Relay Selection with Channel Probing for Geographical Forwarding in WSNs

K. P. Naveen and Anurag Kumar
Dept. of Electrical Communication Engg., Indian Institute of Science, Bangalore 560 012, India.
{naveenkp, anurag}@ece.iisc.ernet.in

***Abstract*—In a large wireless sensor network (WSN) with sleep-wake cycling nodes, we are interested in the local decision problem faced by a node that has "custody" of a packet and has to choose one among a set of next-hop relay nodes. Each of the relays is associated with a "reward" that summarizes the cost/benefit of forwarding the packet through that relay. We seek a locally optimal solution to this problem, the idea being that such a solution, if adopted by every node, could provide a reasonable local heuristic for the end-to-end forwarding problem. Towards this end, we propose a *local forwarding problem* where the relays wake-up at random times, at which instants they reveal the probability distributions of their rewards. To determine a relay's exact reward, the source has to further probe the relay, incurring a probing cost. Thus, at each relay wake-up instant, the forwarding node, given the reward value of an already probed relay and the reward distribution of an unprobed relay, has to decide whether to stop (and forward the packet to the probed relay), continue waiting for further relays to wake-up, or probe the unprobed relay. We formulate this local forwarding problem as a Markov decision process (MDP) and obtain some interesting structural results on the optimal policy. Our problem can be considered as a new variant of the *asset selling problem* studied in the operations research literature.**

I. Introduction

Consider a wireless sensor network deployed for the detection of a *rare event* e.g., forest fire, intrusion in border areas, etc. To conserve energy, the nodes in the network *sleep-wake* cycle whereby they alternate between an ON state and a low power OFF state. We are further interested in *asynchronous* sleep-wake where the point processes of wake-up instants of the nodes are mutually independent [1], [2].

In such networks, whenever an event is detected, an alarm packet (containing the event location and a time stamp) is generated and has to be forwarded, through multiple hops, to a control center (*sink*) where appropriate action could be taken. Since the network is sleep-wake cycling, a forwarding node (i.e., a node with an alarm packet) has to wait for its neighbors to wake-up before it can choose a neighbor for the next hop. Thus, there is a delay incurred, due to the sleep-wake process, at each hop enroute to the sink. We are interested in minimizing the total average end-to-end delay subject to a constraint on some global metric of interest such as the average hop counts, or the average total transmission power (sum of the transmission power used at each hop). Such a global problem can be considered as a stochastic shortest path problem [3], for which a distributed Bellman-Ford algorithm (e.g., the LOCAL-OPT algorithm proposed by Kim et al. in [1]) can be used to obtain the optimal solution. A major drawback with such an approach is that a pre-configuration phase is required to run such algorithms, which would involve exchange of several control messages.

The focus of our research is, instead, towards designing *simple forwarding rules* using only the *local information* available at a forwarding node. We have already made efforts in this direction [2], [4] by proposing a *local forwarding problem* where we minimize one-hop delay subject to a constraint on the reward offered by the chosen relay, where the reward associated with a relay is a function of the transmission power and the progress, towards sink, made by the packet when forwarded to that relay. Through simulations we found that, in some region of operation, the end-to-end performance (i.e., total delay and total transmission power) obtained by applying the solution to the local problem at each hop is comparable with that obtained by the global solution (i.e., the LOCAL-OPT proposed by Kim et al. [1]), giving us the confidence that it is reasonable to solve the local forwarding problem.

However in our earlier work we assumed that the gain of the channel between the forwarding node and a relay is a deterministic function of the distance between the two, whereas, in practice the channel is random and the forwarding node has to send probe packets to learn the channel gain [5], thus incurring an additional energy cost. In the current work we have incorporated these features (namely, channel probing and the associated energy cost) while choosing a relay for the next hop, leading to an interesting variant of the asset selling problem [6].

Outline and Contributions: In Section II we will formally describe our system model, following which we will discuss the related work. Sections III and IV are devoted towards characterizing the structure of the policy RST-OPT (ReSTricted-OPTimal) which is optimal within a restricted class of relay selection policies. In Section V we will informally discuss the globally optimal policy, UnRST-OPT (UnREStricted-OPTimal). Our main technical contributions are:

- We characterize the optimal policy, RST-OPT, in terms of *stopping* sets. We prove that these stopping sets have threshold structure (Theorem 1).
- We further prove that the stopping sets are identical across the decision stages (Theorem 2 and 3). This can be considered as a generalization of the *one-step-look ahead* rule (see the Remark following Theorem 2).

- Through numerical work (Section VI) we find that the performance of RST-OPT is close to that of the UnRST-OPT, which is more computationally intensive and energy comsuming than RST-OPT.

For most of the proofs we refer to our technical report [7].

II. System Model: The Local Forwarding Problem

In this section we will describe the system model in the context of *geographical forwarding*, also known as location aware routing, [2], [8], [9]. In geographical forwarding it is assumed that each node in the network knows its location (with respect to some reference) as well as the location of the sink.

Consider a forwarding node (henceforth referred to as the *source*) at a distance v_0 from the *sink* (see Fig. 1). The *communication region* of the source is the set of all locations where reliable exchange of *control messages* can take place between the source and a receiver, if any, at these locations. In Fig. 1 we have shown the communication region to be circular, but in practice this region can be arbitrary. The set of nodes within the communication region are referred to as *neighbors*. Let v_ℓ represent the distance of a location ℓ (a location is a point in $\Re^2$) from the sink. Then define the *progress* of the location ℓ as $Z_\ell := v_0 - v_\ell$. The source is interested in forwarding the packet only to a neighbor within the *forwarding region* $\mathcal{L}$ where, $\mathcal{L} = \Big\{\ell \in \text{communication region} : Z_\ell \geq 0\Big\}$. The forwarding region is shown hatched in Fig. 1. We will refer to the nodes in the forwarding region as *relays*.

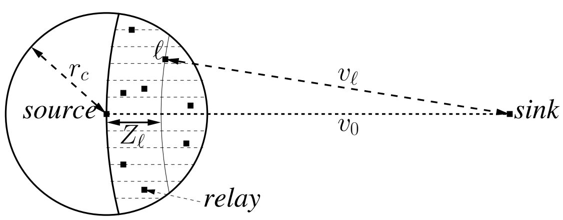


Fig. 1. The hatched area is the forwarding region $\mathcal{L}$. For a location $\ell \in \mathcal{L}$, the progress Z_ℓ is the difference between the source-sink distance, v_0, and the location-sink distance, v_ℓ.

