Communication of Energy Harvesting Tags

Zhe Wang, Ali Tajer, *Member, IEEE*, and Xiaodong Wang, *Fellow, IEEE*

***Abstract*—We solve the problem of designing an affordable optimal transmission strategy for the recently proposed system of energy-harvesting active networked tags (EnHANTs), that is adapted to the identification request and the energy harvesting dynamic. We assume that the system operates in a time-slotted fashion, so that the problem is formulated as a Markov decision process (MDP). Both a static exhaustive search method and a modified policy iteration algorithm are employed to obtain the optimal transmission policy. Simulation results are provided to demonstrate that the obtained transmission policy can considerably improve the overall system performance which takes into consideration of both the system activity-time and the communication reliability.**

***Index Terms*—Energy harvesting tags, transmission strategy, Markov decision process, policy iteration.**

I. INTRODUCTION

THE system of energy-harvesting active networked tags (EnHANTs) has been recently proposed as small devices that can be attached to small objects that are not traditionally networked, e.g., books, clothes, and keys [1], [2]. The EnHANTs system represents a futuristic transition from the radio frequency identification (RFID) technology [3] to a novel one with two main features. First, it enables communications among tag-equipped objects and secondly, the objects are autonomous and self-sufficient from an energy consumption perspective as they harvest and store energy from ambient light, motion, and temperature gradients.

The EnHANTs system mainly facilitates object tracking applications that are not viable through the existing technologies that either lack networking capability (e.g., RFID) or do not satisfy the size or energy autonomy constraints (e.g., Bluetooth). Examples of such tracking applications by energy autonomous networked objects include disaster recovery, emergency alert, and collecting temporal and spatial proximity information. This system enjoys the main features of both the RFID and wireless sensor network (WSN) technologies. In particular, the tags are designed to provide a timely response to any request for their identification information, as done by RFIDs, and also to report their functioning states and surrounding environment information, as done in a WSN.

Paper approved by C.-L.Wang, the Editor for Equalization of the IEEE Communications Society. Manuscript received April 28, 2011; revised November 1, 2011.

Z. Wang is with the Electronic Engineering Department, Shanghai Jiao Tong University, Shanghai, 200240, and the Electrical Engineering Department, Columbia University, New York, 10027 (e-mail: zhewang@ee.columbia.edu).

X. Wang is with the Electrical Engineering Department, Columbia University, New York, 10027 (e-mail: wangx@ee.columbia.edu).

A. Tajer is with the Electrical Engineering Department, Princeton University, Princeton, NJ 08544 (e-mail: tajer@princeton.edu).

This work was supported in part by the U.S. National Science Foundation (NSF) under grant CCF-0964497.

Digital Object Identifier 10.1109/TCOMM.2012.022912.110298

The major challenge in designing the communication protocols for EnHANTs pertains to managing the energy resources. For such energy management, there exists a tension between maximizing the activity-time[1] of the tags on one hand, which necessitates a conservative consumption of the energy resources, and increasing the communication reliability on the other hand, which suggests consuming more energy. The optimal consumption of the energy resources, therefore, requires striking a balance between maximizing the activity-time and communication reliability. Maintaining such a balance becomes more complicated due to the fact that the tags harvest energy on an ad-hoc basis, depending on the physical conditions of the environment (e.g., light, temperature, or motion). Therefore, an object might not have adequate energy for responding to any communication request it receives, and more importantly, even if it does, it might not be necessarily optimal to respond to such a request as preserving the energy for subsequent communications might bring about more overall communication reliability and activity-time.

The problems of activity-time maximization and reliability maximization have been treated independently in the contexts of RFID and WSN, respectively. For example, [4] and [5] consider maximizing the activity-time and coverage range (readability) of the RFID tags, respectively. Specifically, [4] proposes a mechanism for jointly energy harvesting and energy saving and [5] introduces a passive RFID system whose tags are equipped with power amplifiers and energy storage devices. Both systems are designed for typical application of tag identification information reading and do not support state information exchange among the tags. On the other hand, [6]–[8] discuss energy optimization for the WSNs, where the optimal transmission schemes subject to the battery state and delay constraints are developed.

In this paper, we propose a transmission strategy for EnHANTs that optimizes a *long-term* average of the communication reliability. The reliability part of this objective reflects the impact of energy management on communications and the long-term average implicitly incorporates the activity-time maximization goal. We show that the energy-spending policy associated with the information transmission can be cast as a Markov decision process (MDP), and we provide an efficient algorithm for computing the optimal policy.

The remainder of the paper is organized as follows. In Section II, we describe the energy harvesting and communi-

[1]The meaning of lifetime for energy-harvesting tags is slightly different from that of more conventional tags. For this reason we have adopted the term "activity-time", which similar to the traditional definitions of lifetime, refers to the time spans during which the tag has enough energy to respond to the inquiries. Its difference, nevertheless, is that activity-time is not of finite-horizon and can potentially extend for a long duration given that the tag is capable of harvesting adequate energy.

cation models of the system. In Section III, we formulate the problem of optimizing the long-term average communication reliability, and show that it has an inherent MDP structure. In Section IV, we provide an iterative algorithm for solving the MDP problem and analyze its convergence. Simulation results are provided in Section V. Finally, Section VI concludes the paper.

II. System Descriptions

A. Communication Model

Consider a network of objects equipped with EnHANTs that communicate with a tag reader. Upon the request of the reader, the objects provide it with their identity and state information about their surrounding conditions. The communications occur in a time-slotted fashion with slots of equal durations. The beginning of a time slot is reserved for the reader to broadcast its inquiries for collecting identification and information. Upon receiving the inquiries, the objects promptly respond to the reader, where they are allowed to use the remaining portion of the time slot for transmitting their information to the reader. A communication error occurs when either the objects fail to respond to the reader's inquires, or the reader fails to correctly decode the data from the objects.

