, is

$$E[N] = \sum_{i=1}^{\infty} i(p_{on})^{i-1}(1-p_{on}) = \frac{1}{1-p_{on}} \tag{1}$$

and the steady-state probability of event occurrence, π_{on}, is

$$\pi_{on} = \frac{1-p_{off}}{2-p_{on}-p_{off}} \tag{2}$$

Similarly, the average length of a period without events and its steady-state probability are $\frac{1}{1-p_{off}}$ and $\pi_{off} = 1 - \pi_{on}$, respectively. Also, the average length of a period with energy harvesting and the steady-state probability of such events are $\frac{1}{1-q_{on}}$ and $\mu_{on} = \frac{1-q_{off}}{2-q_{on}-q_{off}}$, respectively. Finally, the expected length of periods without recharging and its steady-state probability are $\frac{1}{1-q_{off}}$ and $\mu_{off} = 1-\mu_{on}$, respectively. Parameters corresponding to the source and relay nodes are denoted with a superscript of s and r, respectively (e.g. p_{on}^s).

The communication strategy of a sensor pair {source, designated relay} is governed by a policy Π that decides on the transmission mode to be used for reporting events. The action taken by the sensor pair in time slot t is denoted by a_t with $a_t \in \{0, 1, 2, 3, 4\}$ denoting {no transmission, no transmission}, {direct, no transmission}, {relay, relay}, {direct, own-traffic}, and {no transmission, own-traffic}. A transmission action can be taken only if the corresponding sensor has enough power (δ_1 for direct mode and δ_2 for relay mode) and an event occurs at the beginning of the slot. A node is said to be *active* in a time slot if the action is taken such that it has a packet transmission (either its own traffic or relaying), and *inactive* if there is no transmission.

We assume that when a sensor transmits a packet, its current state information is included in the packet's header. Then, a source sensor can obtain state information about the relay including: (a) the relay's current energy level, (b) whether there is an event generated at the relay in current slot, and (c) whether the relay's battery is currently recharging or not. However, if the relay is inactive in a slot, the source will not have the updated state information of the relay.

We assume that the communication policy is decided at the source sensor. The decision may be based on: (a) the current battery level, the states of the energy and the event generation processes at the source; (b) the partial information of the battery, the energy and event generation processes at the relay which was obtained when the relay was last active. We refer to such system as a *partially observable* system. The objective of the decision policy Π is to maximize the *packet delivery ratio*, defined as the long term ratio of the total number of events reported, to the total number of events generated.

III. Partially Observable Markov Decision Process Formulation

For the partially observable system, we first formulate the decision problem as a POMDP, and then present the equivalent MDP formulation.

A. System States and Observations

Denote the system state at time t by $X_t = (L_t^s, E_t^s, Y_t^s, L_t^r, E_t^r, Y_t^r)$ where $L_t^s, L_t^r \in \{0, 1, 2, \cdots, K\}$ represents the energy available at the sensors at time t. $Y_t^s \in \{0, 1\}$ equals one if the source is being charged during time interval $[t, t+1)$ and zero otherwise. Also, $E_t^s \in \{0, 1\}$ equals one if an event to be reported during time interval $[t, t+1)$ is generated at time t at the source and zero otherwise. The variables $Y_t^r \in \{0, 1\}$ and $E_t^r \in \{0, 1\}$ are defined similarly for the relay, but equal one if the recharge and event processes, respectively, are *on* during time interval $[t-1, t)$. The state of the relay at time t is defined in terms of the previous slot since that is the latest information the source may have about the relay. We assume that the battery at a sensor has a finite capacity K. Then the state space $\mathcal{X}$ is given by, $\mathcal{X} = \{(0,0,0,0,0,0), (1,0,0,0,0,0), \cdots, (K,1,1,K,1,1)\}$ with $|\mathcal{X}| = 16(K+1)^2$. In subsequent discussions, we also refer to $X_t^s = (L_t^s, E_t^s, Y_t^s)$ as the source sensor's state and $X_t^r = (L_t^r, E_t^r, Y_t^r)$ as the relay's state at time t. Denote the set of actions described in Section II as $\mathcal{A} = \{0, 1, 2, 3, 4\}$. The action taken at time t is denoted by $a_t \in \mathcal{A}$.