Let H_ℓ represent the random gain of the channel between the source and location $\ell \in \mathcal{L}$. We will assume that the channel gains, $\{H_\ell : \ell \in \mathcal{L}\}$, are independent and identically distributed (i.i.d.). We will further assume that the channel coherence time is such that during the local decision process the channel gains in the forwarding area remain unchanged. The rate at which packets arrive to be forwarded through a region is low enough so that the channel gains change in between each forwarding instance. Such low forwarding rates would arise in sensor networks whose main purpose is to detect infrequent events (fires, intrusions, gas leaks, etc.).

Given H_ℓ, the minimum power required to achieve an SNR threshold of Γ at location ℓ is $P_\ell = \frac{\Gamma N_0 d_\ell^\beta}{|H_\ell|^2}$, where d_ℓ is the distance between the source and the location ℓ, β is the path loss attenuation factor, and N_0 is the noise variance.

Reward Structure: Several metrics have been proposed in the literature [10] to enable a forwarding node to evaluate its neighbors before choosing one for the next hop. In the current work we prefer to use a reward metric, which is a combination of the progress, Z_ℓ, and the minimum power required, P_ℓ. Formally, to each location $\ell \in \mathcal{L}$ we associate a *reward* R_ℓ as

$$R_\ell = \frac{Z_\ell^a}{P_\ell^{(1-a)}} = \frac{Z_\ell^a}{(\Gamma N_0 d_\ell^\beta)^{(1-a)}} (|H_\ell|^2)^{(1-a)}, \tag{1}$$

where $a \in [0,1]$ is used to tradeoff between Z_ℓ and P_ℓ. The reward being inversely proportional to P_ℓ is clear, because it is advantageous to use less power to get the packet across. R_ℓ is made proportional to Z_ℓ to promote progress towards the sink while choosing a relay for the next hop. Further motivation for choosing the particular structure for the reward is available in [7, Appendix VIII-A]. Finally, let F_ℓ represent the distribution of R_ℓ. Thus, there is a collection of reward distributions $\mathcal{F}$ indexed by $\ell \in \mathcal{L}$. From (1), note that, given a location ℓ it is only possible to know the reward distribution F_ℓ. To know the exact reward R_ℓ, the source has to send probe packets and learn the channel gain H_ℓ.

Sleep-Wake Process: Since we focus on local forwarding, we will assume that the source gets a packet to forward (either from an upstream node or by detecting an event) at time 0. There are N relays in the forwarding set $\mathcal{L}$ that wake-up sequentially at the points of a renewal process, $W_1 \leq W_2 \leq \cdots \leq W_N$. Let $U_k := W_k - W_{k-1}$ ($U_1 := W_1$) denote the *inter-wake-up time* (renewal lifetime) between the k-th and the $(k-1)$-th relay. Then $U_1, U_2, \cdots, U_N$, are independent with their means, equal to τ. For example, U_k, $1 \leq k \leq N$, could be exponentially distributed with mean τ, or could be constant (deterministic) with value τ.

Sequential Decision Problem: Let $L_1, L_2, \cdots, L_N$, denote the relay locations which are assumed to be i.i.d. uniform over the forwarding set $\mathcal{L}$. Let $A(\cdot)$ denote the uniform distribution over $\mathcal{L}$ so that, for $k = 1, 2, \cdots, N$, the distribution of L_k is $A(\cdot)$. The source (with a packet to forward at time 0) only knows that there are N relays in its forwarding set $\mathcal{L}$, but does not know their locations, L_k, nor their channel gains, H_{L_k}. When the k-th relay wakes up, we assume that its location L_k and hence the reward distribution F_{L_k} is revealed to the source. This can be accomplished by including the location information L_k within a control packet (sent using low rate robust modulation technique, and hence, assumed to be error free) transmitted by the k-th relay upon wake-up. However, if the source wishes to learn the channel gain H_{L_k} (and hence the exact reward value R_{L_k}), it has to send additional probe packets (indeed several packets, in order to obtain a reliable estimate of the channel gain) incurring an energy cost of δ units. Thus, when the k-th relay wakes up (referred to as stage k) the actions available to the source are:

- stop and forward the packet to a previously probed relay, thereby accruing the reward of that relay. It is clear that it is optimal to forward, among all the probed relays, to the relay with the maximum reward. Thus, henceforth, the action stop always implies that the best relay (among those that have been probed) is chosen. With the stop action the decision process terminates.

- continue to wait for the next relay to wake-up (the average waiting time is τ) and reveal its reward distribution, at which instant the decision process is regarded to have entered stage $k+1$.
- probe a relay from the set of all unprobed relays (provided there is at least one unprobed relay). The probed relay's reward value is revealed allowing the source to update maximum reward among all the probed relays. After probing, the decision process is still at stage k and again the source has to decide upon an action.

In the model, for the sake of analysis, we neglect the time taken to probe a relay and learn its channel gain. We also neglect the time taken for the exchange of control packets. This is reasonable for very low duty cycling networks where the average inter-wake-up time is much larger than the time taken for probing, or the exchange of control packets.

A *decision rule* or a *policy* is a mapping, at each stage, from all histories of the decision process to the set of actions. Let Π represent the class of all policies. For a policy $\pi \in \Pi$ the delay incurred, D_π, is the time until a relay is chosen (which is one of the W_k). Let R_π represent the reward of the chosen relay and M_π be the total number of relays that were probed during the decision process. Recalling that δ is the cost of probing, δM_π represents the total cost of probing if policy π is used. We would like to think of $(R_\pi - \delta M_\pi)$ as the *effective reward* obtained using policy π.

The problem we are interested in is the following: $\min_{\pi\in\Pi} \mathbb{E}D_\pi$ subject to $\Big(\mathbb{E}R_\pi - \delta\mathbb{E}M_\pi\Big) \geq \gamma$, where $\gamma > 0$ is a constraint on the effective reward. We introduce a "Lagrange" multiplier $\eta > 0$ and focus our attention towards solving the following unconstrained problem:

$$\min_{\pi\in\Pi} \Big(\mathbb{E}D_\pi - \eta\mathbb{E}R_\pi + \eta\delta\mathbb{E}M_\pi\Big). \qquad (2)$$

We will call Π the *unrestricted class* of policies since Π contains policies that are allowed, for each stage k, to keep all the relays until stage k awake. Formally, if $b_k = \max\{R_{L_i} : i \leq k, \text{ relay } i \text{ has been probed}\}$ and $\mathcal{F}_k = \{F_{L_i} : i \leq k, \text{ relay } i \text{ is unprobed}\}$, then the decision at stage k is based on $(b_k, \mathcal{F}_k)$. Thus, the set of all possible states at stage k is large. Hence, for analytical tractability, we first consider (in Sections III and IV) solving the problem in (2) over a *restricted class* of policies $\overline{\Pi} \subseteq \Pi$ where a policy in $\overline{\Pi}$ is restricted to take decisions keeping only two relays awake, one the best among all probed relays and one among the unprobed ones, i.e., the decision at stage k is based on (b_k, G_k) where $G_k \in \mathcal{F}_k$.

