Opportunistic Relay Selection for Cooperative Energy Harvesting Communication Networks

Yong Xiao*, Zhu Han† and Luiz A. DaSilva‡§
*Singapore University of Technology and Design, Singapore
†Electrical and Computer Engineering, University of Houston, TX
‡CTVR, Trinity College Dublin, Ireland
§Department of Electrical and Computer Engineering, Virginia Tech, VA

***Abstract*—We consider cooperative energy harvesting communication networks in which a set of source-to-destination pairs competes for a limit number of relay nodes with energy harvesting ability. The performance of each source has been affected by two interactions: the interaction between the sources and relay nodes and the interaction among sources. We model the first interaction as a college admission market and then fits this market into a stochastic environment. We formulate an interactive partially observable Markov decision process (I-POMDP) to study the second interaction. We derive the optimal policy for the sources to sequentially optimize their decisions. Numerical results show that our proposed policy significantly improves the performance of sources.**

***Index Terms*—Energy harvesting, college admission, stable matching, opportunistic, relay selection, cooperation, POMDP.**

I. Introduction

Recently, energy harvesting technology has attracted significant interest because of its potential to provide a sustainable energy source for electronic devices. This is critically important for mobile communication systems where power charging/recharging service for mobile devices cannot be always available. This motives recent research in energy harvesting communication networks in which each device transmits signals using the energy harvested from the surrounding natural environment such as sun light, wind, radio frequency energy, etc. As the ambient energy available for harvesting is generally sporadic and random, optimization of this system is generally a challenging task.

Most of the existing work focuses on a fixed network structure and assumes each energy harvesting device can have the prefect information about the future trend of the energy replenish process of the nature. Specifically, in [1], an optimal scheduling strategy has been proposed for a single transmitter that can harvest and store unlimited energy. The assumption of the unlimited energy storage has been relaxed in [2]. The power scheduling problem has been studied for broadcast channels and multiple access channels with energy harvesting transmitters in [3] and [4], respectively. Motivated by the observation that the network performance can be further improved by allowing multiple closely-located devices to cooperate and relay signals for each other, the cooperative energy harvesting system has been studied in [5], [6]. Specifically, the multi-hop relaying channel with an energy harvesting relay node has been studied from the information theoretic perspective in [5]. The model has been further extended in [6] by allowing one-way energy transfer from the source to relay node.

We focus on the relay selection problem for a cooperative energy harvesting communication network in a stochastic environment. This is different from the relay selection in a time-invariant environment in which the sources simply compete for the relay nodes that can provide the highest instantaneous performance. In a stochastic environment, each source should also take into account the future evolution of the environment. For example, suppose a source believes the energy available for the transmission in the next time slot will be much lower than that in the current time slot. This source may not choose the relay node that can provide the highest energy saving in the current time slot but can save this relay node for the next time slot of transmission. The optimization of the relay selection problem becomes more complex when multiple sources can compete for the same set of relay nodes because it is possible that the relay nodes with high battery levels saved by some sources for the future transmission are used by other sources in the current time slot. In other words, the performance of each source can be affected by two interactions: the *interaction between sources and relay nodes* and the *interaction among source-to-destination pairs*. The first interaction specifies the negotiation process between each source-to-destination pair and relay nodes as well as the condition for which a mutual agreement can be established. The second interaction characterizes the competition among sources for the limited number of relay nodes.

In our model, both sources and relay nodes can decide their partners and a source-and-relay cooperation pair can only be formed when such a pairing is mutually agreed by both pairing sources and relay nodes. We assume each source cannot know the exact future development of the energy harvesting process or the decisions and payoffs of other sources or relay nodes.

We establish a college admission market to analyze the interaction between sources and relay nodes. In our model, each source can establish and maintain a belief about the future evolution of the environment and probability distribution of the possible decisions of other sources. The

This work is partially supported by the National Science Foundation under Grants CNS-1443917, ECCS-1405121, CNS-1265268, and CNS-0953377 and the Science Foundation Ireland under Grant 10/IN.1/I3007.

preference of each source over the relay nodes is decided by both the current observation and the belief. We fit this stable matching market into a stochastic environment where we establish an interactive partially observable Markov decision process (I-POMDP) framework to analyze the interaction among the source-to-destination pairs. We derive the optimal policy for each source to improve the long term expected payoff by sequentially optimizing its relay selection decision. Our model is general and the payoff of each source can be any performance measure. To the best of our knowledge, this is the first work that the cooperative energy harvesting communication system has been modeled and analyzed using a joint framework of the college admission market and I-POMDP.