To ensure low energy consumption, we assume that the ultra wideband (UWB)-based the pulse-position modulation (PPM) [9] is employed at each tag for sending information to the reader. Specifically, the information is encoded to the different positions of a single pulse (or a group of pulses) within a given time interval T. Given an encoded PPM symbol

$$\boldsymbol{s} = [s_1, s_2, \ldots, s_J], \;\; s_i \in \{0, 1\} \;,$$

and a pulse $p(t)$ of duration T_p, where $T_p < T/J$, the received signal corresponding to $\boldsymbol{s}$ is given by

$$x(t) = \sum_{i=1}^{J} s_i p(t - iT/J) + v(t), \quad 0 \leq t \leq T \;, \qquad (1)$$

where $v(t)$ is the ambient Gaussian noise. We assume that all encoded symbols are mutually orthogonal. Assuming that the pulses in a symbol are all unit pulses, i.e., $\int p^2(t)dt = 1$, then we define the weight w of the symbol as the number of non-zero pulses in the symbol, which is also the energy of the symbol.

In order for the reader to process the received PPM signal from the tagged object, conventionally a front-end A/D converter is employed which requires a very high-sampling rate for the UWB PPM signal. In particular, the sampling rate is the inverse of the pulse width T_p, e.g., $1/T_p = 5\text{GHz}$, which is prohibitively high. Alternatively, given the sparsity of the PPM signal, the compressive sensing technique [10] together with the signal detection method with compressive measurements [11] can be employed at the reader to significantly reduce the sampling rate. The basic idea is to project the received UWB PPM signal to some (random) basis waveforms at the analog front-end. The resulting projections constitute the compressive measurements based on which the original transmitted PPM signal can be detected. Mathematically the projection operation is characterized by a (random) projection matrix $\boldsymbol{\Phi} \in \mathbb{R}^{M \times N}$ [10]. After projection, the original received PPM signal $\boldsymbol{x} \in \mathbb{R}^N$, corresponding to the samples of the received PPM waveform $x(t)$ at the $1/T_p$ sampling rate, is converted to the compressed samples $\tilde{\boldsymbol{x}} = \boldsymbol{\Phi}\boldsymbol{x} \in \mathbb{R}^M$ with a compression ratio of M/N. Note that no sampling at rate $1/T_p$ is needed; instead, we obtain the compressed samples $\tilde{\boldsymbol{x}}$ directly by the analog projection operation.

Assume that there are totally K PPM symbols $\boldsymbol{s}_1, \boldsymbol{s}_2, \ldots, \boldsymbol{s}_K$. Denote their corresponding projections as $\tilde{\boldsymbol{x}}_i = \boldsymbol{\Phi}\boldsymbol{x}_i, i = 1, 2, \ldots, K$. Then the receiver implements the following decision rule on the compressed signal to decide the PPM symbol that was transmitted:

$$\hat{i} = \arg \min_{1 \leq i \leq K} (\tilde{\boldsymbol{x}} - \tilde{\boldsymbol{x}}_i)^T \boldsymbol{\Psi} (\tilde{\boldsymbol{x}} - \tilde{\boldsymbol{x}}_i) \;, \qquad (2)$$

where $\boldsymbol{\Psi} = (\boldsymbol{\Phi}\boldsymbol{\Phi}^T)^{-1}$. Under this classification method, the probability of mis-detecting a symbol of weight w, denoted by $P_{\text{md}}(w)$, is well-approximated by [11]

$$P_{\text{md}}(w) = 1 - Q\left(-\sqrt{\frac{M}{N}} \frac{w}{\sigma^2}\right)^{K-1} \;, \qquad (3)$$

where σ^2 is the variance the additive white Gaussian noise, and $Q(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-t^2/2} dt$.

A good timer synchronization is necessary for the EnHANTs as they work on a time-slot basis. Designing the appropriate synchronizers follow the same principles as those needed in the more conventional RFID systems. The system architectures provided in [1] and [2] employ a simple schemes in which the tags and readers use an analog circuit to detect the reader's inquiries and the tag's responses. These inquires and responses occur in the forms of a single pulse or a train of pulses. By locating the positions of the pulses, the system can obtain the underlying time reference, which in turn serves as the basis for synchronized communication. More information on implementing these synchronization methods is available in [12] and [13].

B. Energy Harvesting Model

We assume that the reader has a passive and continuous power source and has no power constraint. For the tags, we assume that they are equipped with rechargeable batteries and light energy harvesting devices. Due to the size constraints, the batteries must be small and consequently, have low capacity. Therefore, a considerable portion of the energy consumed by the tags should be harvested from the environment and the battery essentially functions as an energy buffer.

We consider probabilistic models for inquiries made by the reader as well as the energy harvesting dynamics of the devices. We aim to optimize the transmission policy from the perspective of each object and therefore restrict the analysis to the case of one reader and one EnHANT equipped object. To model the identification request state of the reader at the beginning of the k-th time slot, for $k \in \mathbb{N}$, we define the random variable

$$a_k \sim \text{Bernoulli}(r) \;,$$

where $a_k = 1$, occurring with probability r, indicates that the reader inquires about the tag's information at the beginning of the k-th time-slot and $a_k = 0$, that occurs with probability

$(1-r)$, indicates otherwise. We also define the indicator b_k to reflect whether the tag is harvesting energy in the k-th time slot ($b_k = 1$) or it is not harvesting energy ($b_k = 0$). Moreover, we model the energy harvesting process as a correlated, two-state process [14]. If the tag harvests energy in a time slot, it will continue to harvest energy in the subsequent time slot with probability p and if no energy is harvested in a time slot, the probability of not harvesting any energy in the subsequent time slot either is q.

We denote the energy level that a tag can harvest and consume in the subsequent time slots by E_h. We also denote the capacity of the battery by $B_{\max}$ and denote the energy level restored in the battery of the object at the beginning of the k-th time slot by B_k, with $B_k \leq B_{\max}$. By defining W_k as the weight of the symbol transmitted in the k-th time slots, we get the following recursive relationship between the energy levels at the beginning of two consecutive time slots

$$B_{k+1} = \min\left\{B_k - a_k \cdot W_k \cdot \mathbf{1}_{\{B_k \geq W_k\}} + b_k \cdot E_h \ , \ B_{\max}\right\} \ , \tag{4}$$

where the indicator function $\mathbf{1}_{\{A\}}$ is defined as $\mathbf{1}_{\{A\}} = 1$ if A is true, and 0 otherwise. We remark that W_k, for all $k \in \mathbb{N}$, take discrete values from $\mathcal{W} = \{0, w_1, \ldots, w_m\}$ which are determined by the design of the hardware.