A. Related Work

The work reported in this paper is a major extension of our earlier work ([2], [4]). The difference is that in [2] we assume that when a relay wakes up its exact reward value is revealed. This is reasonable if the reward is simply the geographical progress (towards the sink) made by a relay, which was the case in [2]. In [4] we studied a variant where the number of relays N is not known to the source. However, in [4], as well, the exact reward value is revealed by a relay upon wake-up.

There has been other work in the context of geographical forwarding and anycast routing, where the problem of choosing one among several neighboring nodes has been studied [9]. The authors in [9] study the greedy policy of always choosing a neighbor that makes the maximum progress towards the sink. Thus, they do not study the tradeoff between the relay selection delay and the progress (or other reward metric), which is the major contribution of our work. For a sleep-wake cycling network, Kim et al. in [1] have developed a distributed Bellman-Ford algorithm (referred to as LOCAL-OPT) to minimize the average end-to-end delay. However a pre-configuration phase, involving lot of control message exchanges, is required to run the LOCAL-OPT algorithm. In all of the above work, the metric that signifies the quality of a relay is always exactly revealed, and hence does not involve a probing action. The action to probe generally occurs in the problem of channel selection [11]. We will discuss about the work in [11], in detail, in Section V.

Finally (but very importantly) our work can be considered as a variant of the *asset selling problem* which is a major class within the *optimal stopping problems* [12], [13] (the other classes being the secretary problem, the bandit problem, etc.,). The basic asset selling problem [6], comprises offers that arrive sequentially over time. The offers are i.i.d. As the offers arrive, the seller has to decide whether to take an offer or wait for future offers. The seller has to pay a cost to observe the next offer. Previous offers cannot be recalled. The decision process ends with the seller choosing an offer. Over the years, several variants of the basic problem have been studied, e.g., problems with uncertain recall, problems with unknown reward distribution, etc. In most of the variants, when an offer arrives, its value is exactly revealed to the seller, while in our case only the offer (i.e., reward) distribution is made available and the source, if it wishes, has to probe to known the exact offer value.

Close to our work is that of Stadje [14] where only some initial information about an offer (e.g., the average size of the offer) is revealed to the decision maker upon its arrival. In addition to the actions, stop and continue, the decision maker can also choose to obtain more information about the offer by incurring a cost. Recalling previous offers is not allowed. A similar problem is studied by Thejaswi et al. in [5], where initially a coarse estimate of the channel gain is made available to the transmitter. The transmitter can choose to probe the channel, the second time, to get a finer estimate. In both of these works ([14], [5]), the optimal policies is characterized by a threshold rule. However, the horizon length of these problems is infinite, because of which the thresholds are stage independent. In general, for a finite horizon problem the optimal policy would be stage dependent. However, for our problem (despite being a finite horizon one) we are able to show that certain optimal stopping sets are identical at every stage. This is due to the fact that we allow the best probed relay to stay awake.

III. Restricted Class $\overline{\Pi}$: An MDP Formulation

In this section we will formulate our local forwarding problem as a Markov decision process [13], which will require us to first discuss the one-step cost and the state transition.

A. One-Step Costs and State Transition

Recall that for any policy in the restricted class $\overline{\Pi}$, the decision at stage k should be based on (b_k, G_k) where b_k is the best reward so far and $G_k \in \mathcal{F}_k$ is the reward distribution of an unprobed relay that is retained until stage k.

If the source's decision is to stop, then the source enters a terminating state ψ incurring a cost of $-\eta b_k$ (recall from (2) that η is the Lagrange multiplier).

If the action is to continue then the source will first incur a waiting cost of U_{k+1} (average cost is τ). When the $(k+1)$-th relay wakes-up (whose reward distribution is $F_{L_{k+1}}$), first the source has to choose between G_k and $F_{L_{k+1}}$, then put the other one to sleep so that the state at stage $k+1$ will again be of the form (b_{k+1}, G_{k+1}). Since the action at stage k was to continue, the set of probed relays at stage $k+1$ remain unchanged so that $b_{k+1} = b_k$.

Finally the source could, at stage k, probe the distribution G_k incurring a cost of $\eta\delta$. After probing, the decision process is still considered to be at stage k where now the state is $b'_k = \max\{b_k, R_k\}$ (where R_k is a sample from the distribution G_k). The source has to now further decide whether to stop (incurring a cost of $-\eta b'_k$ and enter ψ) or continue (cost is U_{k+1} and the next state is $(b'_k, F_{L_{k+1}})$). Note that for a policy π, the sum of all the one-step costs, starting from stage 1, will equal the total cost[1] in (2).

B. Cost-to-go Functions and the Bellman Equation

Let $J_k(\cdot)$, $k = 1, 2, \cdots, N$, represent the optimal cost-to-go function at stage k. Thus, $J_k(b)$ and $J_k(b, F_\ell)$ denote the cost-to-go, depending on whether there is or is not an unprobed relay. For the last stage, N, we have, $J_N(b) = -\eta b$, and

$$\begin{aligned} J_N(b, F_\ell) &= \min\Big\{-\eta b, \eta\delta + \mathbb{E}_\ell\Big[J_N(\max\{b, R_\ell\})\Big]\Big\} \\ &= \min\Big\{-\eta b, \eta\delta - \eta\mathbb{E}_\ell\Big[\max\{b, R_\ell\}\Big]\Big\} \quad (3)\end{aligned}$$

where $\mathbb{E}_\ell$ denotes expectation w.r.t. the distribution F_ℓ. The first term in the min-expression above is the cost of stopping and the second term is the average cost of probing and then stopping (the action continue is not available at the last stage N). Next, for stage $k = 1, 2, \cdots, N-1$, denoting by $\mathbb{E}_A$ the expectation over the distribution, $A(\cdot)$, of the index, L_{k+1}, of the next relay to wake up, we have

$$J_k(b) = \min\Big\{-\eta b, \tau + \mathbb{E}_A\Big[J_{k+1}(b, F_{L_{k+1}})\Big]\Big\}, \quad (4)$$

$$\begin{aligned} J_k(b, F_\ell) = \min\Big\{&-\eta b, \eta\delta + \mathbb{E}_\ell\Big[J_k(\max\{b, R_\ell\})\Big], \\ &\tau + \mathbb{E}_A\Big[\min\{J_{k+1}(b, F_\ell), J_{k+1}(b, F_{L_{k+1}})\}\Big]\Big\}. \quad (5)\end{aligned}$$

[1] Since every policy has to invariably wait for the first relay to wake-up, at which instant the decision process begins, U_1 is not accounted for in the total cost by any policy π.