The *system observation* at time t at the source sensor is denoted by Y_t. The source is assumed to always have full information about itself. If the action taken at time $t-1$ is $2, 3$ or 4, then the relay was active, and the observation matches the state and equals X_t. However, if the action taken was 0 or 1, the relay was inactive and its current state information is unknown. Thus the observation Y_t is characterized by,

$$Y_t = \begin{cases} X_t & \text{if } a_{t-1} \in \{2,3,4\} \\ (L_t^s, E_t^s, Y_t^s, \phi_L, \phi_E, \phi_Y) & \text{if } a_{t-1} \in \{0,1\} \end{cases}$$

where ϕ_ω denotes that a variable ω is unknown. The observation space $\mathcal{Y}$ is given by,

$$\mathcal{Y} = \mathcal{X} \cup \{(0,0,0,\phi_L,\phi_E,\phi_Y), (1,0,0,\phi_L,\phi_E,\phi_Y), \\ \cdots, (K,1,1,\phi_L,\phi_E,\phi_Y)\}, \tag{3}$$

with $|\mathcal{Y}| = 4(K+1)(4K+5)$. Let $q_{x,y}(a)$ be the probability distribution of the observation ($Y_t = y$) at time t, conditioned on the current state ($X_t = x$) and the action taken at time $t-1$ ($a_{t-1} = a$). Thus, $q_{x,y}(a) = Pr[Y_t = y | X_t = x, a_{t-1} = a]$. If relay was active in time interval $[t-1, t)$, the source has perfect information at time t. Then we have

$$q_{x,y}(0) = q_{x,y}(1) = \begin{cases} 1 & y = (x.L^s, x.E^s, x.Y^s, \phi_L, \phi_E, \phi_Y) \\ 0 & \text{otherwise} \end{cases}$$

$$q_{x,y}(2) = q_{x,y}(3) = q_{x,y}(4) = \begin{cases} 1 & y = x \\ 0 & \text{otherwise} \end{cases} \quad (4)$$

where $x.L^s = L_t^s$, $x.E^s = E_t^s$ and $x.Y^s = Y_t^s$ when $x = (L_t^s, E_t^s, Y_t^s, L_t^r, E_t^r, Y_t^r)$.

B. POMDP Formulation

In the presence of only partial observations, the optimal action depends on the current and past observations, and on past actions. Existing work has shown that a POMDP may be formulated as a completely observable MDP with the same finite action set [1], [7], [8]. The state space for the equivalent MDP comprises of the space of probability distributions on the original state space. Thus in the general case, the state space of the equivalent MDP may become uncountable or infinite. In our case, the structure of the POMDP leads to a countable state space for the equivalent MDP, guaranteeing the existence of an optimal solution to the average cost (reward) optimality equation [7]. As a result, the solution to the equivalent MDP with complete state information provides the optimal actions to take in the POMDP, and with the optimal reward.

Denote the state space of the equivalent MDP as Δ, and its state at time t as Z_t. Then $Z_t \in \Delta$ is a information vector of length $|\mathcal{X}|$, whose i-th component is given by,

$$Z_t^{(i)} = Pr[X_t = i | y_t, \cdots, y_1; a_{t-1}, \cdots, a_0], \quad i \in \mathcal{X}. \quad (5)$$

We have $I' Z_t = 1$, where I' denotes a row vector of length $|\mathcal{X}|$ with all elements equal to 1, since the elements of Z_t are mutually exclusive whose union is the universal set. The state Z_{t+1} is recursively computable given the state transition probability matrices $P(a)$, action taken a_t, and the observation y_{t+1} and is given by [8],

$$Z_{t+1} = \sum_{y \in \mathcal{Y}} \frac{\bar{Q}_y(a_t) P'(a_t) Z_t}{I' \bar{Q}_y(a_t) P'(a_t) Z_t} I[Y_{t+1} = y], \quad (6)$$