III. Problem Statement

A. Performance Measure

We define $S_k \triangleq (B_k, a_k, b_k)$, as the *state* of the tag in the k-th time slot. Since all components of S_k, i.e., B_k, a_k, and b_k, take discrete values and are all bounded, there are a finite number of possible states. We denote the number of such possible states by $|\mathcal{S}|$ and the set of possible states by $\mathcal{S} \triangleq \{s_1, \ldots, s_{|\mathcal{S}|}\}$. Due to the structure of PPM that encodes the data in the positions of the non-zero pulses, the data to be transmitted govern the positions of the pulses, and the state of the tag determines the energy of the pulse. As a result, irrespective of the data content to be conveyed to the reader, the energy of the tag in the k-th time interval is uniquely determined by S_k. Therefore, identical states $S_k = S_l$ for $k \neq l$ will give rise to identical symbol weights. We denote a transmission policy ϕ as a mapping from the set of states $\mathcal{S}$ to the set of weights $\mathcal{W}$, so that $\phi(s_k)$ is the symbol weight corresponding to the state s_k. Our objective is to determine the optimal design of $\phi(\cdot)$ such that a performance measure, that incorporates both the tag activity-time and communication reliability, is optimized.

Erroneous communication has two origins, namely *no-response errors* and *mis-detection errors*. The no-response error in the k-th time slot occurs when the battery cannot afford the energy required for sending a response to the reader, i.e., $B_k < W_k$, or when the tag operates under a certain policy that may voluntarily give up responding to the reader's request. For this reason, in order to allow for the possibility of letting $\phi(s_k) = 0$, we must have $0 \in \mathcal{W}$. For any given transmission policy ϕ, these two factors combined give rise to the following long-term average no-response error, where the average is taken over all time-slots,

$$\hat{P}_{\rm nr}(\phi) = \lim_{N\to\infty} \frac{\sum_{k=1}^{N} \mathbf{1}_{\{W_k=0\}} \cdot a_k}{\sum_{k=1}^{N} a_k} \ , \tag{5}$$

where $W_k = \phi(S_k)$. The mis-detection errors take place when the reader cannot successfully decode the data transmitted by the object and its pertinent infinite-horizon average error for the given transmission policy ϕ is

$$\hat{P}_{\rm md}(\phi) = \lim_{N\to\infty} \frac{\sum_{k=1}^{N} P_{\rm md}(W_k) \cdot \mathbf{1}_{\{W_k>0\}} \cdot a_k}{\sum_{k=1}^{N} a_k} \ , \tag{6}$$

where $W_k = \phi(S_k)$. Finally, in order to incorporate the no-response and mis-detection error probabilities under the same performance measure, we define a weighed average of the two error probabilities as

$$P_{\rm err}(\phi) = \beta \hat{P}_{\rm nr}(\phi) + (1-\beta)\hat{P}_{\rm md}(\phi) \ , \tag{7}$$

where $\beta \in [0, 1]$ is the weighting factor. By changing β one can adjust the error probability $P_{\rm err}(\phi)$ based on the application of interest depending on whether the no-response or mis-detection error is more important. Equations (5)-(7) provide the equation (8)

Therefore, the optimization problem that we strive to solve can be formalized as follows:

$$\mathcal{P} = \begin{cases} \min_\phi & P_{\rm err}(\phi) \\ \text{s.t.} & \text{the battery states satisfy (4)} \end{cases} \ . \tag{9}$$

B. Markov Decision Process

The optimization problem as formulated in (9) designs the optimal policy ϕ, which is valid throughout the activity-time of the tag. In other words, the solution we have is stationary in the sense that it does not change over time. This means that we can solve (9) offline and provide the tags with the corresponding look-up tables, without requiring them to spend their energy resources on computations. We next show that the optimization problem that finds a stationary policy, which is the mapping from the states in $\mathcal{S}$ to the weights in $\mathcal{W}$, can be modeled as a standard Markov decision process (MDP) problem.

A standard MDP, which provides a framework for decision-making in situations where outcomes are partly random, can be defined via a quadruplet $\big(\mathcal{S}, \mathcal{W}, p_{w_i}(s_i, s_j), R_{w_i}(s_i, s_j)\big)$, where in our settings $\mathcal{S}$ denotes the set of states S; $\mathcal{W}$ is the set of actions taken based on the states, i.e., the set of weights assigned to the states; $p_{w_i}(s_i, s_j)$ denotes the probability of transition from state s_i to state s_j when action $w_i \in \mathcal{W}$ is taken. Note that the transition probabilities satisfy

$$\forall w_i \in \mathcal{W}, \forall i \in \{1, 2, \ldots, M\} : \sum_{j=1} p_{w_i}(s_i, s_j) = 1 \ .$$

Finally, $R_{w_i}(s_i, s_j)$ denotes the penalty (or reward) associated with the transition from s_i to s_j under action w_i. The objective of an MDP is to choose a policy $\phi : \mathcal{S} \to \mathcal{W}$ that assigns an action to each state such that the average penalty is minimized. Specifically, the policy of interest minimizes the infinite horizon penalty

$$R_{ih}(\phi) \triangleq \lim_{N\to\infty} \frac{1}{N} \sum_{k=1}^{N} R_{W_k}(S_k, S_{k+1}) \ , \tag{10}$$

where $W_k = \phi(S_k)$.

$$P_{\mathrm{err}}(\phi) = \lim_{N\to\infty} \frac{\sum_{k=1}^{N} \left(\beta \cdot \mathbf{1}_{\{W_k=0\}} + (1-\beta)\cdot P_{\mathrm{md}}(W_k)\cdot \mathbf{1}_{\{W_k>0\}}\right)\cdot a_k}{\sum_{k=1}^{N} a_k} \ . \quad (8)$$

C. The MDP Formulation

As described in Section II, the reader operates in a time-slotted fashion with slots of equal durations. During the k-th time slot, the tag selects the symbol weight $W_k = \phi(S_k)$ for the signal to be sent to the reader. Under the choice of W_k the tag's state changes from S_k to S_{k+1}, as shown in Fig. 1. Therefore, the penalty associated with this transition is $R_{W_k}(S_k, S_{k+1})$.