The last term in the min-expression of (4) and (5) is the average cost of continuing. When the state at stage k, $1 \le k \le N-1$, is (b, F_ℓ) and, if the source decides to continue, then the reward distribution of the next relay is $F_{L_{k+1}}$. Now, given the distributions F_ℓ and $F_{L_{k+1}}$, if the source is asked to retain one of them, it is always optimal to go with the distribution that fetches a lower cost-to-go from stage $k+1$ onwards, i.e., it is optimal to retain F_ℓ if $J_{k+1}(b, F_\ell) \le J_{k+1}(b, F_{L_{k+1}})$, otherwise retain $F_{L_{k+1}}$[2]. Later in this section (Lemma 1-(ii)) we will show that if F_ℓ is *stochastically greater* than F_u then $J_{k+1}(b, F_\ell) \le J_{k+1}(b, F_u)$.

First, for simplicity let us introduce the following notations. For $k = 1, 2, \cdots, N-1$, let $cc_k(\cdot)$ represent the cost of continuing, i.e.,

$$cc_k(b) \quad = \quad \tau + \mathbb{E}_A\Big[J_{k+1}(b, F_{L_{k+1}})\Big] \quad (6)$$

$$\begin{aligned} &cc_k(b, F_\ell) = \\ &\qquad \tau + \mathbb{E}_A\Big[\min\{J_{k+1}(b, F_\ell), J_{k+1}(b, F_{L_{k+1}})\}\Big] \quad (7)\end{aligned}$$

and, for $k = 1, 2, \cdots, N$, the cost of probing, $cp_k(\cdot)$, is

$$cp_k(b, F_\ell) = \eta\delta + \mathbb{E}_\ell\Big[J_k(\max\{b, R_\ell\})\Big]. \quad (8)$$

Using the above notations, the cost-to-go functions can be written as, for $k = 1, 2, \cdots, N-1$,

$$\begin{aligned} J_k(b) &= \min\Big\{-\eta b, cc_k(b)\Big\} \quad (9) \\ J_k(b, F_\ell) &= \min\Big\{-\eta b, cp_k(b, F_\ell), cc_k(b, F_\ell)\Big\}. \quad (10)\end{aligned}$$

Definition 1 (Stochastic Ordering): Given two distributions F_ℓ and F_u we say that F_ℓ is stochastically greater than F_u, denoted as $F_\ell \ge_{st} F_u$, if $F_\ell(x) \le F_u(x)$ for all x. Alternatively, one could use the following definition [15]: $F_\ell \ge_{st} F_u$ if and only if for every non-decreasing function $f: \Re \to \Re$, $\mathbb{E}_\ell[f(R_\ell)] \ge \mathbb{E}_u[f(R_u)]$ where the distributions of R_ℓ and R_u are F_ℓ and F_u, respectively. ■

We end this section by listing out the various ordering properties of the cost-to-go function in the following lemma.

Lemma 1: For $1 \le k \le N$ (for part (iii), $1 \le k \le N-1$),

(i) $J_k(b)$ and $J_k(b, F_\ell)$ are decreasing in b.
(ii) If $F_\ell \ge_{st} F_u$ then $J_k(b, F_\ell) \le J_k(b, F_u)$.
(iii) $J_k(b) \le J_{k+1}(b)$ and $J_k(b, F_\ell) \le J_{k+1}(b, F_\ell)$.

Proof: Part (i) and (iii) follow easily by straightforward induction. To prove part (ii) we need to use part (i) and the definition of stochastic ordering (Definition 1). Formal proof is available in [7, Appendix VIII-B]. ■

[2] Formally one has to introduce an intermediate state of the form $(b, F_\ell, F_{L_{k+1}})$ at stage $k+1$ where the only actions available are, choose F_ℓ or $F_{L_{k+1}}$. Then $J_{k+1}(b, F_\ell, F_{L_{k+1}}) = \min\{J_{k+1}(b, F_\ell), J_{k+1}(b, F_{L_{k+1}})\}$, which, for simplicity, we are directly using in (5).

IV. Restricted Class $\overline{\Pi}$: Structural Results

We begin by defining, at stage $1 \leq k \leq N-1$, the *optimal stopping set* $\mathcal{S}_k$ as

$$\mathcal{S}_k := \Big\{ b : -\eta b \leq cc_k(b) \Big\}, \tag{11}$$

Similarly, for a given distribution F_ℓ we define the *optimal stopping set* $\mathcal{S}_k^\ell$ as, for $1 \leq k \leq N-1$,

$$\mathcal{S}_k^\ell \quad := \quad \Big\{ b : -\eta b \leq \min\{cp_k(b, F_\ell), cc_k(b, F_\ell)\} \Big\}. \tag{12}$$

From (9) it follows that the optimal stopping set $\mathcal{S}_k$ is the set of states b (states of this form are obtained after probing at stage k) where it is better to stop than to continue. Similarly from (10), the set $\mathcal{S}_k^\ell$ has to be interpreted as, for a given distribution F_ℓ, the set of b such that at state (b, F_ℓ) it is better to stop than to probe or continue.

From the definition of these sets, and using Lemma 1 one can prove the following set inclusion properties.

Lemma 2: (i) For $1 \leq k \leq N-2$ and any F_ℓ we have $\mathcal{S}_k \subseteq \mathcal{S}_{k+1}$ and $\mathcal{S}_k^\ell \subseteq \mathcal{S}_{k+1}^\ell$.
(ii) For $1 \leq k \leq N-1$ and any F_ℓ we have $\mathcal{S}_k^\ell \subseteq \mathcal{S}_k$.
(iii) For $1 \leq k \leq N-1$, if $F_\ell \geq_{st} F_u$ then $\mathcal{S}_k^\ell \subseteq \mathcal{S}_k^u$.

Proof: See [7, Proof of Lemma 4] ■

Our first main result is to show that the stopping sets can be characterized by thresholds (Theorem 1). Next we prove that the stopping sets are identical across the stages, i.e., $\mathcal{S}_k = \mathcal{S}_{k+1}$ and $\mathcal{S}_k^\ell = \mathcal{S}_{k+1}^\ell$ (Theorem 2 & 3).

A. Stopping Sets: Threshold Structure

To prove the threshold structure of the stopping sets the following key lemma is required, where we show that the increments in the various costs are bounded by the increment in the cost of stopping.

Lemma 3: For $1 \leq k \leq N-1$ and for $b_2 > b_1$ we have
(i) $cc_k(b_1) - cc_k(b_2) \leq \eta(b_2 - b_1)$,
and for any distribution F_ℓ we have,
(ii) $cp_k(b_1, F_\ell) - cp_k(b_2, F_\ell) \leq \eta(b_2 - b_1)$
(iii) $cc_k(b_1, F_\ell) - cc_k(b_2, F_\ell) \leq \eta(b_2 - b_1)$.