where $I[B]$ denotes the indicator function of the event B and the matrices $\bar{Q}_y(a) = \text{diag}\{q_{x,y}(a)\}$, with $q_{x,y}(a)$ as defined in Eqn. (4). $P(a)$ is the state transition probability matrix with action a, and for $i, j \in \mathcal{X}$, is defined as

$$\begin{aligned} [P(a)]_{ij} &= Pr[X_{t+1} = j | X_t = i, X_{t-1}, \cdots, X_0, \\ &\qquad a_t = a, a_{t-1}, \cdots, a_0] \\ &= Pr[X_{t+1} = j | X_t = i, a_t = a] \triangleq P_{ij}(a). \end{aligned} \quad (7)$$

We use $\bar{T}(y, Z_t, a_t) = \bar{Q}_y(a_t) P'(a_t) Z_t$ to denote the numerator and $V(y, Z_t, a_t) = I' \bar{Q}_y(a_t) P'(a_t) Z_t$ to denote the denominator in Eqn. (6). Denote,

$$W(y, Z_t, a_t) = \frac{\bar{T}(y, Z_t, a_t)}{V(y, Z_t, a_t)}. \quad (8)$$

Then Eqn. (6) can be written as,

$$Z_{t+1} = \sum_{y \in \mathcal{Y}} W(y, Z_t, a_t) I[Y_{t+1} = y]. \quad (9)$$

Thus $\{Z_t\}$ forms a completely observable controlled Markov process with state space Δ.

In Section III-C, we will show that the state space of the equivalent MDP is countable and that the state at time t, Z_t, can be represented in the form $Z_t = (L_t^s, E_t^s, Y_t^s, L^r, E^r, Y^r, i)$, representing the following: (a) the relay had no transmissions in the past i slots; (b) the state of the relay when it last transmitted was (L^r, E^r, Y^r); (c) the current state at the source is (L_t^s, E_t^s, Y_t^s).

The POMDP is then transformed to an equivalent MDP with state space Δ and the optimality equations for this MDP are given by [8]:

$$\Gamma^* + h^*(Z) = \max_{a \in \mathcal{A}} \left[\bar{R}(Z, a) + \sum_{y \in \mathcal{Y}} V(y, Z, a) h^*(W(y, Z, u)) \right], \quad \forall Z \in \Delta. \quad (10)$$

where $h^*(Z)$ is the optimal reward when starting at state Z and $\bar{R}(Z, a)$ is the reward function which will be discussed in Section III-D.

C. Formulation of the State Space

Given $X_t^s = (L_t^s, E_t^s, Y_t^s)$ as the state of the source at time t and observation $y_t = (L_t^s, E_t^s, Y_t^s, \phi_L, \phi_E, \phi_Y)$, the observation vector has $4(K+1)$ possibilities with different combinations of $0 \le \phi_L \le K$, $\phi_E \in \{0, 1\}$ and $\phi_Y \in \{0, 1\}$. We number these states as: state $1 = (L_t^s, E_t^s, Y_t^s, 0, 0, 0)$, state $2 = (L_t^s, E_t^s, Y_t^s, 0, 0, 1)$, $\cdots$, and state $4(K+1) = (L_t^s, E_t^s, Y_t^s, K, 1, 1)$.

Let e^j denote the unit column vector with all zeros except the j^{th} element being one. Then $Z_t = e^{y_t} = e^{x_t}$ for $a_{t-1} \in \{2, 3, 4\}$. However, if $a_{t-1} \in \{0, 1\}$, the state Z_t of the equivalent MDP has maximum of $4(K+1)$ non-zero components, and can be represented by,

$$Z_t = \alpha_1 e^1 + \alpha_2 e^2 + \cdots + \alpha_{4(K+1)} e^{4(K+1)} \quad (11)$$

where $\sum_{i=1}^{4(K+1)} \alpha_i = 1$. To obtain the values of $\alpha_i, 1 \le i \le 4(K+1)$, we first evaluate the transition probabilities of the event process and then those for the energy generation process and the battery level.