II. Network Model and Problem Formulation

Consider an energy harvesting communication system with K source-to-destination pairs, denoted as s_1 to d_1, s_2 to d_2, ..., s_K to d_K, and a set of M relay nodes, denoted as $\mathcal{R} = \{r_1, r_2, \ldots, r_M\}$, that can receive and forward signals for the source-to-destination pairs using their harvested energy. We write $\mathcal{S} = \{s_1, s_2, \ldots, s_K\}$ and $\mathcal{D} = \{d_1, d_2, \ldots, d_K\}$. We assume the entire transmission time can be divided into time slots. We use superscript t to denote parameters in the tth time slot. Sources and relay nodes can harvest energy from their surrounding natural environment and use the collected energy to support their signal transmission and relaying. Let $\hat{e}^t_{s_k}$ and $\hat{e}^t_{r_m}$ be the amount of energy harvested by s_k and r_m during time slot t, respectively. Each source s_k and relay node r_m has been equipped with a rechargeable battery that can store no more than $\bar{e}_{s_k}$ and $\bar{e}_{r_m}$ amounts of energy, respectively. We consider an energy harvesting system with *causality constraints*, that is, each source or relay node cannot use the energy that will be harvested in the future, i.e., the battery levels of source s_k and relay node r_m at the beginning of each time slot t are given by $e^t_{s_k} = \min\left\{\bar{e}_{s_k}, \left(e^{t-1}_{s_k} + \hat{e}^{t-1}_{s_k} - w^{t-1}_{s_k}\right)\right\}$ and $e^t_{r_m} = \min\left\{\bar{e}_{r_m}, \left(e^{t-1}_{r_m} + \hat{e}^{t-1}_{r_m} - w^{t-1}_{r_m}\right)\right\}$ where $w^{t-1}_{s_k}$ is the transmit power of source s_k in time slot $t-1$ and $w^{t-1}_{r_m}$ is the transmit power of relay node r_m used to forward signal received from the source during time slot $t-1$ for $0 \le w^{t-1}_{s_k} \le e^{t-1}_{s_k}$ and $0 \le w^{t-1}_{r_m} \le e^{t-1}_{r_m}$.

We assume in each time slot each source can choose at most one relay node and each relay node r_m can only help the transmission for a limited number of source-to-destination pairs. Let q_m be the maximum number of source-to-destination pairs that can be supported by relay node r_m in each time slot. In this paper, we mainly focus on the performance improvement brought by sequential optimization of the relay selection process for the sources. We assume the transmit power of each source and relay node during each time slot can be regarded as an one-to-one correspondence of its battery level, i.e., $w^t_{s_k} = f\left(e^t_{s_k}\right)$ and $w^t_{r_m} = g\left(e^t_{s_k}\right)$ where $f(\cdot)$ and $g(\cdot)$ are two fixed functions decided by the source s_k and relay node r_m, respectively.

We consider a stochastic environment, that is, the physical state of the environment such as the energy that is available to be harvested by both sources and relay nodes, channel gains between sources and relay nodes as well as that between relay nodes and destinations, etc., changes from time to time. Let $h^t_{s_k r_m}$ be the ratio of the channel gain between s_k and r_m to the additive noise received by r_m in time slot t. Let $h^t_{r_m d_k}$ and $h^t_{s_k d_k}$ be the ratio of the channel gain between r_m and d_k and that between s_k and d_k to the additive noise received by d_k in time slot t, respectively. We define the *physical state* of the energy harvesting communication system as the state of the system environment denoted as $\eta^t = \langle \boldsymbol{e}^t_s, \boldsymbol{e}^t_r, \boldsymbol{h}^t_{sd}, \boldsymbol{h}^t_{sr}, \boldsymbol{h}^t_{rd} \rangle$ where $\boldsymbol{e}^t_s = \{e^t_{s_k}\}_{s_k \in \mathcal{S}}$, $\boldsymbol{e}^t_r = \{e^t_{r_m}\}_{r_m \in \mathcal{R}}$, $\boldsymbol{h}^t_{sd} = \left\{h^t_{s_k d_k}\right\}_{s_k \in \mathcal{S}, d_k \in \mathcal{D}}$, $\boldsymbol{h}^t_{sr} = \left\{h^t_{s_k r_m}\right\}_{s_k \in \mathcal{S}, r_m \in \mathcal{R}}$, and $\boldsymbol{h}^t_{rd} = \left\{h^t_{r_m d_k}\right\}_{r_m \in \mathcal{R}, d_k \in \mathcal{D}}$. Note that the harvested energy and the channel gains are generally continuous variables and may have infinite number of possible values. However, because of the limit of the accuracy for digital communication devices, we can assume the energy available at the sources and relay nodes as well as the possible channel gain of the relay and direct transmission channel in each time slot can only be a discrete value from a finite set [7]. Let Υ be the set of possible physical states of the energy harvesting communication system, i.e., $\eta^t \in \Upsilon$.

Each source needs to decide whether or not to use the help of relay nodes and which relay node it should choose at the beginning of each time slot and cannot change its decision during the rest of the time slot. It can be easily observed that each source prefers different relay nodes to forward its signal under different physical states. For example, different energy harvested by the sources and relay nodes in different time slots can affect transmit powers $w^t_{s_k}$ and $w^t_{r_m}$ as well as the final payoffs obtained by both pairing sources and relay nodes. Let $a^t_{s_k}$ be the relay node decided by source s_k to send signals at the beginning of time slot t. We write $a^t_{s_k} = \emptyset$ if source s_k decides to directly send its signal without the help of any relay node. Let $\mathcal{A}_{s_k} = \mathcal{R} \cup \{\emptyset\}$ be the set of possible decisions of source s_k. At the beginning of each time slot t, each source s_k sends signal to relay node $a^t_{s_k}$ according to its decision. If relay node $a^t_{s_k} \in \mathcal{R}$ agrees to relay signal for source s_k, it will forward the received signal to destination d_k. However, each relay node can also decide to not help the transmission of source s_k. In this case, the signal sent by the source s_k will be discarded and relay node $a^t_{s_k}$ will feedback a rejection message to source s_k. Note that, allowing each relay node to ignore the signal sent by some sources may cause energy waste. A handshaking-like protocol can be enforced at the beginning of every transmission to avoid such a waste of energy. For example, each source first sends a request signal to the relay node and then waits for the acknowledgement from the relay node it requests. The main shortcoming of this handshaking protocol is that it will always cost extra energy and time for communication at the beginning of each time slot. As will be proved later, we will propose a distributed policy for each source to learn from the past observation and avoid choosing the relay node who will ignore its signal. In other words, the energy waste caused by directly sending signals to the relay node without handshaking protocols could be ignored if the