Proof: See [7, Appendix VIII-C]. ■

Theorem 1: For $1 \leq k \leq N-1$ and for $b_2 > b_1$,
(i) If $b_1 \in \mathcal{S}_k$ then $b_2 \in \mathcal{S}_k$.
(ii) For any F_ℓ, if $b_1 \in \mathcal{S}_k^\ell$ then $b_2 \in \mathcal{S}_k^\ell$.

Proof: Part (i): Using Lemma 3-(i) we can write,

$$-\eta b_2 \leq -\eta b_1 - cc_k(b_1) + cc_k(b_2).$$

Since $b_1 \in \mathcal{S}_k$, from (11) we know that $-\eta b_1 \leq cc_k(b_1)$, using which in the above expression, we obtain $-\eta b_2 \leq cc_k(b_2)$ implying that $b_2 \in \mathcal{S}_k$. *Part (ii)* similarly follows using Lemma 3-(ii) and 3-(iii). ■

Discussion: Thus, the stopping set $\mathcal{S}_k$ can be characterized in terms of a lower bound x_k as illustrated in Fig. 2(a). Similarly for distributions F_ℓ and F_u, there are (possibly different) thresholds x_k^ℓ and x_k^u (Fig. 2(b) and 2(c)). Using Lemma 2-(ii) and 2-(iii), for $F_\ell \geq_{st} F_u$ we can write, $x_k \leq x_k^u \leq x_k^\ell$. These thresholds are for a given stage k.

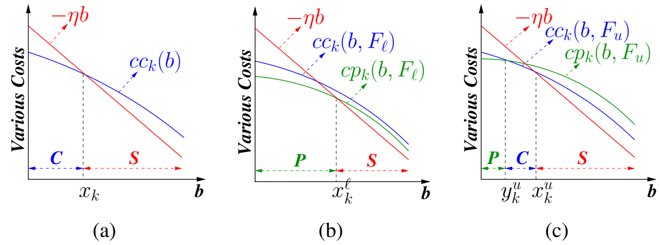


Fig. 2. Illustration of the threshold property. S, P and C in the figure represents the actions stop, probe and continue, respectively. (a) $\mathcal{S}_k$ is characterized by the threshold x_k. (b) and (c) depicts the stopping sets corresponding to the distributions F_ℓ and F_u, respectively, where $F_\ell \geq_{st} F_u$.

From Lemma 2-(i), we know that the thresholds x_k and x_k^ℓ are decreasing with k. The main result in the next section (Theorem 2 and 3) is to show that these thresholds are, in fact, equal (i.e., $x_k = x_{k+1}$ and $x_k^\ell = x_{k+1}^\ell$).

B. Stopping Sets: Stage Independence Property

In Lemma 2-(i) we have already shown that $\mathcal{S}_k \subseteq \mathcal{S}_{k+1}$, and $\mathcal{S}_k^\ell \subseteq \mathcal{S}_{k+1}^\ell$. In this section we will prove the inclusion in the other direction, leading to the result that the stopping sets are identical across the stages. Towards this end we will define the *optimal stopping/probing set* $\mathcal{Q}_k^\ell$ as, for $1 \leq k \leq N-1$,

$$\mathcal{Q}_k^\ell \quad := \quad \Big\{ b : \min\{-\eta b, cp_k(b, F_\ell)\} \leq cc_k(b, F_\ell) \Big\}. \tag{13}$$

From (10) it follows that $\mathcal{Q}_k^\ell$ is, for a given distribution F_ℓ, the set of b such that at state (b, F_ℓ) it is better to either stop or probe than to continue. From the definition of the sets $\mathcal{S}_k^\ell$ and $\mathcal{Q}_k^\ell$ (in (12) and (13), respectively) it immediately follows that $\mathcal{S}_k^\ell \subseteq \mathcal{Q}_k^\ell$. Also from Lemma 2-(ii) we know that $\mathcal{S}_k^\ell \subseteq \mathcal{S}_k$. However it is not immediately clear how the sets $\mathcal{Q}_k^\ell$ and $\mathcal{S}_k$ are ordered. Under the assumption that $\mathcal{F} = \{F_\ell : \ell \in \mathcal{L}\}$ is *totally stochastically ordered* (to be defined next), we show that $\mathcal{S}_k \subseteq \mathcal{Q}_k^\ell$ (Lemma 5), which plays a key role while proving our main theorems.

Assumption 1 (Total Stochastic Ordering Property of $\mathcal{F}$): From here on, we will assume that $\mathcal{F}$ is *totally stochastically ordered* meaning, any two distributions from $\mathcal{F}$ are always stochastically ordered. Formally, if $F_\ell, F_u \in \mathcal{F}$ then either $F_\ell \geq_{st} F_u$ or $F_u \geq_{st} F_\ell$. We further assume the existance of a *minimal distribution* $F_m \in \mathcal{F}$ such that for every $F_\ell \in \mathcal{F}$ we have $F_\ell \geq_{st} F_m$. ■

Note that such a restriction on $\mathcal{F}$ is not very stringent, in the sense that the set of distributions arising in our local forwarding problem in Section II, $\Big\{F_\ell : \ell \in \mathcal{L}\Big\}$, is totally stochastically ordered.

Before proceeding to our main theorems, we need the following results:

Lemma 4: Suppose $\mathcal{S}_k \subseteq \mathcal{Q}_k^\ell$ for some $k = 1, 2, \cdots, N-1$, and some F_ℓ, then for every $b \in \mathcal{S}_k$ we have $J_k(b, F_\ell) = J_N(b, F_\ell)$.

Proof: See [7, Proof of Lemma 6] ■

Lemma 5: For $1 \leq k \leq N-1$ and for any $F_\ell \in \mathcal{F}$ we have $\mathcal{S}_k \subseteq \mathcal{Q}_k^\ell$.

Proof Outline: For a formal proof see [7, Appendix VIII-D]. We only provide an outline here.

Step 1: First we show that if there exists an F_u such that, for $1 \leq k \leq N-1$, $\mathcal{S}_k \subseteq \mathcal{Q}_k^u$ (thus satisfying the hypothesis in Lemma 4), and if $F_\ell \geq_{st} F_u$ then $\mathcal{S}_k \subseteq \mathcal{Q}_k^\ell$. Lemma 4 and the total stochastic ordering of $\mathcal{F}$ are required for this part.

Step 2: Next we show that the minimal distribution F_m satisfies, for every $1 \leq k \leq N-1$, $\mathcal{S}_k \subseteq \mathcal{Q}_k^m$. The proof is completed by recalling that $F_\ell \geq_{st} F_m$ for every $F_\ell \in \mathcal{F}$ and then using in *Step 1*, F_m in the place of F_u. ■

The following are the main theorems of this section:

Theorem 2: For $1 \leq k \leq N-2$, $\mathcal{S}_k = \mathcal{S}_{k+1}$.