Let $F_{1,0}^{(i)}$ be the probability that the event process at a sensor (source or relay) at time $t+i$ is *off*, given that it was *on* at time t. Similarly, $F_{0,1}^{(i)}$ denotes the i-step transition probabilities of the event generation process at the node from *off* to *on* state in i time slots. Then the $(i+1)$-step transition probabilities are recursively given by,

$$\begin{aligned} F_{1,0}^{(i+1)} &= p_{off}F_{1,0}^{(i)} + (1-p_{on})(1-F_{1,0}^{(i)}) \\ F_{0,1}^{(i+1)} &= p_{on}F_{0,1}^{(i)} + (1-p_{off})(1-F_{0,1}^{(i)}). \end{aligned}$$

Equivalently,

$$F_{E,1-E}^{(i+1)} = p_{1-E}F_{E,1-E}^{(i)} + (1-p_E)(1-F_{E,1-E}^{(i)}) \quad (12)$$

where $E \in \{0,1\}$, $p_0 = p_{off}$ and $p_1 = p_{on}$. We also have $F_{E,E} = 1 - F_{E,1-E}^{(i)}$, $F_{1,0}^{(1)} = 1 - p_{on}$ and $F_{0,1}^{(1)} = 1 - p_{off}$. Since $0 < p_{off} + p_{on} - 1 < 1$, it can be shown that,

$$F_{E,1-E}^{(i)} = \frac{(1-p_E)[1-(p_E+p_{1-E}-1)^i]}{2-p_E-p_{1-E}}, \quad (13)$$

and $\lim_{i\to\infty} F_{0,1}^{(i)} = \pi_{on}$ and $\lim_{i\to\infty} F_{1,0}^{(i)} = \pi_{off}$.

The state of the battery level at the relay is related to the recharge process, given the last known battery level. In the slots where the relay does not transmit, the battery level increases whenever the recharge process is *on*. The i-step transition probabilities of the recharge process at the relay, $G_{E,1-E}^{(i)}$, are then also given by Eqn. (13) with p_E replaced by q_E. To obtain the battery level at time $t+i$, given the battery level and recharging state at time t, we need to evaluate the number of slots with recharge events during the i-step time interval. This problem is solved recursively as follows.

Consider an interval of u slots. Let $R(u,v,0)$ and $R(u,v,1)$ denote the probabilities that in v out of u slots, the recharge process at a sensor was in the *on* state and the state in the u-th (the final) slot is *off* ($Y^r = 0$) and *on* ($Y^r = 1$), respectively. These probabilities can be recursively written as [9],

$$\begin{aligned} R(u,v,0) &= (1-q_{on})R(u-1,v,1) + q_{off}R(u-1,v,0) \\ R(u,v,1) &= q_{on}R(u-1,v-1,1) + (1-q_{off})R(u-1,v-1,0), \end{aligned}$$

while satisfying the following initial conditions:

$$\begin{aligned} R(u,0,1) &= 0 \\ R(u,u,0) &= 0 \\ R(u,u,1) &= (q_{on})^u R_1(0) + (1-q_{off})(q_{on})^{u-1}R_0(0) \\ R(u,0,0) &= (q_{off})^u R_0(0) + (1-q_{on})(q_{off})^{u-1}R_1(0). \end{aligned}$$

with $u, v \in \{1, 2, \cdots\}$, $u \geq v$ and $R_1(0) = 1$ if the energy harvesting process was in the *on* state at the beginning of the first slot of the u-slot interval, 0 otherwise. Also, $R_0(0) = 1 - R_1(0)$.

Given that a sensor did not transmit for i slots, let $P_{j,k}^{(i)}$ denote the probability of transition of the relay from state j to state k in i slots. Let the states be $j = X_t^r = (L, E, Y)$ and $k = X_{t+i}^r = (L', E', Y')$. Then, $P_{j,k}^{(i)}$ is given by,

$$P_{j,k}^{(i)} = \begin{cases} F_{E,E'}^{(i)} R(i,v,Y') & \text{if } L' = L + vc < K \\ \sum_{v=\lceil \frac{K-L}{c} \rceil}^{i} F_{E,E'}^{(i)} R(i,v,Y') & \text{if } L' = K \\ 0 & \text{otherwise.} \end{cases}$$

for all $v \in [0, 1, \cdots, i]$. Since the state of the source is independent of that of the relay, given that the relay was in state j the time last it transmitted and that the relay has been inactive for the last i slots, we can rewrite Eqn. (11) as,

$$Z_t = P_{j,1}^{(i)}e^1 + P_{j,2}^{(i)}e^2 + \cdots + P_{j,4(K+1)}^{(i)}e^{4(K+1)} \quad (14)$$