A natural choice for the penalty term $R_{W_k}(S_k, S_{k+1})$ is the probability that such a transition is sensed by the reader erroneously. In particular, we aim to associate $R_{W_k}(S_k, S_{k+1})$ with the communication error probability given by

$$R_{W_k}(S_k, S_{k+1}) = \left(\beta \cdot \mathbf{1}_{\{W_k=0\}} + (1-\beta)\cdot P_{\mathrm{md}}(W_k)\cdot \mathbf{1}_{\{W_k>0\}}\right)\cdot a_k \ . \quad (11)$$

By invoking (10), the infinite-horizon penalty becomes

$$R_{ih}(\phi) = \lim_{N\to\infty} \frac{1}{N}\sum_{k=1}^{N} \left(\beta \cdot \mathbf{1}_{\{W_k=0\}} + (1-\beta)\cdot P_{\mathrm{md}}(W_k)\cdot \mathbf{1}_{\{W_k>0\}}\right)\cdot a_k. \quad (12)$$

By comparing (12) and $P_{err}(\phi)$ in (8), we find that $R_{ih}(\phi)$ and $P_{err}(\phi)$ are identical up to a scaling factor. This scaling factor is $\lim_{N\to\infty} \sum_{K=1}^{N} a_k/N$, which by considering the distribution of a_i and the law of large numbers is equal to Nr. Therefore, the optimal weight assignment policy ϕ, which is the solution to (9) can be equivalently found by solving the following the MDP problem

$$\hat{\mathcal{P}} = \left\{ \begin{array}{ll} \min_\phi & R_{ih}(\phi) \\ \text{s.t.} & \text{the battery states satisfy (4)} \end{array} \right. . \quad (13)$$

This statement is formalized in the following proposition.

Proposition 1: The solution to the optimization problem in (9) can be obtained by solving the MDP problem in (13).

IV. Computing the Optimal Transmission Policy

In this section we discuss how to solve (9). We denote the set of all possible transmission policies as $\Phi = \{\phi : \mathcal{S} \to \mathcal{W}\}$. Then we have $|\Phi| = |\mathcal{W}|^{|\mathcal{S}|}$.

We first consider a naive exhaustive search method. Assuming the MDP process starts form the 0-th time slot and is continuously observed for N time slots, we can simulate the $\{a_0, a_1, \ldots, a_N\}$, which is the sequence of the identification request state of each time slot, and $\{b_0, b_1, \ldots, b_N\}$, which is the sequence of the energy harvesting state of each time slot, based on their respective underlying statistical models. Based on the battery state transition process in (4), a finite-horizon state sequence $\mathcal{S}(\phi) = \{S_0, S_1, \ldots, S_N\}$ can then be generated under each possible policy $\phi \in \Phi$.

Using (8) for finite N, we can calculate the average penalty associated with the state-sequence $\mathcal{S}(\phi)$, which we denote as $P_{\mathrm{err}}(\phi)$. The optimal policy is then

$$\phi^* = \arg\min_{\phi\in\Phi} P_{\mathrm{err}}(\phi) \ . \quad (14)$$

Obviously if we choose the sequence length N to be large enough, ϕ^* can be considered as a close approximation to the solution to the original problem in (9).

A. Modified Policy Iteration Algorithm

The complexity of the exhaustive search method becomes prohibitive when $|\mathcal{W}|$ or $|\mathcal{S}|$ is large. We next apply the *modified policy iteration* (MPI) algorithm [17] to compute the optimal transmission policy. The basic idea is to iterate the policy search process until an iteration variable converges. This variable is calculated in each iteration by another value iteration process.

All iterations in the MPI algorithm are based on the state transition probabilities. According to the state definition and the battery state transition process in (4), at any state S_k an action w_k leads to a transition to the following four possible next state S_{k+1}, $S^1_{k+1} = (B_{k+1}, 0, 0)$, $S^2_{k+1} = (B_{k+1}, 0, 1)$, $S^3_{k+1} = (B_{k+1}, 1, 0)$, and $S^4_{k+1} = (B_{k+1}, 1, 1)$, where

$$B_{k+1} = \min\{B_k + E_h b_k - w_k a_k, B_{\max}\} \ . \quad (15)$$

The transition probabilities from S_k to S^j_{k+1}, $j = 1, 2, 3, 4$, depend on the state S_k and the system parameters r, p and q. Assuming that the current state is $S_k = (B_k, a_k, b_k)$ and the next state is $S_{k+1} = (B_{k+1}, a_{k+1}, b_{k+1})$, when the action $w_k = \phi(S_k)$ is taken, we have

$$p_{w_k}(S_k, S_{k+1}) = p(a_{k+1})p(b_{k+1} \mid b_k) \ . \quad (16)$$

The transition probabilities are summarized in Table I.

The MPI algorithm consists of two phases, policy improvement and partial policy evaluation. In the policy improvement phase, the algorithm searches for a policy based on the iteration variable, the current *penalty iteration value*. Specifically, at the n-th iteration, we have the iteration variables $v^{(n-1)}(s_i),\ s_i \in \mathcal{S}$, which are the penalty iteration values corresponding to different states calculated in the previous iteration (We set $v^{(0)}(s_i) = 0,\ s_i \in \mathcal{S}$). Denote

$$\boldsymbol{v}^{(n)} \triangleq \left[v^{(n)}(s_1), v^{(n)}(s_2), \ldots . v^{(n)}(s_{|\mathcal{S}|})\right] ,$$

and the equation (17).