Proof: From Lemma 2-(i) we already know that $\mathcal{S}_k \subseteq \mathcal{S}_{k+1}$. Here, we will show that $\mathcal{S}_k \supseteq \mathcal{S}_{k+1}$. Fix a $b \in \mathcal{S}_{k+1} \subseteq \mathcal{S}_{k+2}$. From Lemma 5 we know that $\mathcal{S}_{k+1} \subseteq \mathcal{Q}_{k+1}^\ell$ and $\mathcal{S}_{k+2} \subseteq \mathcal{Q}_{k+2}^\ell$, for every F_ℓ. Then, applying Lemma 4 we can write, $J_{k+1}(b, F_\ell) = J_{k+2}(b, F_\ell) = J_N(b, F_\ell)$. Thus,

$$\begin{aligned} cc_{k+1}(b) &= \tau + \mathbb{E}_A\Big[J_{k+2}(b, F_{L_{k+2}})\Big] \\ &= \tau + \mathbb{E}_A\Big[J_{k+1}(b, F_{L_{k+1}})\Big] \\ &= cc_k(b) \end{aligned}$$

Now, since $b \in \mathcal{S}_{k+1}$ we have $-\eta b \leq cc_{k+1}(b) = cc_k(b)$ which implies that $b \in \mathcal{S}_k$. ■

Remark: It is worth pointing out the parallels and the differences of the above result with that in [13, Section 4.4, pp-165], where it is shown that the *one-step-look ahead* rule is optimal, but for the case where the rewards are exactly revealed. There, as in our Lemma 4, the key idea is to show that the cost-to-go functions, at every stage k, are identical for every state within the stopping set. For our case, to apply Lemma 4, it was further essential for us to prove Lemma 5 showing that for every F_ℓ, $\mathcal{S}_k \subseteq \mathcal{Q}_k^\ell$. Now, for $\delta = 0$, Lemma 5 trivially holds since at (b, F_ℓ) it is always optimal to probe (so that $\mathcal{Q}_k^\ell = \Re_+$). Further, $\delta = 0$ can be thought of as the case where the rewards are exactly revealed. Thus, Theorem 2 can be considered as a generalization of the result in [13, Section 4.4] for the case, $\delta > 0$. ■

Theorem 3: For $1 \leq k \leq N-2$ and for any F_ℓ, $\mathcal{S}_k^\ell = \mathcal{S}_{k+1}^\ell$.

Proof: Similar to the proof of Theorem 2, see [7, Appendix VIII-E]. ■

C. Optimal Probing Sets

Similar to the stopping set $\mathcal{S}_k^\ell$, one can define an *optimal probing set* $\mathcal{P}_k^\ell$ as, the set of all b such that at (b, F_ℓ) it is better to probe than to stop or continue, i.e.,

$$\mathcal{P}_k^\ell := \Big\{b : cp_k(b, F_\ell) \leq \min\{-\eta b, cc_k(b, F_\ell)\}\Big\}. \quad (14)$$

In our numerical work we observed that, similar to the stopping set, the probing set was characterized by an upper bound y_k^ℓ as illustrated in Fig. 2(b) (where $y_k^\ell = x_k^\ell$) and 2(c). At the time of this writing we have not proved such a result. However, we strongly believe that it is true and make the following conjecture,

Conjecture 1: For $1 \leq k \leq N-1$, for any F_ℓ, if $b_2 \in \mathcal{P}_k^\ell$ then for any $b_1 < b_2$ we have $b_1 \in \mathcal{P}_k^\ell$. ■

The above conjecture, along with Lemma 5 (where we have shown that, for any F_ℓ, $\mathcal{S}_k \subseteq \mathcal{Q}_k^\ell$), is the reason for the illustration in Fig. 2(b) (where $y_k^\ell = x_k^\ell$) to not contain a region where it is optimal to continue. Unlike x_k^ℓ, the thresholds y_k^ℓ are stage dependent. In fact, from our numerical work, we observe that y_k^ℓ are increasing with k.

D. Policy Implementation

To summarize, due to Theorem 1, the stopping sets $\mathcal{S}_k$ and $\mathcal{S}_k^\ell$ ($F_\ell \in \mathcal{F}$ and $k = 1, 2, \cdots, N-1$) are characterized by lower bounds x_k and x_k^ℓ. In Theorem 2 and 3 we proved that these thresholds are equal for every k. Hence it is sufficient to compute only x_{N-1} and x_{N-1}^ℓ, thus simplifying the overall computation of the optimal policy. Further, if Conjecture 1 is true, then the upper bounds y_k^ℓ are sufficient to characterize the probing sets $\mathcal{P}_k^\ell$.

The source, after computing these thresholds, operates in the following manner: It initially sets $b = -\infty$. At stage $k = 1, 2, \cdots, N-1$, with the state being (b, F_ℓ), (**1**) if $b > x_{N-1}^\ell$ then stop and forward the packet to the probed relay, (**2**) if $b < y_k^\ell$ then probe the unprobed relay and update the best reward to $b' = \max\{b, R_\ell\}$. If $b' > x_{N-1}$ stop, otherwise continue to wait for the next relay, (**3**) otherwise, continue to wait for the next relay to wake-up, at which instant choose, between F_ℓ and $F_{L_{k+1}}$, whichever is stochastically greater while putting the other unprobed relay to sleep.

If the decision process enters the last stage N then the source simply compares the "cost of stopping, $-\eta b$," with the "cost of probing and then stopping, $cp_N(b, F_\ell)$," and chooses the action whichever has minimum cost.

V. Unrestricted Class Π: An Informal Discussion

In this section we will consider the complete class Π. Here we will only informally discuss the possible structure of the optimal policy within Π. Formal analysis is still in progress at the time of this writing.

Recall that a policy within Π, at stage k, is in general allowed to take decisions based on the entire history. Formally, at stage k, let $b_k = \max\{R_{L_i} : i \leq k, \text{ relay } i \text{ has been probed}\}$ and $\mathcal{F}_k = \{F_{L_i} : i \leq k, \text{ relay } i \text{ is unprobed}\}$, then the decision at stage k is based on $(b_k, \mathcal{F}_k)$. Thus the set of all possible states (State Space SS_k) at stage k is

$$\begin{aligned} SS_k = \Big\{&(b, \mathcal{G}) : b \in \Re_+, \mathcal{G} = \{G_1, G_2, \cdots, G_j\}, \\ &0 \leq j \leq k, G_i \in \mathcal{F}, 1 \leq i \leq j\Big\}. \end{aligned} \quad (15)$$

Again the actions available are stop, probe and continue. Further, if the action is to probe then one has to decide which relay to probe, among the several ones awake at stage k.