Now we can represent Z_t as,

$$Z_t = (L_t^s, E_t^s, Y_t^s, L^r, E^r, Y^r, i) \quad (15)$$

where $X_t^s = (L_t^s, E_t^s, Y_t^s)$ is the source's state at time t and $X^r = (L^r, E^r, Y^r)$ is the relay's state when in last transmitted, which was i slots ago. We then have the following result.

Lemma 1: The state-space Δ is countable.

Proof: Let $Z_t = (L_t^s, E_t^s, Y_t^s, L^r, E^r, Y^r, i)$ for some $(L_t^s, E_t^s, Y_t^s, L^r, E^r, Y^r) \in \mathcal{X}$ and integer $i \geq 0$. Let $X_{t+1} = (L_{t+1}^s, E_{t+1}^s, Y_{t+1}^s, L_{t+1}^r, E_{t+1}^r, Y_{t+1}^r)$.

Case(i) : $a_t = \{0,1\}$. Since the relay is not observable, $y_{t+1} = (L_{t+1}^s, E_{t+1}^s, Y_{t+1}^s, \phi_L, \phi_E, \phi_Y)$. Then, Z_{t+1} in the form of Eqn. (14) can be expanded as,

$$\begin{aligned} Z_{t+1} &= \sum_{k=1}^{4(K+1)} P_{j,k}^{(i)}P_{k,1}^{(1)}e^1 + \cdots + \sum_{k=1}^{4(K+1)} P_{j,k}^{(i)}P_{k,4(K+1)}^{(1)}e^{4(K+1)} \\ &= (L_{t+1}^s, E_{t+1}^s, Y_{t+1}^s, L^r, E^r, Y^r, i+1). \end{aligned} \quad (16)$$

Case(ii) : $a_t = 2$. In this case the transmission of a source's packet relies on the relay and the action can be taken if the relay has enough power (by assumption, the relay discards its own traffic, if any). Thus,

$$Z_{t+1} = \begin{cases} (L_{t+1}^s, E_{t+1}^s, Y_{t+1}^s, L_{t+1}^r, E_{t+1}^r, Y_{t+1}^r, 0) \\ \qquad \text{w.p. } Pr[L_t^r \geq \delta_2^r | L_{t-i}^r = L^r]. \\ (L_{t+1}^s, E_{t+1}^s, Y_{t+1}^s, L^r, E^r, Y^r, i+1) \\ \qquad \text{w.p. } 1 - Pr[L_t^r \geq \delta_2^r | L_{t-i}^r = L^r]. \end{cases} \quad (17)$$

Case(iii) : $a_t = \{3,4\}$. In these two cases, the relay will transmit its own traffic if it has enough power and has an event to report. Thus,

$$Z_{t+1} = \begin{cases} (L_{t+1}^s, E_{t+1}^s, Y_{t+1}^s, L_{t+1}^r, E_{t+1}^r, Y_{t+1}^r, 0) \\ \quad \text{w.p. } Pr[L_t^r \geq \delta_2^r, E_t^r = 1 | L_{t-i}^r = L^r, E_{t-i}^r = E^r] \\ (L_{t+1}^s, E_{t+1}^s, Y_{t+1}^s, L^r, E^r, Y^r, i+1) \\ \quad \text{w.p. } 1 - Pr[L_t^r \geq \delta_2^r, E_t^r = 1 | L_{t-i}^r = L^r, E_{t-i}^r = E^r] \end{cases}$$