Then the policy $\phi^{(n)}$ at this iteration is computed as

$$\phi^{(n)}(s) = \arg\min_{w\in\mathcal{W}} f(s, w, \boldsymbol{v}^{(n-1)}),\ s \in \mathcal{S} \ . \quad (18)$$

And the penalty iteration value is updated as

$$v^{(n)}(s) = f(s, \phi^{(n)}(s), \boldsymbol{v}^{(n-1)}),\ s \in \mathcal{S} \ . \quad (19)$$

In the partial policy evaluation phase, the algorithm determines whether $\phi^{(n)}$ found in the policy improvement phase

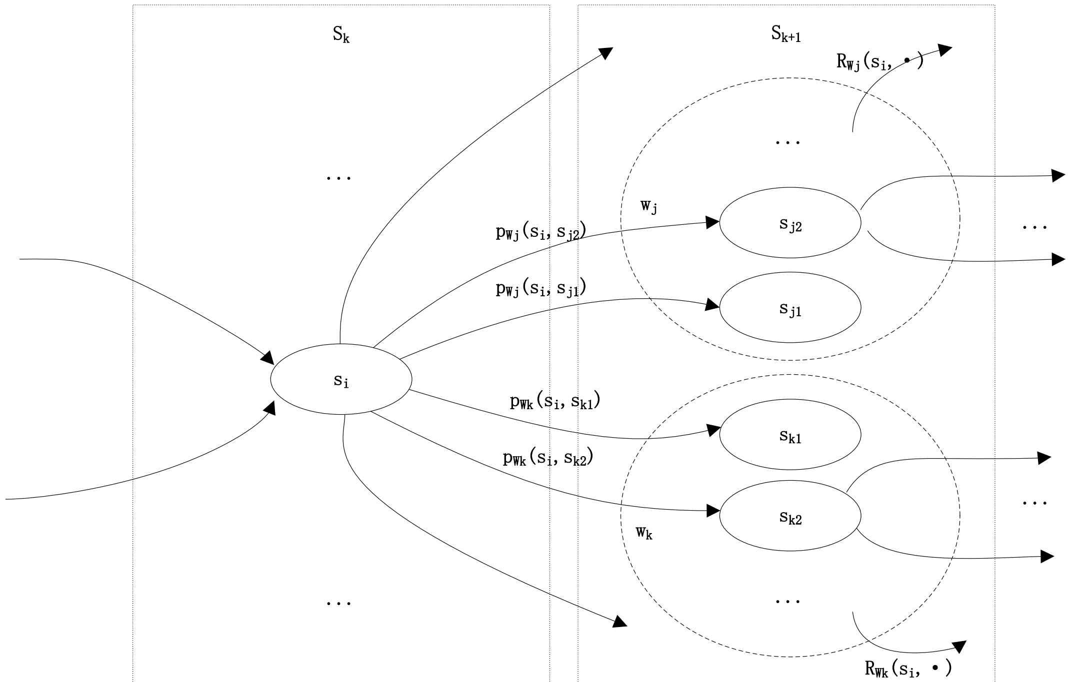

Fig. 1. The state transition diagram.

TABLE I
STATE TRANSITION PROBABILITIES $p_{w_k}(S_k, S_{k+1})$.

	$S_k=(B_k,0,0)$	$S_k=(B_k,0,1)$	$S_k=(B_k,1,0)$	$S_k=(B_k,1,1)$
$S^1_{k+1}=(B_{k+1},0,0)$	$(1-r)q$	$(1-r)(1-p)$	$(1-r)q$	$(1-r)(1-p)$
$S^2_{k+1}=(B_{k+1},0,1)$	$(1-r)(1-q)$	$(1-r)p$	$(1-r)(1-q)$	$(1-r)p$
$S^3_{k+1}=(B_{k+1},1,0)$	rq	$r(1-p)$	rq	$r(1-p)$
$S^4_{k+1}=(B_{k+1},1,1)$	$r(1-q)$	rp	$r(1-q)$	rp

$$f(S_k, w, \boldsymbol{v}^{(n-1)}) = \sum_{j=1}^{4} p_w(S_k, S^j_{k+1})\Big(R_w(S_k, S^j_{k+1}) + v^{(n-1)}(S^j_{k+1})\Big),\ S_k \in \mathcal{S},\ w \in \mathcal{W}\ . \tag{17}$$

is the overall optimal policy. If not, the algorithm starts a sub-iteration process to update the penalty iteration values $v^{(n)}(s_i)$ and then goes back to the policy improvement phase for another iteration. In order to determine whether $\phi^{(n)}$ is the optimal policy, we compute

$$u^{(1)}(s) \triangleq f(s, \phi^{(n)}(s), \boldsymbol{v}^{(n)}),\ s \in S\ . \tag{20}$$

Denote $\boldsymbol{u}^{(n)} = [u^{(n)}(s_1), u^{(n)}(s_2), \ldots, u^{(n)}(s_{|\mathcal{S}|})]$. Given a small value ϵ, if

$$\|\boldsymbol{u}^{(1)} - \boldsymbol{v}^{(n)}\| < \epsilon\ , \tag{21}$$

then we consider $\phi^{(n)}$ as the overall optimal policy ϕ^*. Otherwise, we perform the following iteration to update the penalty iteration value,

$$u^{(m)}(s) = f(s, \phi^{(n)}(s), \boldsymbol{u}^{(m-1)}),\ s \in \mathcal{S},\ m = 1, 2, \ldots, M\ . \tag{22}$$

Finally we set $v^{(n)}(s_i) = u^{(M)}(s_i),\ s_i \in \mathcal{S}$ and go back to the policy improvement phase for another iteration.

The MPI algorithm for solving the MDP problem in (13) is summarized as follows.

Algorithm - Modified Policy Iteration Algorithm for Solving (13)

1: Initialization
 $\boldsymbol{v}^{(0)} = \mathbf{0}$
 $n = 1$
2: Policy Improvement
 FOR $s \in \mathcal{S}$
  $\phi^{(n)}(s) = \arg\min_{w \in \mathcal{W}} f(s, w, \boldsymbol{v}^{(n-1)})$
  $v^{(n)}(s) = f(s, \phi^{(n)}(s), \boldsymbol{v}^{(n-1)})$
 ENDFOR
3: Partial Policy Evaluation
 FOR $s \in \mathcal{S}$
  $u^{(1)}(s) = f(S_k, \phi^{(n)}(s), \boldsymbol{v}^{(n)})$
 IF $\|(u^{(1)} - v^{(n+1)}\| < \epsilon$, **GOTO STEP 4**
 ELSE FOR $m = 1, 2, \ldots, M$
  FOR $s \in \mathcal{S}$
   $u^{(m)}(s) = f(s, \phi^{(n)}(s), \boldsymbol{u}^{(m-1)})$
  ENDFOR
 ENDFOR, ENDIF
 $\boldsymbol{v}^{(n)} = \boldsymbol{u}^{(M)}$
 $n \leftarrow n + 1$, **GOTO STEP 2**
4: Choose Policy
 $\phi^* = \phi^{(n)}$

TABLE II
SYMBOL MIS-DETECTION PROBABILITIES FOR DIFFERENT SYMBOL WEIGHTS.

w	0	1	2	4
$P_{\mathrm{md}}(w)$	1.0000	0.6874	0.1625	0.0054

This algorithm combines the features of both policy iteration and value iteration. The most significant feature is its low computational complexity, compared to the exhaustive search. On the other hand, as will be shown in Section V, its performance is similar to that of the exhaustive search.

