Transmission Policies for Energy Harvesting Sensors Based on Markov Chain Energy Supply

Wenxiang Zhu, Pingping Xu,
Maozong Zheng, Guilu Wu
National Mobile Communications Research
Laboratory,Southeast University
Nanjing, China
{zwx,xpp}@seu.edu.cn

Honglei Wang
School of Information and Electrical Engineering
Xuzhou College of Industrial Technology
Xuzhou, China
wanghl@mail.xzcit.cn

ABSTRACT

Due to the small energy harvesting rates and stochastic energy harvesting processes, energy management of energy harvesting senor is still crucial for body network. Transmission polices for energy harvesting sensors with Markov chain energy supply over time varying channels is formulated as an infinite discounted reward Markov Decision Problem under the assumption of geometric distribution of sensors' lifetime. In this paper, we firstly propose a low-storage transmission policy based on probability of successful transmission for body network. Then we narrow the feasible region of parameters in our policies from the real domain to a discrete set with limited number, which makes the method of combing optimal equations and enumeration algorithm to obtain optimal parameters workable. Finally, numerical results show that our presented transmission policies can achieve a good approximated performance of optimal policies, which can be derived by policy iteration algorithm. Compared with the optimal policies, our presented policies has the advantage of low storage.

Keywords

Energy Harvesting, Wireless Sensor Networks, Energy Management, Markov Decision Process

1. INTRODUCTION

With the development of energy harvesting techniques, harvesting energy from the environment has received increasing attention in wireless sensor networks. Wireless sensor networks with energy harvesting nodes can operate for long periods. However, due to the low replenishing rate and the stochastic energy renewal process, energy management is still crucial [1].

Many previous works have been conducted in the optimization of a point-to-point energy harvesting communication system in which the transmitter is equipped with an energy harvesting device. They can be categorized into two classes based on the assumption of whether or not the source knows the information about the energy and data arrival processes. The review about optimization in energy harvesting communication system has been conducted in [2]. Our paper mainly focuses on online scheduling.

Online scheduling assumes that the source contains statistical information about the energy harvesting processes[3]-[7]. In [3], a threshold algorithm is proposed for utilizing available energy to transmit packets with different rewards based on a memoryless Markov chain for battery state, and time correlation in the energy supply is introduced in [4]. In [5], energy management polices for energy harvesting sensor nodes to maximize the date rate while the data queue stays stable are proposed. An online heuristic policy[6] that performs closely to the online optimal is proposed for the finite horizon optimal packet scheduling problem. In [7], a Markov Decision Problem(MDP) is formulated to maximize the number of successfully delivered packets per time slot for an energy harvesting node with infinite capacity. It is proved that the optimal policy is a threshold policy depending on the state of the channel and the length of the energy queue.

However, on one hand, it is not possible to know the noncausal information before transmitting; on the other hand, the parameters of the underlying stochastic processes change over time in many practical systems. So, neither offline scheduling or online scheduling will be applicable. In [8], a learning approach is proposed to learn the optimal transmission policy that maximizes the expected sum of the data transmitted during the transmitter's life over a fixed channel.

In our paper, we consider a point-to-point communication system in a time-slot fashion with time-correlated energy supply, which is similar to [4]. However, in [4], the model did not consider the impact of limited source lifetime. In our paper, we model the lifetime of the source as a random variable, which means the source can stop its operation at any time slot with fixed probability. At the beginning of each time slot, the source node decides to transmit a packet or drop it depending on the channel state and the energy queue length. The objective for the source is to maximize the average amount of data successfully transmitted during its lifetime.

The rest of the paper is organized as follows. In section II, a system model is presented and the problem is formulated as an infinite-horizon discounted reward Markov Decision Processes(MDP). In section III, we define a low storage transmission policy and present an algorithm to find the optimal policy. Finally, numerical results are presented in section IV, followed by our concluding in section V.