To sum up, Z_{t+1} is completely described by Z_t, a_t and y_{t+1}. Since L_t^s, L_t^r, E_t^s, E_t^r, Y_t^s, Y_t^r and i are individually finite or countable, and all $Z \in \Delta$ has the form of Eqn. (15), we have the result. ■

D. Equivalent MDP Reward Function

Let θ^s and θ^r denote the rewards gained by the system for each source sensor and relay sensor event that is successfully reported, respectively. For the partially observable system, the reward associated with the states $Z \in \Delta$ of the equivalent MDP, denoted as $\bar{R}(Z,a)$, is the same as that of the optimal reward for the original POMDP [7]. Then, the reward function of the equivalent MDP at time t is given by,

$$\bar{R}(Z,a) = \begin{cases} \theta^s & \text{if } a_t=1, E_t^s=1, L_t^s \geq \delta_1^s \\ \theta^s Pr[L_t^r \geq \delta_2^r | L_{t-i}^r = L^r] & \text{if } a_t=2, E_t^s=1, L_t^s \geq \delta_2^s \\ \theta^s + \theta^r Pr[L_t^r \geq \delta_2^r, E_t^r=1 | & \text{if } a_t=3, E_t^s=1, L_t^s \geq \delta_1^s \\ \quad L_{t-i}^r = L^r, E_{t-i}^r = E^r] & \\ \theta^r Pr[L_t^r \geq \delta_2^r | L_{t-i}^r = L^r] \cdot & \text{if } a_t=4 \\ \quad Pr[E_t^r = 1 | E_{t-i}^r = E^r] & \\ 0 & \text{otherwise} \end{cases}$$

IV. Simulation Results

This section explores the impact of various parameters on the performance of the proposed scheduler using simulations. We consider a network where a three-node-group model (i.e. source, relay and destination) is applied and each group is independent of others. The simulations are based on a single group and the results can be generalized. We generated the optimal policy using value iteration, and simulated its performance using C language. All simulations were run for a duration of 5000000 time units. All figures show the packet delivery ratio as defined in Section II.

Figure 2 demonstrates the effect of the event generation process on the performance and all parameters are specified in the caption. Since $0.5 \leq p_{on}, p_{off} \leq 1$, the four choices of $(0.6, 0.6)$, $(0.6, 0.9)$, $(0.9, 0.6)$ and $(0.9, 0.9)$ give an indication of the performance in diverse settings of low-low, low-high, high-low and high-high correlation probabilities at the relay. In all of the four cases, the packet delivery ratio decreases as p_{on}^s increases and as p_{off}^r decreases. Overall, the packet delivery ratio is higher when the relay has a lower p_{on}^r and higher p_{off}^r. The percentage of relay usage in four cases are shown in Table I. The table shows the overall packet delivery ratio (PDR) defined earlier along with the individual PDRs defined as the following: *Source PDR* is the ratio of the number of packets transmitted to the number of packets generated by the source; *Relay PDR* is the ratio of the number of packets transmitted (its own) to the number of packets generated by the relay. *Relay usage* is defined as the ratio of the number of source packets transmitted using the relay to the total number of packets transmitted by the source. For cases $(0.6, 0.6)$ and $(0.9, 0.9)$, the steady-state probabilities of the event occurrence at the relay (π_{on}^r) are the same ($1/2$), but as the length of continuous events at the relay ($E(N)$, defined in Eqn. (1)) increases, the source tends to transmit the traffic directly as long as it has enough energy. When both $E(N)$ and π_{on}^r are low ($(0.6, 0.9)$ case), the relay is used intensively by the source.

V. Conclusions

This paper addressed the problem of developing transmission strategies for WSNs when energy harvesting devices are used by sensors to generate energy. We consider the case where a node may use either a direct transmission or a cooperative relay for its transmission, under the limitation that the state of the relay is not fully observable. The problem is formulated as POMDP and a scheduling policy is then developed by converting the problem to an equivalent, fully observable MDP.