B. Convergence of the MPI Algorithm

The MPI algorithm is designed for solving a class of MDP problems that has finite state space, finite decision space, non-discount average reward, and infinite-horizon [17]. Obviously, the MDP problem in (13) belongs to this class.

A sufficient condition for the MPI algorithm to converge is given in [17]. In particular, if

$$\min_{\phi_1,\phi_2\in\Phi}\ \min_{(u,v)\in S\times S}\sum_{j\in S}\min\left\{p^J_{\phi_1}(j\mid u),p^J_{\phi_2}(j\mid v)\right\}>0\ , \tag{23}$$

where $p^J_\phi(j\mid u)$ is the transition probability from $u\in\mathcal{S}$ to $j\in\mathcal{S}$ after J state transitions under the policy ϕ, then the optimal policy can be found by the MPI algorithm within a finite number of iterations. For our problem, we make the reasonable assumption that the rates of both energy harvesting and identification request are positive, i.e., $0<p,q<1$ and $0<r<1$, and the tag will be silent if there is no request, i.e., $a_k=0$. Then we have the following convergence result.

Proposition 2: The ϵ-optimal solution to the MDP problem in (13) can be obtained by performing the MPI algorithm within a finite number of iterations.

Proof: According to (4) and the system state definition S, if $0<p,q<1$ and $0<r<1$, and the tag consumes no energy when there is no request, for any stationary policy $\forall s\in\mathcal{S}$ can transit to $s'=(B_{\max},0,1)$ within a finite number of transitions, i.e., $p^j_\phi(s'\mid s)>0$. Therefore, (23) is satisfied and the MDP problem in (13) can be solved by the MPI algorithm. ■

V. SIMULATION RESULTS

We assume that each response message from the tag is encoded using 15 bits, transmitted in 3 PPM symbols with the symbol modulation order of $K=32$. Each symbol may contain 1, 2 or 4 non-zero pulses, i.e., $\mathcal{W}=\{0,1,2,4\}$. The durations of the pulse and the symbol are $T_p=5ns$ and $T=6.4\mu s$ respectively. Assuming a compression ratio of $M/N=0.1$, we obtain the compressed signal samples at the affordable rate of 200MHz. The symbol mis-detection probabilities $P_{\mathrm{md}}(w)$ corresponding to different symbol weights are calculated using (3) and given in Table II, for the pulse SNR $\frac{1}{\sigma^2}=6dB$. Furthermore, we set the battery capacity $B_{\max}=10$ and the error weight parameter $\beta=0.5$.

For the purpose of performance comparison, we consider two simple transmission strategies, a *conservative policy* and a *greedy policy*. The conservative policy always chooses the minimum available energy $w\in\mathcal{W}$ to transmit the response, such that the probability of no-response is minimized. On the other hand, the greedy policy targets for the best detection performance and always chooses the maximum available $w\in\mathcal{W}$ for responding to the reader's inquiry. For each simulation, the number of simulated time slots is $N=10^6$. The convergence threshold of the MPI algorithm is $\epsilon=10^{-5}$.

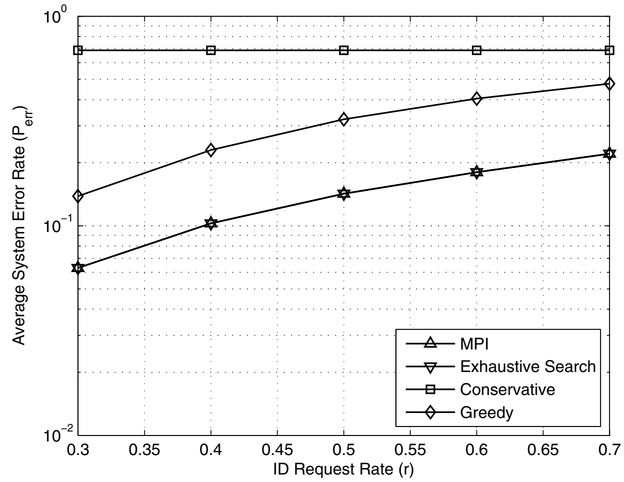

Fig. 2. Performance comparisons for the energy-balanced scenario.

We first consider an energy-balanced scenario where the energy harvesting parameters are $p=q=0.5$ and $E_h=3$. Under such a condition, the battery is neither empty nor full in most time slots. The battery acts as an energy buffer and the scheduling algorithm pursues the best trade-off between the mis-detection errors and the no-response errors. The simulation results for this scenario are shown in Fig. 2. For the second scenario, we consider an energy-deficient environment, where $p=0.3$, $q=0.7$ and $E_h=3$, corresponding to the case that a tag has a small probability to obtain energy from its environment at any time slot. In this case, the battery is empty in most time slots and the scheduling algorithm is apt to trade the detection performance for the activity-time. The simulation results for this scenario are shown in Fig. 3. In the last scenario, we simulate the policies in the energy-overflow environment, where $p=0.7$, $q=0.3$ and $E_h=6$. This environment ensures that the tags are strongly capable of being over-charged in most time slots. So the scheduling algorithm is apt to spend more energy to reduce the mis-detection errors. The simulation results for this scenario are shown in Fig. 4. The optimality of the proposed scheme relies on the knowledge about p and q, which in practice might not be known accurately. Further simulations on the sensitivity of the performance on the design values p and q are demonstrated in Fig. 5. This figures shows that perturbations the values of p and q by as much as 30% imposes only negligible performance losses.