signal sent by sources can always be forwarded by relay nodes of their choices.

In energy harvesting communication systems, both sources and relay nodes try to maximize the efficiency of their energy utilization. Specifically, each source prefers to choose the relay node that provides the highest energy saving for sending its data signal. On the other hand, each relay node prefers to serve as the relay for a set of sources that has the largest amount of data packets to forward. Let $\varpi_{s_k,r_m}(\eta^t)$ and $\varpi_{s_k,s_k}(\eta^t)$ be the instantaneous payoffs of each source s_k with and without the help of a relay node $r_k \in \mathcal{R}$ at time slot t, respectively. We assume $\varpi_{s_k,r_m}(\eta^t) \neq \varpi_{s_k,r_n}(\eta^t)$ for $m \neq n$, $r_m, r_n \in \mathcal{R}$ and $\eta^t \in \Upsilon$. We consider a general model and the payoffs of sources and relay nodes can be any performance measure or function. For example, in a two-hop relay channel-based cooperative energy harvesting network [5], [6] with each source s_k trying to maximize its transmission rate, we can write the payoff of the each source as

$$\begin{aligned}\varpi^t_{s_k,r_m}(\eta^t) &= 0.5\min\left\{\log\left(1+h^t_{s_k r_m}w^t_{s_k}\right),\right.\\ &\qquad \left.\log\left(1+h^t_{r_m d_k}w^t_{r_m}\right)\right\}, \quad (1)\\ \varpi^t_{s_k,s_k}(\eta^t) &= \log\left(1+h^t_{s_k d_k}w^t_{s_k}\right). \quad (2)\end{aligned}$$

Similarly, each relay node r_m can also obtain a positive payoff, denoted as $\varpi_{r_m,\mathcal{K}}(\eta^t)$ when it helps the transmission of a set $\mathcal{K}$ of sources in time slot t. Let $\mathcal{K}^t_m \subseteq \mathcal{S}$ be the set of sources that are helped by relay node r_m in time slot t. For example, if the payoff of relay r_m corresponds to the data bit per energy to forward the received data signal, i.e., we can write the payoff of relay node r_m obtained from set $\mathcal{K}^t_m$ of sources as

$$\varpi_{r_m,\mathcal{K}^t_m}(\eta^t) = \frac{\beta_m}{w^t_{r_m}}\sum_{s_k\in\mathcal{K}^t_m}\log\left(1+h^t_{s_k r_m}w^t_{s_k}\right) \quad (3)$$

where β_m is the pricing coefficient decided by relay node r_m.

As each relay node can only help the transmission of q_m sources at a time, a *conflict* will happen if more than q_m sources transmit signals to the same relay node. Let $\mathcal{U}^t_{r_m} = \{s_k : a^t_{s_k} = r_m,\ \forall s_k \in \mathcal{S}\}$ be the set of sources sent signals to relay node r_m at the beginning of time slot t. In this paper, we assume each relay node is myopic and always tries to maximize its instantaneous payoff in each time slot. We introduce the following *conflict-solving rule* for the relay node: if more than q_m sources send signal to the same relay node r_m, r_m will only receive and forward the signals sent by a set $\mathcal{K}^t_m$ of sources that can provide the maximum payoff, i.e., for each relay node $r_m \in \mathcal{R}$ with $|\mathcal{U}^t_{r_m}| \geq q_m$, the set $\mathcal{K}^{*t}_m$ of sources whose signals will be forwarded by relay node r_m is given by $\mathcal{K}^{*t}_m = \max_{\mathcal{K}^t_m\subseteq\mathcal{U}^t_{r_m},|\mathcal{K}^t_m|=q_m}\varpi_{r_m,\mathcal{K}^t_m}(\eta^t)$[1]. We assume $\varpi_{r_m,\mathcal{K}^t_m}(\eta) \neq \varpi_{r_m,\mathcal{L}^t_m}(\eta)\ \forall\ \mathcal{L}^t_m \neq \mathcal{K}^t_m$ and $\mathcal{L}^t_m, \mathcal{K}^t_m \subseteq \mathcal{S}$.