A. Discussion on the Last Stage N

Consider the scenario where the decision process enters the last stage N. Given the best reward value b, among the relays that have been probed, and the set $\mathcal{G}$ of reward distributions of the unprobed relays, the source has to decide whether to stop or probe a relay (note that there is no continue action available at the last stage). This decision problem is similar to the one studied by Chaporkar and Proutiere in [11] in the context of channel selection, which we briefly describe now. Given a set of channels with different channel gain distributions, a transmitter has to choose a channel for its transmissions. The transmitter can probe a channel to know its channel gain. Probing all the channels will enable the transmitter to select the best channel but at the cost of reduced effective transmission time within the coherence period. On the other hand, probing only a few channels may deprive the transmitter of the opportunity to transmit on a better channel. The transmitter is interested in *maximizing* its *throughput* within the coherence period, which is analogous to the *cost* (combination of delay and effective reward, see (2)) in our case, which we are trying to *minimize*.

The authors in [11], for their channel probing problem, prove that the one-step-look-ahead (OSLA) rule is optimal. Thus, given the channel gain of the best channel (among the channels probed so far) and a collection of channel gain distributions of the unprobed channels, it is optimal to stop and transmit on the best channel if and only if the throughput obtained by doing so is greater than the average throughput obtained by probing any unprobed channel and then stopping (i.e., transmitting on the new-best channel). Further they prove that if the set of channel gain distributions is totally stochastically ordered (see Assumption 1), then it is optimal to probe the channel whose distribution is stochastically largest among all the unprobed channels. Applying the result of [11] to our model we can conclude that OSLA is optimal once the decision process enters the last stage N. Thus given a state, $(b, \mathcal{G})$, at stage N it is optimal to stop if the cost of stopping is less than the cost of probing any distribution from $\mathcal{G}$ and then stopping. Otherwise it is optimal to probe the stochastically largest distribution from $\mathcal{G}$.

B. Discussion on Stages $k = N-1, N-2, \cdots, 1$

For the other stages $k = N-1, N-2, \cdots, 1$, one can begin by defining the stopping sets $\mathcal{S}_k$, $\mathcal{S}_k^{\mathcal{G}}$ and stopping/probing set $\mathcal{Q}_k^{\mathcal{G}}$ analogous to the ones in (11), (12) and (13). Note that this time we need to define $\mathcal{S}_k^{\mathcal{G}}$ and $\mathcal{Q}_k^{\mathcal{G}}$ for a set of distributions $\mathcal{G}$ unlike in the earlier case where we had defined these sets only for a given distribution F_ℓ. We conjecture that it is possible to show the results analogous to the ones in Section IV, namely Theorem 2 and 3 where we prove that the stopping sets are identical for every stage k. Formal analysis is still in progress at the time of writing this paper. However, in the numerical results section while performing value iteration we observed that our conjecture is true, at least for the example considered.

Finally, from the discussion in the previous sub-section we know that, at stage N, suppose it is optimal to probe when the state is $(b, \mathcal{G})$ then it is best to probe the stochastically largest distribution from $\mathcal{G}$. We also conjecture that such a result will hold for every stage k, which is true for the numerical example considered in Section VI.

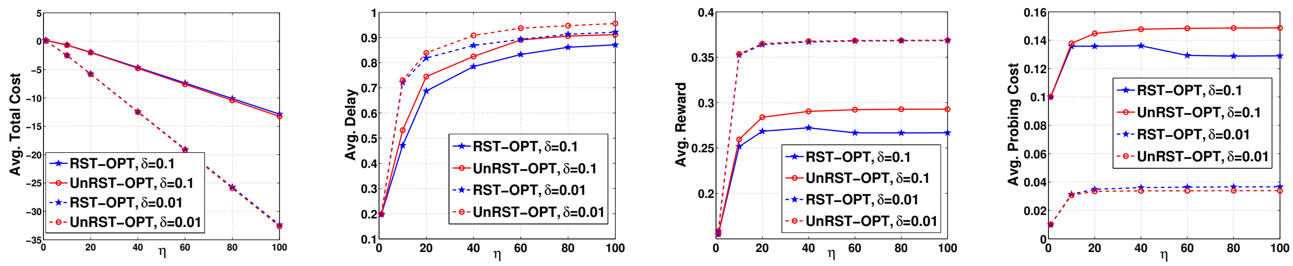
VI. Numerical Results

The optimal policy within the restricted class (Sections III and IV) is allowed to keep only one unprobed relay awake in addition to the best probed relay, while within the unrestricted class (Section V), the optimal policy can keep all the unprobed relays awake. We will refer to the former policy as RST-OPT (to be read as, ReSTricted-OPTimal) and the latter as UnRST-OPT (for UnReSTricted-OPTimal). In this section we will compare the performance of RST-OPT against UnRST-OPT. First we will briefly describe the relay selection example considered for the numerical work.

Recall the local forwarding problem described in Section II. The source and sink are seperated by a distance of $v_0 = 10$ unit (see Fig. 1). The radius of the communication region is set to 1 unit. There are $N = 5$ relays within the forwarding region $\mathcal{L}$. These are uniformly located within $\mathcal{L}$. To enable us to perform value iteration (i.e., recursively solve the Bellman equation to obtain optimal value and the optimal policy), we discretize the forwarding region $\mathcal{L}$ by considering a grid of 20 equally spaced points within $\mathcal{L}$ and then rounding the location of each relay to the respective closest point. Recall the reward expression from (1). We have fixed, $\Gamma N_0 = 1$, $\beta = 2$ and $a = 0.5$. The distribution of $|H_\ell|^2$ is truncated exponential with mean 1. Finally we normalize the reward to take values within the interval $[0, 1]$ and then quantize it to one of the 100 equally spaced points within $[0, 1]$. The inter-wake-up times $\{U_k\}$ are exponential with mean $\tau = 0.2$.

In Fig. 3(a) we plot the average total cost (see (2)), incurred by RST-OPT and UnRST-OPT, as a function of the Lagrange multiplier η for two different values of the probing cost δ (namely $\delta = 0.1$ and $\delta = 0.01$). The total cost is decreasing with η. First, observe that the total cost incurred, by either policy, for $\delta = 0.01$ is smaller than for $\delta = 0.1$. This is because, when the probing cost is smaller, each of the policy will end up probing more relays, thus accruing a larger reward and yielding a lower cost. Next, since UnRST-OPT is optimal over a larger class of policies, we know that, for a given δ, UnRST-OPT should incur a smaller cost than RST-OPT. However, interestingly from the plot we observe that the difference between the two costs is small. Also, from the figure, note that the cost difference for $\delta = 0.01$ is smaller than that for $\delta = 0.1$. This is because, as the probing cost is decreased, the two policies will start behaving identically by probing most of the relays until they stop. Finally, when there is no cost for probing, i.e., for $\delta = 0$, we expect the two policies to be identical.

