ACM Transactions on Sensor Networks, accepted for publication

Relay Selection with Channel Probing in Sleep-Wake Cycling Wireless Sensor Networks

K.P. NAVEEN, Indian Institute of Science
ANURAG KUMAR, Indian Institute of Science

In geographical forwarding of packets 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 to forward the packet towards the sink. Each relay is associated with a "reward" that summarizes the benefit of forwarding the packet through that relay. We seek a solution to this local problem, the idea being that such a solution, if adopted by every node, could provide a reasonable heuristic for the end-to-end forwarding problem. Towards this end, we propose a local *relay selection problem* comprising a forwarding node and a collection of relay nodes, with the relays waking up sequentially at random times. At each relay wake-up instant the forwarder can choose to *probe* a relay to learn its reward value, based on which the forwarder can then decide whether to *stop* (and forward its packet to the chosen relay) or to *continue* to wait for further relays to wake-up. The forwarder's objective is to select a relay so as to minimize a combination of waiting-delay, reward and probing cost. The local decision problem can be considered as a variant of the asset selling problem studied in the operations research literature. We formulate the local problem as a Markov decision process (MDP) and characterize the solution in terms of *stopping sets* and *probing sets*. We prove results illustrating the structure of the stopping sets, namely, the (lower bound) threshold and the stage-independence properties. Regarding the probing sets, we make an interesting conjecture that these sets are characterized by upper bounds. Through simulation experiments we provide valuable insights into the performance of the optimal local forwarding and its use as an end-to-end forwarding heuristic.

Additional Key Words and Phrases: Wireless sensor networks, sleep-wake cycling, geographical forwarding, stopping sets, Markov decision processes, stochastic ordering, asset selling problem

ACM Reference Format:
K.P. Naveen and Anurag Kumar, 2014. Relay Selection with Channel Probing in Sleep-Wake Cycling Wireless Sensor Networks. *ACM Trans. Sensor Netw.* V, N, Article A (January YYYY), 38 pages.
DOI:http://dx.doi.org/10.1145/0000000.0000000

1. INTRODUCTION

Consider a wireless sensor network deployed for the detection of *rare events*, e.g., forest fires, human intrusion in border areas, etc. In these networks, since the events of interest are rare, continuous monitoring by the nodes is unnecessary. Instead, the nodes can conserve their battery power by *sleep-wake cycling*, whereby they alternate

This work is an extended version of our conference paper [Naveen and Kumar 2012], presented at WiOpt 2012, held in Paderborn, Germany.

This work was supported in part by the Indo-French Centre for the Promotion of Advanced Research (IFCPAR Project 4000-IT-1), and in part by the Department of Science and Technology (DST) via a J.C. Bose Fellowship.

Authors' addresses: K.P. Naveen, (current address) INRIA Saclay, Palaiseau 91120, France (email: naveenkp@inria.fr); Anurag Kumar, Department of Electrical Communication Engineering, Indian Institute of Science, Bangalore 560012, India (email: anurag@ece.iisc.ernet.in).

DOI:http://dx.doi.org/10.1145/0000000.0000000

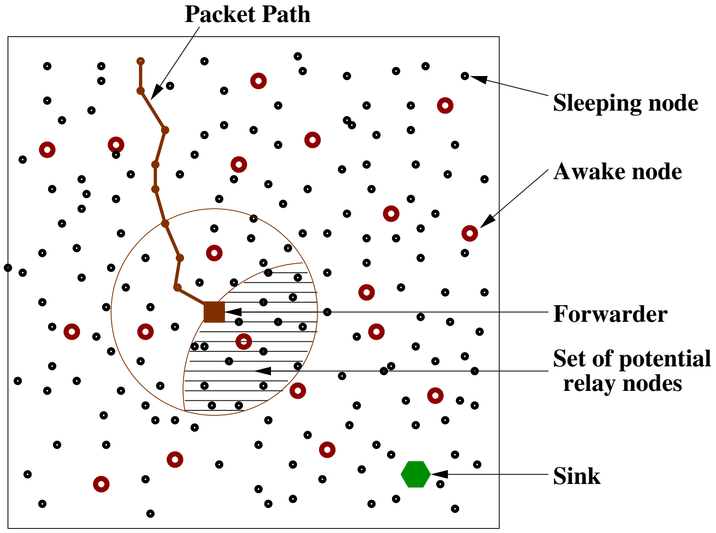


Fig. 1. A snap-shot of a packet being forwarded to the sink node (green hexagon) through a sleep-wake cycling network. The square node (labeled as forwarder) is the current custodian of the packet. The nodes in the hatched area are the set of potential relays for the forwarder.