It is seen from Fig. 2–4 that the optimal polices based on the MPI algorithm and exhaustive search give the best performance for all three scenarios. And the conservative policy gives the worst performance. Moreover, the greedy

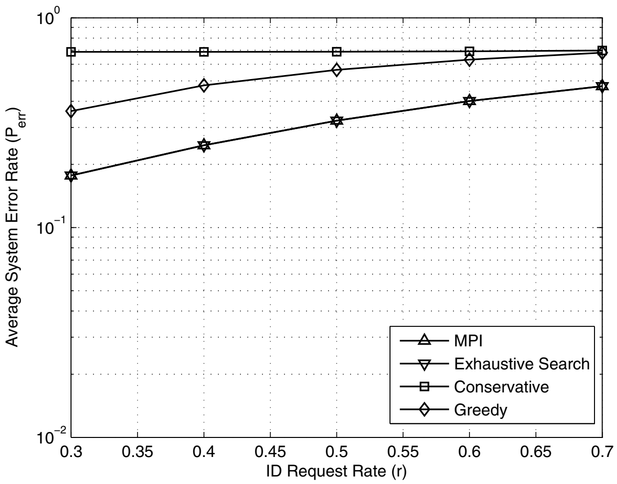

Fig. 3. Performance comparisons for the energy-deficient scenario.

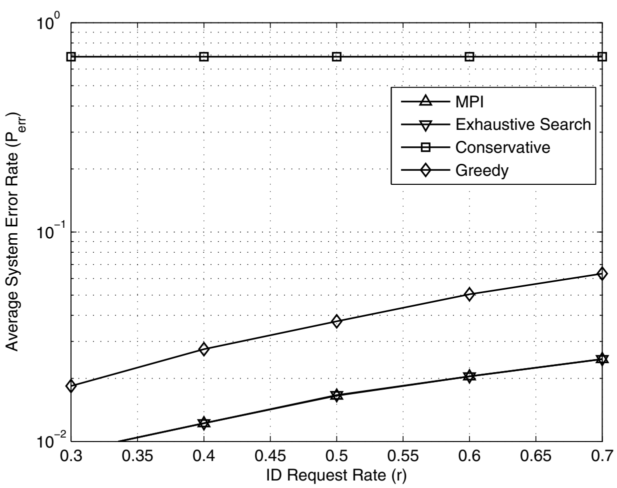

Fig. 4. Performance comparisons for the energy-overflow scenario.

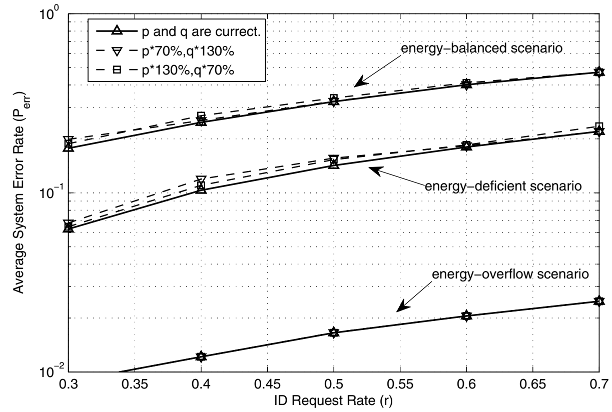

Fig. 5. Performance comparisons for inaccurate p and q.

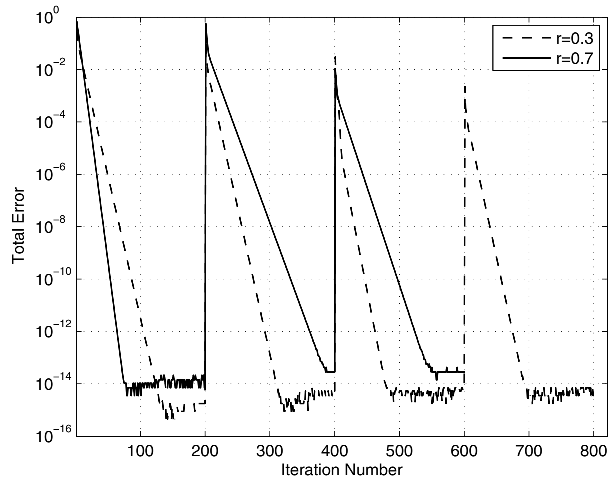

Fig. 6. The convergence of the MPI algorithm under the energy-balanced scenario.

policy performs worse than the optimal policy because it fails to balance the mis-detection errors and the no-response errors by simply ignoring the latter. On the other hand, as the identification request rate increases, the performances of both the optimal policy and the greedy policy degrade due to the energy constraints. Another observation is that when the tag's energy harvesting capability becomes stronger, the performances of all these policies improve, since the tag can use more energy to reduce the no-response errors and to improve the detection performance. In addition, by comparing the optimal transmission policies found by the MPI algorithm and the exhaustive search method, it is observed that their performance conform precisely in most simulation scenarios and there exist slight discrepancy in rare situations.

Fig. 6 shows the MPI algorithm's convergence under the energy-balanced scenario. The number of iterations for the partial policy evaluation phase is $M = 200$. It is seen that the optimal policies are obtained at the 4-th and 5-th policy improvement iterations for $r = 0.7$ and $r = 0.3$, respectively. At each policy improvement iteration, the policy is updated based on the current penalty values, which are converged in the previous evaluation phase. Then, the penalty values are updated for the updated policy. Also, at the last policy improvement iteration, the total error is below the threshold ϵ and the algorithm stops.

VI. Conclusions

We have formed a system model for the recently proposed system of energy-harvesting active networked tags (EnHANTs), including the communication model and the energy harvesting model, where the events of identification request and energy harvesting are assumed to follow simple Markov processes. A typical application of the EnHANTs system is for the tags to respond to the request by sending some simple information about their own identifications and their surrounding environment. For such an application, we formulate the problem of optimizing the transmission policy to maximize both the reliability and activity-time of the system. We have shown that the optimization problem has an inherent MDP structure and therefore can be solved using the modified policy iteration method. Finally simulation studies have demonstrated the effectiveness of the proposed optimal transmission policy in terms of making efficient use of the limited energy to improve the system reliability.