TABLE I

Relay usage summary. (Parameters used: $q_{on}^s = q_{on}^r = 0.85$, $q_{off}^s = q_{off}^r = 0.7$, $p_{on}^s = 0.85$, $p_{off}^s = 0.7$, $c^s = c^r = 1$, $\delta_1^s = \delta_1^r = 2$, $\delta_2^s = \delta_2^r = 1$)

(p_{on}^r, p_{off}^r)	(0.6,0.6)	(0.6,0.9)	(0.9,0.6)	(0.9,0.9)
PDR	0.5662	0.7309	0.4510	0.5179
Source PDR	0.5945	0.7765	0.4934	0.4932
Relay usage	0.3372	0.7477	0.0000	0.0000
Relay PDR	0.5284	0.5793	0.4157	0.5507

References

[1] D. Bertsekas, *Dynamic Programming and Optimal Control,* Athena Scientific, Belmon MA, 2000.

[2] J. Paradiso and T. Starner, “Energy scavenging for mobile and wireless electronics,” *IEEE Perv. Comp.,* vol. 4, no. 1, pp. 18-27, Jan-Mar 2005.

[3] I. Akyildiz, W. Su, Y. Sankarasubramaniam and E. Cayirci, “Wireless Sensor Networks and Applications: a Survey,” *Int. J. Comp. Sci. and Network Security,* vol. 7, no. 3, pp. 264-273, March 2007.

[4] G. Kramer, M. Gastpar, and P. Gupta, “Cooperative Strategies and Capacity Theorems for Relay Networks,” *IEEE Transactions on Information Theory,* Vol. 51, Issue 9, pp. 3037- 3063, August 2005.

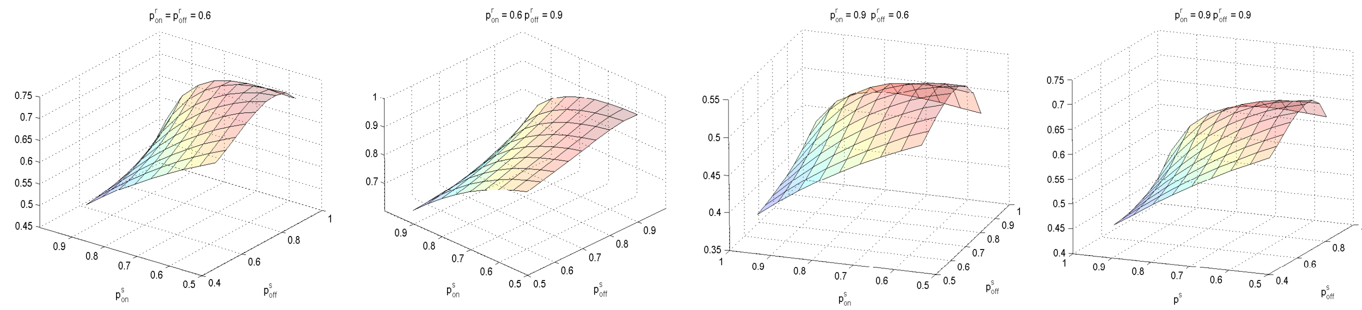
[5] E. C. van der Meulen, “Three-terminal communication channels,” *Adv. Appl. Prob.,* vol. 3, pp. 120-154, 1971.

[6] T. Cover and A. El Gamal, “Capacity theorems for the relay channel,” *IEEE Trans. Inf. Theory,* vol. 25, no. 5, pp. 572-584, September 1979.

[7] A. Cassandra, L. Kaelbling and M. Littman, “Acting optimally in partially observable stochastic domains,” *Proc. National Conference on Artificial Intelligence,* pp. 1023-1028, Seattle, WA, 1994.