2. SYSTEM MODEL

In this section, we present our system model for a point-to-point energy harvesting communication system and formulate the optimal transmission problem as an infinite-horizon discounted reward MDP. A time slotted system is considered, which the k-th time slot means $[k, k+1)$, $k \in Z$. Fig. 1 gives the model of the source node with a finite battery. The data queue is assumed to be saturated, that means there is always data to be sent at the beginning of each time slot. We use E to indicate the energy queue length.And the capacity of the battery is N units. And we also assume that the life time of source is geometric random variable with mean of $\frac{1}{1-\beta}$ $(0 < 1 - \beta \le 1)$. In other words, at each slot, the source node can terminate its operation with small probability 1-β, which is independent and identically distributed.
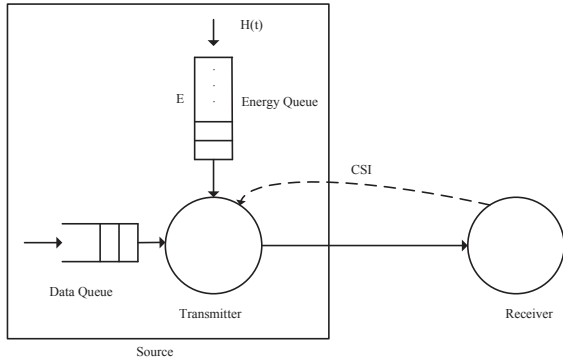

Figure 1: system model

We model the time correlation energy harvesting process $\{B_k\}$ as two-states Markov chain, as shown in Fig.2. In state 1, node can harvest one unit energy with probability λ_1, and can not harvest energy with probability $1 - \lambda_1$; in state 0, node can not harvest energy. Define $p_1 = P(B_k = 1|B_{k-1} = 1)$ as transmission probability from state 1 to state 1, $p_0 = P(B_k = 0|B_{k-1} = 0)$ as transmission probability from state 0 to state 0.
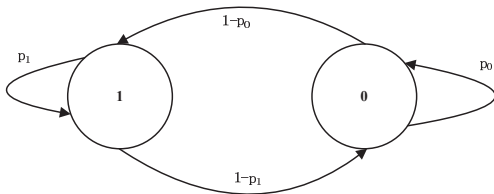

Figure 2: energy supply model

Let H_k be the channel state during the time slot k. We assume that H_k follows a Markov model and p_{lj} is the transmission probability from state l to state j. $h_k(0 \le h_k \le 1)$ is the successful transmission probability during slot k, which has M values. Here we assume the channel state process and energy harvesting process are independent.

At the beginning of each slot, the source node decides to either transmit a packet or discard it. One unit energy is used from the energy queue when a packet is transmitted. We assume the source node can receive the feedback from the receiver about the CSI at the beginning of each slot, then the action is taken by the source node based on $\{E_k, B_{k-1}, H_k\}$. We use a to indicate the action, where $a = T$ means the action is to transmit a packet and $a = D$ means the action is to discard a packet. The objective of the source node is to find the optimal policy that maximizes the expected total number of packets that are received successfully during the lifetime of the source node.

Next, define S as state space, which contains $\{E_k, B_{k-1}, H_k\}$, $A(i)$ as possible action set at state i, π as the policy used by the source node. We define $V(i, \pi)$ as the expected total reward with the initial state i at a given policy π. It can be calculated by the following expression

$$V(i, \pi) = E_i^\pi \left\{ E_Y \left[\sum_{k=1}^{Y} r(s_k, a_k) \right] \right\}, i \in S \qquad (1)$$

where k is the time slot index, s_k is the state at time slot k, a_k is the action taken at the time slot k, random variable Y is the lifetime of the source, E_Y denotes the statistical expectation and E_i^π denotes the expectation of total reward. The term $r(s_k, a_k)$ is reward when the state is s_k and the action is a_k, and has the following expression

$$r(s_k, a_k) = \left\{ \begin{array}{ll} h_k, a_k = T \\ 0, \ \ a_k = D \end{array} \right. \qquad (2)$$

The reward represents the number of packets received correctly when the action is taken. If the source chooses action T, which means to transmit a packet, the reward will be h_k depending the channel state at time slot k. If the source choose action D, which means to discard a packet, the reward will be 0 since there is no packet to be transmitted.

Due to the geometric distribution of the lifetime Y of the source with mean $\frac{1}{1-\beta}$, equation (1) can be equal to the objective function of infinite-horizon discounted reward, which has the following expression[9]:

$$V(i, \pi) = E_i^\pi \left\{ \sum_{k=0}^{\infty} \beta^k r(s_k, a_k) \right\}, i \in S \qquad (3)$$

The optimal problem can be expressed as follows:

$$V^*(i) = \max_\pi V(i, \pi) \qquad (4)$$

3. ONLINE TRANSMISSION POLICIES

The optimal problem is an infinite-horizon discounted reward MDP. Because in our system, state space S is countable and discrete, and the number of actions for each state is

finite, there exists an optimal deterministic stationary policy. Furthermore, policy iteration algorithm(PIA)[10] can be used to solve the problem and will terminate in a finite number of iterations with an optimal stationary policy.