Let $\mu(s_k,\eta^t)$ be the relay node that accepts the request of source s_k in time slot t for $\mu(s_k,\eta^t) \in \mathcal{R}\cup\{s_k\}$. We use $\mu(s_k,\eta^t) = s_k$ to mean that s_k has been rejected by its requesting relay node. $\mu(s_k,\eta^t)$ can be regarded as the result of decisions $\boldsymbol{a}^t_s = \left\{a^t_{s_k}\right\}_{s_k\in\mathcal{S}}$ of all sources and physical state η^t in each time slot t, i.e., we can introduce a mapping function $\mu(s_k,\eta^t) = P_{s_k}(\boldsymbol{a}^t_s,\eta^t)$ for all $s_k \in \mathcal{S}$. It is generally unrealistic to assume each source can know the decisions of other sources before it makes its own decision at the beginning of each time t. It is however possible for each source to eavesdrop the decision of other sources in the past. Therefore, in this paper, we assume each each source can observe the decisions of other sources in the previous time slots. However, each source cannot know the conflict-solving rules of the relay nodes as well as the payoffs or the current decisions of other sources. We also assume each sources cannot observe the prefect information of the physical state but can obtain an common observation denoted as o^t about the physical state of the environment at the beginning of each time slot t. We assume the set Ω of possible observation is a finite set, i.e., $o^t \in \Omega$.

Since each source cannot have the prefect information about current physical state and the actions of others, it cannot know which relay node will maximize its payoff. However, if each source s_k can establish a belief function about the physical state and the distribution of the joint decision of other sources in the current time slot, it can estimate the expected payoff obtained by choosing each of the relay nodes. More specifically, if each source can establish a belief function $B_{s_k}\left(\eta^t,\boldsymbol{a}^t_{-s_k}\right)$ about the uncertainty of the current physical state and decisions of other sources, we can write the expected payoff of each source for a given $a^t_{s_k}$ as

$$\bar{\varpi}^t_{s_k}\left(a^t_{s_k}\right) = \sum_{\boldsymbol{a}^t_{-s_k},\eta^t} B_{s_k}\left(\eta^t,\boldsymbol{a}^t_{-s_k}\right)\varpi_{s_k,\mu(s_k,\eta^t)} \quad (4)$$

where $\mu(s_k,\eta^t) = P_{s_k}(\boldsymbol{a}^t_s,\eta^t)$ and $\varpi_{s_k,\mu(s_k,\eta^t)}$ is given in (1) and (2). Note that each decision $a^t_{s_k}$ of source s_k corresponds to the relay node that source s_k decides to send signal. Each source s_k can then establish a preference over all relay nodes for the given belief by ranking its expected payoff $\bar{\varpi}^t_{s_k}\left(a^t_{s_k}\right)$ for each of its decisions $a^t_{s_k} \in \mathcal{R}$ from the highest to the lowest.

In this paper, we follow the same line as [8] and assume the belief of each source about the decisions of other sources follows a Dirichlet distribution function. The main objective for each source s_k is to maximize its long-term expected payoff $\bar{\varpi}_{s_k}$ given by

$$\bar{\varpi}_{s_k} = \lim_{T\to\infty}\sum_{t=1}^{T}\rho^t\bar{\varpi}^t_{s_k} \quad (5)$$

through sequentially optimizing its decision about the relay node where $0 < \rho < 1$ is a discount factor.

III. A College Admission Market for Cooperative Energy Harvesting Communication Systems

As observed at the end of Section III, the interaction among sources plays an important role in cooperative energy

[1]Note that there will be cross-interference when multiple sources send their signals to the same relay node. However, as will be shown later in this paper, if each source can learn from the past history and avoid such conflicts in the future, this cross-interference will disappear.

harvesting communication systems. We can model the relay selection problem of sources as a multi-agent (subintentional) interactive partially observable Markov decision process (I-POMDP) [9] as follows. An *action* of each source s_k in time slot t, denoted as $a^t_{s_k}$, is the relay node it decided to send signals. In this paper, we use terms "action" and "decision" interchangeably. Let $\boldsymbol{\mathcal{A}} = \{\mathcal{A}_{s_k}\}_{s_k \in \mathcal{S}}$ be the set of possible joint action profiles of the sources. We define the *interactive state* of the system in each time slot t as $\lambda^t_{s_k} = \langle \eta^t, \theta^t_{s_k} \rangle$ where η^t is the physical state of the system defined in Section II and $\theta^t_{s_k}$ is the set of possible models for the opponents of source s_k, defined as $\theta_{s_k} = \langle H_{-s_k}, F_{-s_k} \rangle$ where H_{-s_k} is the observation history of all sources other than s_k and $F_{-s_k} : H_{-s_k} \rightarrow \Delta(\boldsymbol{\mathcal{A}}_{-s_k})$ is the mapping from H_{-s_k} to the distribution over the joint action of all sources other than source s_k. Let Λ_{s_k} be the set of all the possible interactive states for source s_k and $\boldsymbol{\Lambda} = \{\Lambda_{s_k}\}_{s_k \in \mathcal{S}}$ be the set of possible state profiles for all sources. We define the *observation function* $O(o, \eta, \boldsymbol{a}_s)$ as the probability of obtaining observation o when the physical state is given by η and the action profile of the sources in the previous time slot is given by $\boldsymbol{a}_s$, i.e., $O(o, \eta, \boldsymbol{a}_s) = \Pr(o|\eta, \boldsymbol{a}_s)$. We follow the same line as [8] and assume that each source knows the observation function and presumes that other sources make independent decisions about their actions according to a fixed but unknown distribution. $\Gamma : \Upsilon \times \mathcal{A} \times \Upsilon \rightarrow [0, 1]$ is the *state transition function* and $\Gamma(\eta', \eta, \boldsymbol{a}_s)$ specifies the probability distribution that, starting at a physical state η and action profile $\boldsymbol{a}_s$, the physical state ends in η'. This transition function can be estimated from the system model or using the training methods [10]. In this paper, we assume this state transition function is perfectly known by all sources.