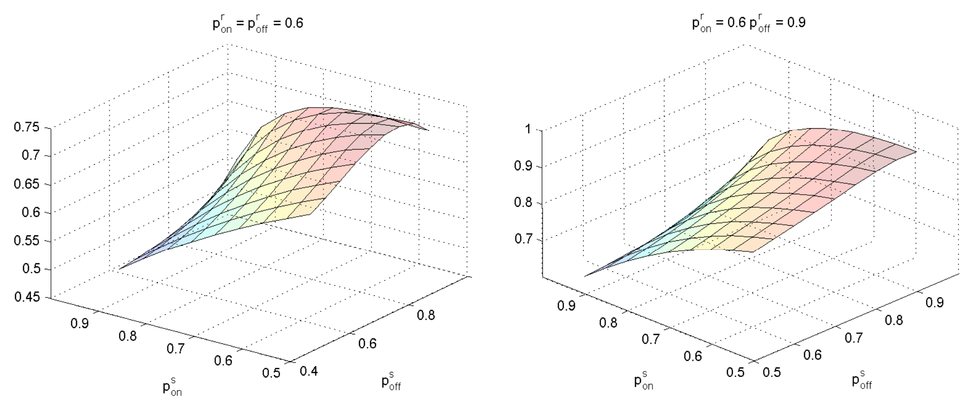
[8] E. Fernandez-Gaucherand, A. Arapostathis and S. Marcus, “On the average cost optimality equation and the structure of optimal policies for partially observable Markov decision processes,” *Annals of Operations Research,* vol. 29, no. 1-4, pp. 439-470, April 1991.

[9] H. Yousefizahed and H. Jafarkhani, “Statistical guarantee of QoS in communication networks with temporally correlated loss,” *Proc. IEEE GLOBECOM,* November 2003.

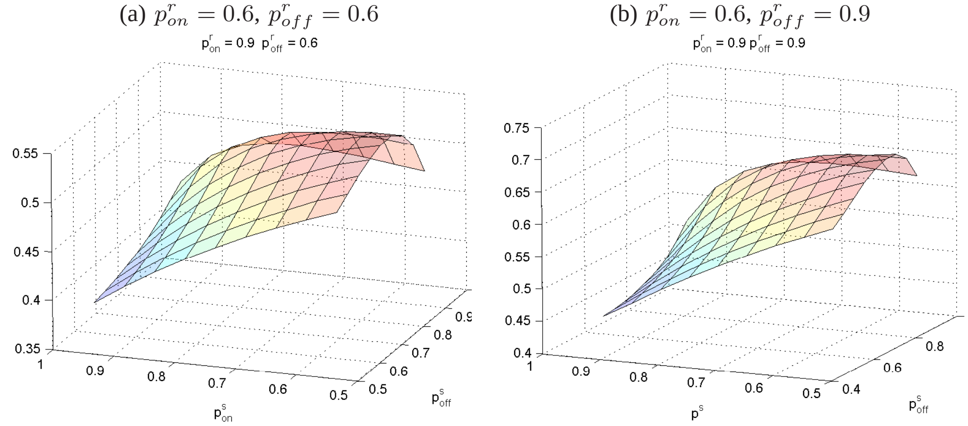


(a) $p^r_{on} = 0.6$, $p^r_{off} = 0.6$ (b) $p^r_{on} = 0.6$, $p^r_{off} = 0.9$ (c) $p^r_{on} = 0.9$, $p^r_{off} = 0.6$ (d) $p^r_{on} = 0.9$, $p^r_{off} = 0.9$

Fig. 1. Effect of p_{on} and p_{off} on the packet delivery ratio (z-axis). Parameters used: $q^s_{on} = q^r_{on} = 0.85$, $q^s_{off} = q^r_{off} = 0.7$, $c^s = c^r = 1$, $\delta^s_1 = \delta^r_1 = 2$, $\delta^s_2 = \delta^r_2 = 1$, $K = 10$.



(a) $p^r_{on} = 0.6$, $p^r_{off} = 0.6$ (b) $p^r_{on} = 0.6$, $p^r_{off} = 0.9$



(c) $p^r_{on} = 0.9$, $p^r_{off} = 0.6$ (d) $p^r_{on} = 0.9$, $p^r_{off} = 0.9$

Fig. 2. Effect of p_{on} and p_{off} on the packet delivery ratio (z-axis). Parameters used: $q^s_{on} = q^r_{on} = 0.85$, $q^s_{off} = q^r_{off} = 0.7$, $c^s = c^r = 1$, $\delta^s_1 = \delta^r_1 = 2$, $\delta^s_2 = \delta^r_2 = 1$, $K = 10$.