However, when the number of states is large, the algorithm needs a lot of calculation to solve the linear system of equations, and also needs large storage to save the optimal policy. So in this section, we present a low storage policy and an algorithm to solve the optimal parameter of the policy.

3.1 Definition of Transmission Policies

The definition of transmission policies as follows,

$$a = \begin{cases} T, E_k \geq 1 \cap h_k \geq \theta_b \\ D, E_k = 0 \cup h_k < \theta_b \end{cases} \quad (5)$$

where E_k is the energy level of k-th slot, h_k is the successful transmission probability during slot k. θ_b is a threshold of previous slot's energy harvesting state b, which b can be 1 or 0, and $\theta_0 \geq \theta_1$.

When energy queue length $E_k = 0$, the source takes action D; When energy queue length $E_k \geq 1$, the source decides action according to h_k. When $h_k < \theta_b$, the source tends to take action D to save energy; when $h_k \geq \theta_b$, the source tends to take action T to get good reward. θ_1 is threshold of $B_{k-1} = 1$, and θ_0 is threshold of $B_{k-1} = 0$. And when the energy harvesting rate is larger, more data should be transmitted, so we have $\theta_0 \geq \theta_1$.

In the optimal policy solved by PIA, source needs to store a look-up table, which contains the states and corresponding action. In our present policy, source only needs to store two thresholds θ_1 and θ_0. Our present policy has the advantages of low storage capacity.

3.2 Algorithm of Optimal Threshold

From the policy, we can see different thresholds lead to different rewards. In this section, we present an algorithm to solve the optimal threshold.

Firstly, we narrow the feasible region of the threshold from the real domain to discrete set with limited number. The range of θ_b is $[0,1]$. Sort the channel state from small to large by the successful transmission probability, denoted by $h_1, h_2, \cdots, h_M$ with $h_0 = 0$. Assumed there is $h_i < \theta_b \leq h_{i+1}, i = 0, 1, \cdots, M-1$, then when the k-th successful transmission probability is any of $\{h_1, h_2, \cdots, h_i\}$, the source takes action D; when the k-th successful transmission probability is any of $\{h_{i+1}, h_{i+2}, \cdots, h_M\}$, the source takes action T. The effect is equal to $\theta_b = h_{i+1}$. And if $\theta_b = h_M + \varepsilon, \varepsilon > 0$, the source can only take action D. So the feasible region of threshold can be narrowed to the set $\{h_1, h_2, \cdots, h_M, h_M + \varepsilon\}, \varepsilon > 0$.

The optimal threshold can be solved by combining optimal equations [10] and enumeration algorithm. The algorithm flowchart is as follows:

1. Sort the channel state from small to large by the successful transmission probability, denoted by $h_1, h_2, \cdots, h_M$.

   So $\theta_b \in \{h_1, h_2, \cdots, h_M, h_M + \varepsilon\}, \varepsilon > 0, b = 0 \ or \ 1$, and $\theta_0 \geq \theta_1$.

2. Enumerate the possible values of θ_1, θ_0, to get policy according to equation (5). Then solve the following equation to get the rewards. The threshold, which make the reward biggest is optimal.

$$(I - \beta P(f)) V = R(f) \quad (6)$$

In equation (6), V is the expected total reward, $R(f)$ is the instant reward of each state given policy f, which can be computed according to equation (2). $P(f)$ is transmission probability matrix at given policy f. Next we introduce computational formula of $P(f)$.

The k-th state is denoted by (E_k, B_{k-1}, H_k). On one hand, when source takes action T to transmission a packet, one unit energy is depleted. On the other hand, if $B_{k-1} = 1$, the source can harvest one unit energy by the probability λ_1, so the energy level can either remain same or decrease by one unit. If $B_{k-1} = 0$, the source can't harvest energy, so the energy level E_k will decrease one unit. Also, channel state process and energy harvesting process are independent. So we can derive the following formula.