The main objective for each source is to transmit signals to a proper relay node that will not only agree to relay its signal but also maximize its long-term expected payoff. We now consider the interaction among sources and that between sources and relay nodes within one time slot. We can model the cooperative energy harvesting communication system in each time as a *college admission market*, referred to as the cooperative energy harvesting communication (CEH) market. The college admission market is also known as two-sided many-to-one matching market in which a set of students (sources in our model) from one side of the market tries to be paired with colleges (relay nodes in our model) from the other side. We present a formal definition as follows:

Definition 1: A *CEH market* $\mathcal{G} = \langle \mathcal{S}, \mathcal{R}, \succ \rangle$ consists of a set $\mathcal{S}$ of sources, a set $\mathcal{R}$ of relay nodes and the preference $\succ$.

As each source-to-destination pair can only choose one relay node and hence the main objective for the cooperative energy harvesting communication system is to find a proper relay node for each source. We refer to this structure as a *matching* which is formally defined as follows.

Definition 2: A *matching* μ of a CEH market in each time slot with physical state η is a function from the set $\mathcal{S} \cup \mathcal{R}$ into the set of unordered families of elements of $\mathcal{S} \cup \mathcal{R}$ such that $|\mu(s_k, \eta)| = 1$ for every $s_k \in \mathcal{S}$ and $\mu(s_k, \eta) = s_k$ if $\mu(s_k, \eta) \notin \mathcal{R}$, $|\mu(r_m, \eta)| \leq q_m$ for every $r_m \in \mathcal{R}$, if $\mu(s_k, \eta) \neq s_k$, then $\mu(s_k, \eta) \in \mathcal{R}$ and $\mu(s_k, \eta) = r_m$ if and only if $s_k \in \mu(r_m, \eta)$.

Our cooperative energy harvesting communication system consists of a stochastic environment and, to find the matching between the sources and relay nodes under different situations, we should establish a decision making rule for each source, referred to as a *policy* $\pi : \boldsymbol{\Lambda} \rightarrow \boldsymbol{\mathcal{A}}$, which is a mapping that specifies, for each interactive state profile, an action to be taken by each source. For each given state, each joint action of the sources can result in a specific matching between sources and relay nodes. However, this matching is generally not stable which means that at least one individual source or relay node, or a pair of source and relay node can choose different partners to further improve its payoff. In this paper, we seek a policy which specifies a sequence of actions taken by each source that can converge to a sequence of *stable matching* between sources and relay nodes in which no source or a pair of source and relay node believes it can further improve its payoff by unilaterally deviating from its existing pairing partner.

One of the main challenges to find this policy is that the stable matching will change with the environment and it is generally difficult to find a sequence of stable matchings that adopts to this change without knowing the instantaneous interactive state and actions of other sources. In addition, there may exist multiple stable matching structures for each specific interactive state and it is possible for the sources to keep jumping between different stable matching structures without converging to a specific matching structure under each resulting state.

To solve the above problem, we needs to extend the CEH market into a stochastic environment by introducing a *belief function* for each source. Specifically, each source can establish and maintain a belief function $b_{s_k}\left(\lambda^t_{s_k}\right)$ about the interactive state $\lambda^t_{s_k}$. This belief function characterizes the subjective probability of each source about the interactive state of the system i.e., we have $b_{s_k}\left(\lambda^t_{S_k}\right) = \Pr\left(\eta^t, \theta^t_{s_k} | o^t, \boldsymbol{a}^{t-1}_s, b^{t-1}_{s_k}\right) = b_{s_k}(\eta^t)\ b_{s_k}\left(\theta^t_{s_k} | \eta^t\right)$, where $b_{s_k}(\eta^t) = \Pr\left(\eta^t | o^t, \boldsymbol{a}^{t-1}_s, b^{t-1}_{s_k}\right)$ and $b_{s_k}\left(\theta^t_{s_k} | \eta^t\right) = \Pr\left(\theta^t_{s_k} | \eta^t, o^t, \boldsymbol{a}^{t-1}_s, b^{t-1}_{s_k}\right)$.

We can now show that the uncertainty of each source about the actions of others and physical states can be converted to its uncertainty about the interactive state. As mentioned previously, the interactive state of each source s_k consists of the physical state and the source's belief about the model of other sources which according to the discussion in Section II can determine the expected matching partner as well as the expected payoff of each source. Therefore, in the rest of this paper, we focus on the belief function $b_{s_k}\left(\lambda^t_{s_k}\right)$ for each source s_k.

IV. An Optimal Policy

In our system, the belief of each source should be updated based on its observation at the beginning of each time slot. The belief function for each source should also be sufficient to summarize its past observation history.