References

[1] M. Gorlatova, P. Kinget, I. Kymissis, D. Rubenstein, X. Wang, and G. Zussman, "Challenge: ultra-low-power energy-harvesting active networked tags (EnHANTs)," in *Proc. 2009 ACM Mobile Telecommum. Conf.*.

[2] M. Gorlatova, P. Kinget, I. Kymissis, D. Rubenstein, X. Wang, and G. Zussman, "Energy harvesting active networked tags (EnHANTs) for ubiquitous object networking," *IEEE Trans. Wireless Commun.*, vol. 17, no. 6, pp. 18–25, Dec. 2010.

[3] V. Chawla and S. H. Dong, "An overview of passive RFID," *IEEE Commun. Mag.*, vol. 45, no. 9, pp. 11–17, Sep. 2007.

[4] A. Janek, C. Steger, R. Weiss, J. Preishuber-Pfluegl, and M. Pistauer, "Lifetime extension of semi-passive UHF RFID tags using special power management techniques and energy harvesting devices," in *Proc. 2007 IEEE Africon Conf.*, pp. 1–7.

[5] F. Iannello, O. Simeone, and U. Spagnolini, "Energy management policies for passive RFID sensors with RF-Energy harvesting," in *Proc. 2010 IEEE Int'l Conf. Commun.*, pp. 1–7.

[6] K. Yang, and X. Wang, "Battery-aware adaptive modulation based on large-scale MDP," *IEEE Trans. Wireless Commun.*, vol. 7, no. 1, pp. 72–77, Jan. 2008.

[7] K. Yang and X. Wang, "Battery-aware adaptive modulation with QoS constraints," *IEEE Trans. Commun.*, vol. 54, no. 10, pp. 1797–1805, Oct. 2006.

[8] V. Sharma, U. Mukherji, V. Joseph, and S. Gupta, "Optimal energy management policies for energy harvesting sensor nodes," *IEEE Trans. Wireless Commun.*, vol. 9, no. 4, pp. 1326–1336, Apr. 2010.

[9] L. Yang and G. B. Giannakis, "Ultra-wideband communications: an idea whose time has come," *IEEE Signal Process. Mag.*, vol. 21, no. 6, pp. 26–54, Nov. 2004.

[10] D. L. Donoho, "Compressed sensing," *IEEE Trans. Inf. Theory*, vol. 52, no. 4, pp. 1289–1306, Apr. 2006.

[11] M. A. Davenport, P. T. Boufounos, M. B. Wakin, and R. G. Baraniuk, "Signal processing with compressive measurements," *IEEE J. Sel. Topics Signal Process.*, vol. 4, no. 2, pp. 445–460, Apr. 2010.

[12] C. Carbonelli and U. Mengali, "M-PPM noncoherent receivers for UWB applications," *IEEE Trans. Wireless Commun.*, vol. 5, no. 8, pp. 2285–2294, Aug. 2006.

[13] C. Hu, R. Khanna, J. Nejedlo, K. Hu, H. Liu, and P.Y. Chiang, "A 90 nm-CMOS, 500 Mbps, 35 GHz fully-integrated IR-UWB transceiver with multipath equalization using pulse injection-locking for receiver phase synchronization," *IEEE J. Solid-State Circuits*, vol. 46, no. 5, pp. 1076–1088, May 2011.

[14] A. Seyedi and B. Sikdar, "Performance modeling of transmission schedulers for sensor networks capable of energy harvesting," in *Proc. 2010 IEEE Int'l Conf. Commun.*, pp. 1–5.

[15] H. Li and B. Sikdar, "Relay usage scheduling in sensor networks with energy harvesting," in *Proc. 2010 IEEE Int'l Conf. Commun.*, pp. 1–5.

[16] W. J. Cook, W. H. Cunningham, W. R. Pulleybland, and A. Schrijver, *Combinatorial Optimization*. Wiley-Iterscience, 1998.

[17] M. L. Puterman, *Markov Decision Processes: Discrete Stochastic Dynamic Programming*. Wiley-Interscience, 2005.

Zhe Wang received the B.S. degree in Electrical Engineering and Business Administration (minor) from Shanghai Jiao Tong University, Shanghai, China in 2007, and the M.S. degree in Electrical Engineering from Shanghai Jiao Tong University, Shanghai, China in 2009. He is now pursuing his Ph.D. degree from Shanghai Jiao Tong University, Shanghai, China and currently studying at Columbia University, New York, NY, US. His research interests are in the areas of the modulation techniques and system optimization for wireless communications.

Ali Tajer received the B.Sc. and M.Sc. degrees in electrical engineering from Sharif University of Technology, Tehran, Iran, in 2002 and 2004, respectively, and the M.Phil. and Ph.D. degrees in electrical engineering from Columbia University, New York, NY in 2010. He is currently a Postdoctoral Research Associate in the Department of Electrical Engineering, Princeton University, Princeton, NJ, and an Adjunct Assistant Professor in the Department of Electrical Engineering, Columbia University, New York, NY. He spent the summers 2008 and 2009 as a Research Assistant with NEC Laboratories America, Princeton, NJ. His research interests lie in the general areas of network information theory and statistical signal processing with applications in wireless communication.

Xiaodong Wang (S'98-M'98-SM'04-F'08) received the Ph.D. degree in electrical engineering from Princeton University. He is a Professor of Electrical Engineering at Columbia University in New York. Dr. Wang's research interests fall in the general areas of computing, signal processing, and communications, and he has published extensively in these areas. Among his publications is a book entitled "Wireless Communication Systems: Advanced Techniques for Signal Reception," published by Prentice Hall in 2003. His current research interests include wireless communications, statistical signal processing, and genomic signal processing. Dr. Wang received the 1999 NSF CAREER Award, the 2001 IEEE Communications Society and Information Theory Society Joint Paper Award, and the 2011 IEEE Communication Society Award for Outstanding Paper on New Communication Topics. He has served as an Associate Editor for the IEEE TRANSACTIONS ON COMMUNICATIONS, the IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS, the IEEE TRANSACTIONS ON SIGNAL PROCESSING, and the IEEE TRANSACTIONS ON INFORMATION THEORY. He is a Fellow of the IEEE and listed as an ISI Highly-cited Author.