$$1 \leq E \leq N$$
$$P((E,1,j)|(E,1,k)) = \lambda_1 p_1 p_{kj}$$
$$P((E,0,j)|(E,1,k)) = \lambda_1 (1 - p_1) p_{kj}$$
$$P((E-1,1,j)|(E,1,k)) = (1 - \lambda_1) p_1 p_{kj} \quad (7)$$
$$P((E-1,0,j)|(E,1,k)) = (1 - \lambda_1)(1 - p_1) p_{kj}$$
$$P((E-1,1,j)|(E,0,k)) = (1 - p_0) p_{kj}$$
$$P((E-1,0,j)|(E,0,k)) = p_0 p_{kj}$$

When source takes action D, energy level E_k will either remain same or increase one unit. However, due to the limit of battery capacity, when $E_k = N$, energy level E_k can only be N. So we can derive the following formula.

$$0 \leq E < N$$
$$P((E+1,1,j)|(E,1,k)) = \lambda_1 p_1 p_{kj}$$
$$P((E+1,0,j)|(E,1,k)) = \lambda_1 (1 - p_1) p_{kj}$$
$$P((E,1,j)|(E,1,k)) = (1 - \lambda_1) p_1 p_{kj}$$
$$P((E,0,j)|(E,1,k)) = (1 - \lambda_1)(1 - p_1) p_{kj}$$
$$P((E,1,j)|(E,0,k)) = (1 - p_0) p_{kj}$$
$$P((E,0,j)|(E,0,k)) = p_0 p_{kj} \quad (8)$$
$$E = N$$
$$P((E,1,j)|(E,1,k)) = p_1 p_{kj}$$
$$P((E,0,j)|(E,1,k)) = (1 - p_1) p_{kj}$$
$$P((E,1,j)|(E,0,k)) = (1 - p_0) p_{kj}$$
$$P((E,0,j)|(E,0,k)) = p_0 p_{kj}$$

4. NUMERICAL RESULTS

In this section, we compare the performance of three transmission policies. They are optimal policies by PIA, our presented adaptive transmission policies, and greedy polices which means that source can take action T only when $B_{k-1} = 1$ and $E_k > 0$. The performance is the expected total number of packets that are received successfully during the lifetime of the source node with initial state $E_k = 0$.

We consider H_k are Gilbert-Elliot channel model[11], which has $p_{11} = 0.8, p_{01} = 0.4, h_1 = 0.5, h_0 = 0.2$. Besides, $N = 10, \beta = 0.95$.
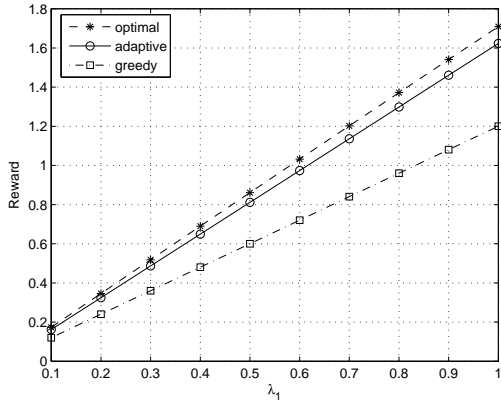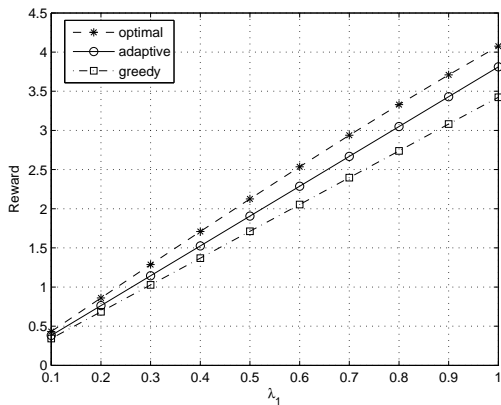
Figure 3: $p_1 = 0.2, p_0 = 0.8$