We introduce the following belief updating function for each

source s_k at the beginning of each time slot t,

$$b^t_{s_k}\left(\lambda^t_{s_k}\right) = \zeta O\left(o^t, \eta^t, \boldsymbol{a}^{t-1}_s\right) \sum_{\eta^{t-1}\in\Upsilon} \Gamma\left(\eta^t, \eta^{t-1}, \boldsymbol{a}^{t-1}_s\right) \quad (6)$$
$$b^{t-1}_{s_k}\left(\eta^{t-1}\right)\left(\frac{\sum_{u=1}^{t-1}\mathbf{1}\left(\boldsymbol{a}^u_{-s_k}=\boldsymbol{a}^{t-1}_{-s_k}, \eta^u=\eta^{t-1}\right)}{\sum_{u=1}^{t-1}\mathbf{1}\left(\eta^u=\eta^{t-1}\right)}\right)$$

where $\mathbf{1}(\cdot)$ is the indicator function and ζ is the normalizing constant.

The above belief function can be used to update the belief of each source in our proposed cooperative energy harvesting communication system. We need to prove that belief $b^t_{s_i}(\lambda^t)$ is a *sufficient statistic* which means that each source can make decisions about its future actions without requiring any further information about the past observation history.

Proposition 1: In our proposed cooperative energy harvesting communication system, the current belief $b^t_{s_i}(\lambda^t)$ of each source s_i is a sufficient statistic for the past history of source s_i's observations.

Proof: Using the Baye's rule, we have

$$\begin{aligned}
b^t_{s_k}\left(\lambda^t\right) &= \Pr\left(\eta^t, \theta^t_{s_k} | o^t, \boldsymbol{a}^{t-1}_s, b^{t-1}_{s_k}\right) \\
&= \zeta \Pr\left(o^t | \eta^t, \theta^t_{s_k}, \boldsymbol{a}^{t-1}_s, b^{t-1}_{s_k}\right) \Pr\left(\eta^t, \theta^t_{s_k} | \boldsymbol{a}^{t-1}_s, b^{t-1}_{s_k}\right) \\
&= \zeta O\left(o^t, \eta^t, \boldsymbol{a}^{t-1}_s\right) \sum_{\eta^{t-1}\in\Upsilon} \Pr\left(\eta^t, \theta^t_{s_k}, \eta^{t-1} | \boldsymbol{a}^{t-1}_s, b^{t-1}_{s_k}\right) \\
&= \zeta O\left(o^t, \eta^t, \boldsymbol{a}^{t-1}_s\right) \sum_{\eta^{t-1}\in\Upsilon} \Pr\left(\eta^t | \eta^{t-1}, \boldsymbol{a}^{t-1}_s\right) \\
&\quad \Pr\left(\theta^t_{s_k}, \eta^{t-1} | \eta^t, \boldsymbol{a}^{t-1}_s, b^{t-1}_{s_k}\right), \\
&= \zeta O\left(o^t, \eta^t, \boldsymbol{a}^{t-1}_s\right) \sum_{\eta^{t-1}\in\Upsilon} \Gamma\left(\eta^t, \eta^{t-1}, \boldsymbol{a}^{t-1}_s\right) \\
&\quad \Pr\left(\theta^t_{s_k} | \eta^t, \boldsymbol{a}^{t-1}_s, b^{t-1}_{s_k}\right) b^{t-1}_{s_k}\left(\eta^{t-1}\right). \qquad (7)
\end{aligned}$$

It has been shown in [8] that if the belief of each source about actions of its opponents follows a Dirichlet distribution, we can write the last term in the right-hand-side of (7) as

$$\begin{aligned}
&\Pr\left(\theta^t_{s_k} | \eta^t, \boldsymbol{a}^{t-1}_s, b^{t-1}_{s_k}\right) \qquad (8) \\
&= \frac{\sum_{u=1}^{t-1}\mathbf{1}\left(\boldsymbol{a}^u_{-s_k}=\boldsymbol{a}^{t-1}_{-s_k}, \eta^u=\eta^{t-1}\right)}{\sum_{u=1}^{t-1}\mathbf{1}\left(\eta^u=\eta^{t-1}\right)}.
\end{aligned}$$

From (7) and (8), we can claim that the current belief of the source depends on the observation and system transition functions as well as the previous observations and belief function of each source all of which can be fully characterized by the parameters from the current and previous time slots. ■

To maximize the payoff defined in (5), each source needs to decide the proper relay node to send the signal. We define a value function $V_{s_i}(\lambda_{s_k})$ as the expected discount sum of the current and future payoff for each source started at current interactive state λ_{s_k}. This value function should contain two parts: 1) the immediate expected payoff which is given by

$$R_{s_k}(\lambda_{s_k}) = \sum_{\lambda_{s_k}\in\Lambda_{s_k}} b_{s_k}(\lambda_{s_k})\, \varpi_{s_k,\mu(s_k,\eta)}, \qquad (9)$$

and 2) the expected discounted payoff in all the possible states and actions in the future started with state λ_{s_k}.