In Fig. 3(b), 3(c), and 3(d) we plot the individual components (namely delay, reward and probing cost, respectively) of the total cost, as a function of η. All the curves in these figures are increasing with η, except for $\delta = 0.1$ where the average reward and probing cost of RST-OPT shows a small decrease when η varies from 40 to 60. As η decreases we observe, from Fig. 3(b), that all the average delay curves converge to 0.2 (recall that $\tau = 0.2$ is the average time until the first relay wakes up). This is because, for small values of η, the



(a) (b) (c) (d)

Fig. 3. (a) Average total cost vs. Lagrange multiplier η. (b), (c) and (d): Individual components of the total cost in Fig. 3(a) as functions of η. (b) Avg. Delay (c) Avg. Reward and (d) Avg. Probing Cost.

source values the delay more (recall (2)) and hence always forwards to the relay that wakes up first, irrespective of its reward. For the same reason the average probing cost curves in Fig. 3(d) converge to their respective δ value which is the cost for probing a single relay. Similarly the average reward in Fig. 3(c) converges to the average reward of the first relay.

Finally on the computational complexity of both the policies. To obtain the policy UnRST-OPT we had to recursively solve the Bellman equations for every stage k and every possible state at stage k, starting from the last stage N. The total number of all possible states at stage k, i.e., the cardinality of the state space SS_k in (15), grows exponentially with the cardinality of $\mathcal{F}$ (assuming that $\mathcal{F}$ is discrete like in our numerical example). It also grows exponentially with the stage index k. While in contrast for computing RST-OPT, since within the restricted class, at any time, only one unprobed relay is kept awake, the state space size grows linearly with the cardinality of $\mathcal{F}$. Further, the size of the state space does not grow with k. From our analysis in Section IV we know that the stopping sets are threshold based and moreover the thresholds (x_k and $\{x_k^\ell : F_\ell \in \mathcal{F}\}$) are stage independent, so that these thresholds have to be computed only once (namely for stage $N-1$), further reducing the complexity of RST-OPT.

RST-OPT can also be regarded as *energy efficient* in the sense that it keeps only one unprobed relay (in addition to the best probed one) awake while instructing the other unprobed ones to switch to a low power OFF state (i.e., sleep state). In contrast, UnRST-OPT operates by keeping all the relays, that have woken up thus far, awake. The close to optimal performance of RST-OPT, its computational simplicity and energy efficiency is motivating us to apply RST-OPT at each hop enroute to the sink in a large sensor network and study its end-to-end performance, which is a part of our ongoing work.

VII. Conclusion

We considered the sequential decision problem of choosing a relay node for the next hop, by a forwarding node, in the presence of random sleep-wake cycling. In the model, we have incorporated the energy cost of probing a relay to learn its channel gain. We have analysed a restricted class of policies where any policy is allowed to keep at most one unprobed relay awake, in addition to the best probed relay. The optimal policy for the restricted class (RST-OPT) was characterized in terms of, the stopping sets and the stopping/probing sets. First, we showed that the stopping sets are threshold in nature. Further, we proved that the thresholds, characterizing the stopping sets, are stage independent, thus simplifying the computation of RST-OPT in comparison with the global optimal, UnRST-OPT (whose decisions at any stage is based on the entire history). Numerical work confirmed that the performance of RST-OPT is close to that of UnRST-OPT.

References

[1] J. Kim, X. Lin, and N. Shroff, "Optimal Anycast Technique for Delay-Sensitive Energy-Constrained Asynchronous Sensor Networks," in *INFOCOM 2009. The 28th Conference on Computer Communications. IEEE*, April 2009.

[2] K. P. Naveen and A. Kumar, "Tunable Locally-Optimal Geographical Forwarding in Wireless Sensor Networks with Sleep-Wake Cycling Nodes," in *INFOCOM 2010. The 29th Conference on Computer Communications. IEEE*, March 2010.

[3] D. P. Bertsekas and J. N. Tsitsiklis, "An Analysis of Stochastic Shortest Path Problems," *Mathematics of Operations Research*, vol. 16, 1991.

[4] K. P. Naveen and A. Kumar, "Relay Selection for Geographical Forwarding in Sleep-Wake Cycling Wireless Sensor Networks," *IEEE Transactions on Mobile Computing*, accepted, to appear.

[5] P. Thejaswi, J. Zhang, M. O. Pun, H. Poor, and D. Zheng, "Distributed Opportunistic Scheduling with Two-Level Probing," *IEEE/ACM Transactions on Networking*, vol. 18, no. 5, October 2010.

[6] S. Karlin, *Stochastic Models and Optimal Policy for Selling an Asset*. Stanford University Press, Stanford, 1962.

[7] K. P. Naveen and A. Kumar, "Optimal Relay Selection with Channel Probing in Wireless Sensor Networks," Dept. of E.C.E., IISc, Bangalore, Tech. Rep. ARXIV Report Number 1107.5778, 2011. [Online]. Available: http://arxiv.org/abs/1107.5778

[8] K. Akkaya and M. Younis, "A Survey on Routing Protocols for Wireless Sensor Networks," *Ad Hoc Networks*, vol. 3, pp. 325–349, 2005.

[9] M. Zorzi and R. R. Rao, "Geographic Random Forwarding (GeRaF) for Ad Hoc and Sensor Networks: Multihop Performance," *IEEE Transactions on Mobile Computing*, vol. 2, pp. 337–348, 2003.

[10] J. Xu, B. Peric, and B. Vojcic, "Performance of Energy-Aware and Link-Adaptive Routing Metrics for Ultra Wideband Sensor Networks," *Mob. Netw. Appl.*, vol. 11, no. 4, pp. 509–519, 2006.

[11] P. Chaporkar and A. Proutiere, "Optimal Joint Probing and Transmission Strategy for Maximizing Throughput in Wireless Systems," *IEEE Journal on Selected Areas in Communications*, vol. 26, no. 8, pp. 1546–1555, October 2008.

[12] T. S. Ferguson, *Optimal Stopping and Applications*. Dept. of Mathematics, UCLA. [Online]. Available: http://www.math.ucla.edu/~tom/Stopping/Contents.html

[13] D. P. Bertsekas, *Dynamic Programming and Optimal Control, Vol. I*. Athena Scientific, 2005.

[14] W. Stadje, "An Optimal Stopping Problem with Two Levels of Incomplete Information," *Mathematical Methods of Operations Research*, vol. 45, pp. 119–131, 1997.

[15] D. Stoyan, *Comparison Methods for Queues and other Stochastic Models*. John Wiley & Sons, New York, 1983.