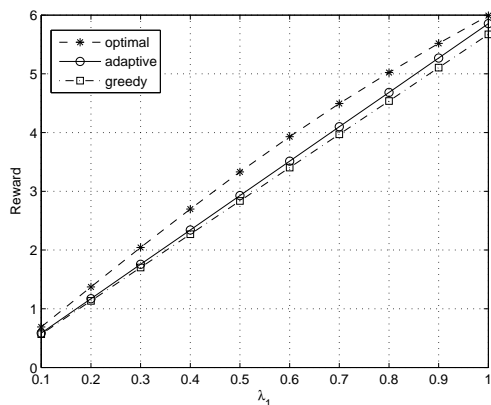
Figure 4: $p_1 = 0.5, p_0 = 0.5$

Figure 5: $p_1 = 0.8, p_0 = 0.2$

Fig. 3-5 show the performance against energy harvesting probability λ_1 over different parameters of the energy harvesting process. In Fig.3, $p_1 = 0.2, p_0 = 0.8$ means source stays on energy harvesting state 0 for a longer time; in Fig.4, $p_1 = 0.5, p_0 = 0.5$ means source stays on energy harvesting state 0 and 1 for the same time; in Fig.5, $p_1 = 0.8, p_0 = 0.2$ means source stays on energy harvesting state 1 for a longer time. First of all, from the three figures, we can see that the performance of transmission policies increases with the increasing of energy harvesting probability λ_1. This is due to the fact that the bigger energy harvesting probability is , more energy can be used to transmit data. Secondly, the performance of our presented adaptive policy is larger than that of Greedy policies, but smaller than that of optimal policies. However, when energy harvesting probability is lower than a threshold, the performance of our presented adaptive policy is a good approximation of optimal polices's performance. Finally, the difference of performance between optimal polices and adaptive polices increases with the increase of energy harvesting probability λ_1. This is because in our presented adaptive policies, we only consider different thresholds for different energy harvesting states. However, the thresholds have nothing to do with energy level.

5. CONCLUSIONS

In this paper, we have considered a point-to-point energy harvesting system with time-correlated energy supply and limited lifetime of source over time varying channel in a time slot system. Our goal is to find optimal transmission policies to maximize the expected sum of successfully transmitted data during the lifetime of the source. First, assumed that the life time of source is geometric random variable, we formulated above problem as an infinite-horizon discounted reward MDP. Then we presented a low storage transmission policies and introduced an algorithm to solve the optimal parameter in our present policies. Also, numerical results show that when the energy harvesting is lower than a threshold, the performance of our presented policies can be a good approximation of that of optimal polices, which can be derived through policy iteration algorithm.

6. ACKNOWLEDGMENTS

7. REFERENCES

[1] Seah, Winston KG, Zhi Ang Eu, and Hwee-Pink Tan. "Wireless sensor networks powered by ambient energy harvesting (WSN-HEAP)-Survey and challenges." Wireless Communication, Vehicular Technology, Information Theory and Aerospace & Electronic Systems Technology, 2009. Wireless VITAE 2009. 1st International Conference on. IEEE, 2009.

[2] Gunduz D, Stamatiou K, Michelusi, Zorzi M. "Design intelligient energy harvesting communication systems[J]." IEEE Communications Magazine, 2014, vol 52(1), pp:210-216.

[3] Jing Lei and Yates, R. and Greenstein, L.,"A generic model for optimizing single-hop transmission policy of

replenishable sensors", Wireless Communications, IEEE Transactions on Feb 2009, vol 8(2), pp.547-551.

[4] Michelusi, N. and Stamatiou, K. and Zorzi, M.,"Transmission Policies for Energy Harvesting Sensors with Time-Correlated Energy Supply",Communications, IEEE Transactions on July, 2013, vol 6(7),pp.2988-3001.

[5] Sharma, V. and Mukherji, U. and Joseph, V. and Gupta, S.,"Optimal energy management policies for energy harvesting sensor nodes", Wireless Communications, IEEE Transactions on April 2010, vol 9(4), pp.1326-1336.

[6] Bacinoglu, B.T. and Uysal-Biyikoglu, E. "Finite horizon online packet scheduling with energy and delay constraints", Communications and Networking (BlackSeaCom), 2013 First International Black Sea Conference on July, 2013, pp. 191-195.

[7] Kashef, M. and Ephremides, Anthony,"Optimal packet scheduling for energy harvesting sources on time varying wireless channels",Communications and Networks, Journal of April 2012, vol 14(2), pp.121-129.

[8] Blasco, P. and Gunduz, D. and Dohler, M.,"A learning theoretic approach to energy harvesting communication system optimization",Globecom Workshops (GC Wkshps), 2012 IEEE, pp.1657-1662.

[9] M. L. Puterman, "Markov Decision Processes: Discrete Stochastic Dynamic Programming", New York,NY:John Wiley and Sons, 2005.

[10] Liu Ke, "Applied Markov Decision processes", Tsinghua University Press, 2004.

[11] Wang H.S., Moayeri N. "Finite-State Markov Channel: A Useful Model for Radio Communication Channels[J]." Vehicular Technology, IEEE Transcation on 1995, VOL 44(1), PP:163-171.