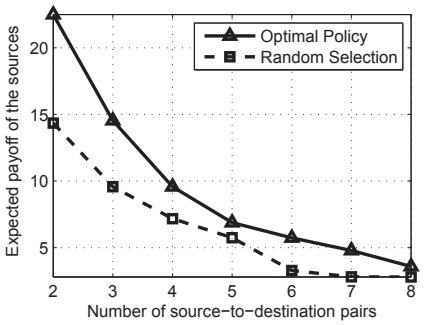

Fig. 1. Average payoff of the sources under different numbers of source-to-destination pairs where there are 6 relay nodes and 5 number of states.

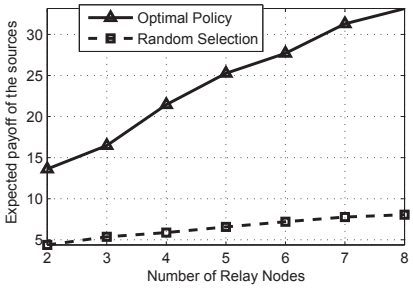

Fig. 2. Expected payoff of the sources under different numbers of source-to-destination pairs where there are 6 relay nodes and 5 number of states.

We can define a value function for source s_k at time slot t with state λ_{s_k} as follows,

$$\begin{aligned}
V_{s_k}(\lambda_{s_k}) &= \max_{a_{s_k}\in\mathcal{A}_{s_k}} R_{s_k}(\lambda_{s_k}) + \rho \sum_{o\in\Omega} \Pr\left(o | \boldsymbol{a}_s, b_{s_k}\right) \\
&\qquad V_{s_k}\left(\tau_{s_k}\left(b_{s_k}, \boldsymbol{a}_s, o\right)\right). \qquad (10)
\end{aligned}$$

where $\tau_{s_k}(b_{s_k}, \boldsymbol{a}_s, o)$ is the updated belief of source s_k if the current action profile, belief and observation are given by $\boldsymbol{a}_s$, $b_{s_k}(\lambda_{s_k})$ and o, respectively.

We can then define the optimal policy π that decides the action of each source s_k under each interactive state as

$$\begin{aligned}
a^*_{s_k} &= \arg\max_{a_{s_k}\in\mathcal{A}_{s_k}} R_{s_k}(\lambda_{s_k}) + \rho \sum_{o\in\Omega} \Pr\left(o | \boldsymbol{a}_s, b_{s_k}\right) \\
&\qquad V_{s_k}\left(\tau_{s_k}\left(b_{s_k}, \boldsymbol{a}_s, o\right)\right). \qquad (11)
\end{aligned}$$

V. Numerical Results

To verify the performance of our results, we consider a cooperative energy harvesting communication network consisting of multiple source-to-destination pairs and relay nodes that are uniformly randomly located in a square-shaped coverage area. We assume the transmission of different source-to-destination pairs are orthogonal. We consider the payoff of sources and relay nodes defined in (1)–(3) and let the channel gain between two nodes be $h_{xy} = \frac{\tilde{h}_{xy}}{\sqrt{D^{\sigma}_{xy}}}$ where $x \in \mathcal{S}\cup\mathcal{R}$ and $y \in \mathcal{R}\cup\mathcal{D}$, $\tilde{h}_{xy}$ is the channel fading coefficient unrelated to the transmission distance, D_{xy} is the distance between x and y, and σ is the pathloss exponent. To simplify our simulation, we assume $\tilde{h}_{xy}$ can be regarded as a constant. We do not consider the

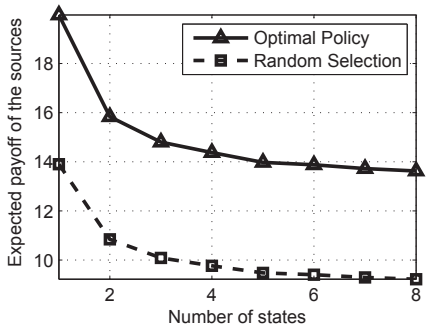

Fig. 3. Expected payoff of the sources under different numbers of states where there are 4 source-to-destination pairs and 4 relay nodes.

optimization of the power scheduling and assume the source and relay node always use all the energy collected from previous time slot to send and forward signals. As the energy that can be harvested by source and relay node is generally a continuous function of time, if the duration of each time slot is short enough, we can assume the probability for the harvested energy level transiting from the current level to the neighboring one is much more likely than other states. Specifically, we consider a system with N levels of energy harvested by each relay node and source and, without loss of generality, we assume the possible energy that can be harvested by the source or relay node can be written as $\hat{e}_{s_k}, \hat{e}_{r_m} \in \{\frac{1}{N}, \frac{2}{N}, \ldots, 1\}$. We assume that if the current energy harvested by source s_k and relay node is $\hat{e}^t_{s_k}$ and $\hat{e}^t_{r_m}$, the probability distribution of the harvested energy in the next time slot is given by $\Pr\left(\hat{e}^{t+1}_{s_k} = \hat{e}^t_{s_k} \pm \frac{1}{N}\right) = p$, $\Pr\left(\hat{e}^{t+1}_{s_k} = \hat{e}^t_{s_k}\right) = 1 - 2p$, $\Pr\left(\hat{e}^{t+1}_{r_m} = \hat{e}^t_{r_m} \pm \frac{1}{N}\right) = q$ and $\Pr\left(\hat{e}^{t+1}_{r_m} = \hat{e}^t_{r_m}\right) = 1 - 2q$ for $\frac{2}{N} \leq \hat{e}^t_{s_k}, \hat{e}^t_{r_m} \leq \frac{N-1}{N}$ and if $\hat{e}^t_{s_k} \in \{0, 1\}$ or $\hat{e}^t_{r_m} \in \{0, 1\}$, we have $\Pr\left(\hat{e}^{t+1}_{s_k} = \frac{1}{N} | \hat{e}^t_{s_k} = 0\right) = \Pr\left(\hat{e}^{t+1}_{s_k} = \frac{N-1}{N} | \hat{e}^t_{s_k} = 1\right) = 2p$ or $\Pr\left(\hat{e}^{t+1}_{r_m} = \frac{1}{N} | \hat{e}^t_{r_m} = 0\right) = \Pr\left(\hat{e}^{t+1}_{r_m} = \frac{N-1}{N} | \hat{e}^t_{r_m} = 1\right) = 2q$ where $0 < p < 0.5$ and $0 < q < 0.5$ are constants.

In Figure 1, we fix the number of relay nodes and physical states and compare the payoff of sources under different number of source-to-destination pairs. We observe that our optimal policy improves the performance of cooperative energy harvesting communication system. This is because, with the increasing of the number of sources, the effects of the interaction among sources on the payoff of each source will also increase. Note that the decreasing of the expected payoff when the number of source-to-destination pairs is small is caused by the averaging process of our simulation.

In Figure 2, we consider the expected payoff of sources under different numbers of the relay nodes. We observed that using the optimal policy to decide the relay node of each source can provide a significant payoff improvement especially when the number of relay nodes becomes large. It is also observed that the random selection can only provide limited payoff improvement when the number of relay nodes grows. This is because as mentioned in the introduction, exploiting the relay node to forward its signal cannot always provides performance improvement [11] and hence deciding the optimal relay node for each source becomes very important especially when the number of relay nodes in the given service area becomes large.

In Figure 3, we fix the number of the source-to-destination pairs and relay nodes and consider the payoff of the sources under different numbers of physical states. As we fixed the value of p, the more the number of physical states, the slower the physical state of the environment will change. If it only has one physical state, the system will become stationary and the energy harvested by the sources and relay nodes will be fixed during the entire transmission process.

VI. Conclusion

This paper studies cooperative energy harvesting communication systems with multiple source-to-destination pairs competing for a set of relay nodes with energy harvesting ability. We model the interaction between the sources and relay nodes as a college admission market and then fit this market into a stochastic environment. We propose an I-POMDP framework to analyze the interaction among the sources. The optimal policy for each source to decide the relay node that can maximize its long-term expected payoff has been derived. We present numerical results to verify the performance of our proposed policy.

Acknowledgment

Yong Xiao would like to thank Professor Leslie Pack Kaelbling and Dr. Christopher Amato for their helpful discussions in the early stage of this work.

References

[1] J. Yang and S. Ulukus, "Optimal packet scheduling in an energy harvesting communication system," *IEEE Trans. Commun.*, vol. 60, no. 1, pp. 220–230, Jan. 2012.

[2] K. Tutuncuoglu and A. Yener, "Optimum transmission policies for battery limited energy harvesting nodes," *IEEE Trans. Wireless Commun.*, vol. 11, no. 3, pp. 1180–1189, Mar. 2012.

[3] M. A. Antepli, E. Uysal-Biyikoglu, and H. Erkal, "Optimal packet scheduling on an energy harvesting broadcast link," *IEEE J. Sel. Areas in Commun.*, vol. 29, no. 8, pp. 1721–1731, Sep. 2011.

[4] 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, Apr. 2012.

[5] A. M. Fouladgar and O. Simeone, "On the transfer of information and energy in multi-user systems," *IEEE Communications Letters*, vol. 16, no. 11, pp. 1733–1736, Nov. 2012.

[6] B. Gurakan, O. Ozel, J. Yang, and S. Ulukus, "Energy cooperation in energy harvesting communications," *IEEE Trans. Commun.*, vol. 61, no. 12, pp. 4884–4898, Dec. 2013.

[7] A. Aprem, C. Murthy, and N. Mehta, "Transmit power control policies for energy harvesting sensors with retransmissions," *IEEE Journal of Selected Topics in Signal Processing*, vol. 7, no. 5, pp. 895–906, Oct 2013.

[8] D. Fudenberg and D. K. Levine, *The theory of learning in games*. MIT Press, Cambridge, MA, 1998.

[9] P. Gmytrasiewicz and P. Doshi, "A framework for sequential planning in multiagent settings," *Journal of Artificial Intelligence Research*, vol. 24, no. 1, pp. 49–79, Jul. 2005.

[10] A. R. Cassandra, "A survey of POMDP applications," in *Working Notes of AAAI 1998 Fall Symposium on Planning with Partially Observable Markov Decision Processes*, Oct. 1998.

[11] Y. Xiao, G. Bi, and D. Niyato, "Game theoretic analysis for spectrum sharing with multi-hop relaying," *IEEE Trans. Wireless Commun.*, vol. 10, no. 5, pp. 1527 –1537, May 2011.