between an ON state and a low power OFF state ([Abrardo et al. 2013; Guo et al. 2009; Liu et al. 2007]). We are interested in low duty-cycle, *asynchronous* sleep-wake cycling where the point processes of wake-up instants of the nodes are not synchronized [Li et al. 2014; Carrano et al. 2014]. We further consider a setting where the nodes are not aware of the sleep schedules of their neighbors [Naveen and Kumar 2013; Kim et al. 2011]. Although, it is possible for the nodes to learn their neighbors' sleep schedules through an initial configuration phase or while forwarding a packet, but, since the events are rare, such learned data would become stale for the next forwarding instant as the nodes' clocks would have randomly drifted. Moreover, addition of new nodes (fresh deployment) or deletion of some existing ones (due to battery drainage) will add to the uncertainty of the times at which the successive neighbors will wake-up.

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 (as illustrated in Fig. 1), to a control center (*sink*) where appropriate action could be taken. Although it is possible for multiple nodes to have detected the event, to avoid flooding, which causes extensive contentions and collisions in the network (referred to as the *broadcast storm problem* [Tonguz et al. 2006; Tseng et al. 2002]), we consider generating only one alarm packet per event. This can be accomplished by allowing the detecting nodes to collaborate among themselves to choose a packet generating node [Kumar et al. 2010].

Now, since the network is sleep-wake cycling, a forwarding node (i.e., a node currently holding the alarm packet) has to wait for its neighbors to wake-up before it can choose one for serving as the next hop relay. As successive potential relays wake up (and then go back to sleep), the forwarding node has the *sequential decision problem* of selecting one of them to forward the packet through, while balancing the trade-off between delay in forwarding and some measure of the *quality* of the relay (e.g., the progress it make towards the sink [Naveen and Kumar 2010], or the channel quality to this relay). With this local trade-off in mind, the end-to-end problem become one of minimizing a combination of total average end-to-end delay and some global metric such as the average hop count, 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 [Bertsekas and Tsitsiklis 1991], for which the distributed Bellman-Ford algorithm (e.g., the LOCAL-OPT algorithm proposed by Kim et al. in [Kim et al. 2011]) can be used to obtain the optimal solution. However, 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. Furthermore, such global configuration would need to be performed each time there is a change in the network topology, such as due to node failures, or long time scale variations in the propagation characteristics.

The focus of our research is instead towards designing *simple forwarding rules* that use only the *local information* available at a forwarding node. In our own earlier work in this direction [Naveen and Kumar 2010; Naveen and Kumar 2013], we formulated the local forwarding problem as one of minimizing the one-hop forwarding delay subject to a constraint on the reward offered by the chosen relay. The reward associated with a relay is a function of the transmission power and the progress towards the sink made by the packet when forwarded via that relay. We considered two variations of the problem, one in which the number of potential relays in the forwarding nodes neighborhood is known [Naveen and Kumar 2010], and the other in which only a probability mass function of the number of potential relays is known [Naveen and Kumar 2013]. In each case, we derived the structure of the optimal policy. Further, through simulation experiments we found that, in some regimes 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. [Kim et al. 2011]), thus providing additional support for the approach of utilizing local forwarding rules, albeit suboptimal.

In our earlier work, however, we assume that the gain of the wireless communication channel between the forwarding node and a relay is a deterministic function of the distance between the two, whereas, in practice, due to the phenomenon called *shadowing*, the channel gain at a given distance from the forwarding node is not a constant, but varies spatially over points at the same distance (the statistical variation being typically modeled as log normally distributed [Rappaport 2001]). In addition to not being just a function of distance, the path-loss between a pair of locations has long term variation with time; in a forest, for example, this would be due to seasonal variations in the foliage. Therefore, in each instance that a node gets custody of a packet, the node has to send probe packets to determine the channel gain to relay nodes that wake up, and thereby "offer" to forward the packet. Such probing incurs additional cost (for instance, see [Thejaswi et al. 2010] where probing allows the transmitter to obtain a finer estimate of the channel gain). Hence, "to probe" or "not to probe" can itself become a part of the decision process. In the current work we incorporate these features (namely, channel probing and the associated power cost) while choosing a relay for the next hop, leading to an interesting variant of the *asset selling problem* ([Bertsekas 2005, Section 4.4], [Karlin 1962]), studied in the operations research literature.

We emphasize that in this work we are addressing the problem of *resource (in particular, relay) allocation*; this is in contrast to the problem of medium access contention resolution that arises when several relays contend for the medium simultaneously, as in [Guo et al. 2009; Kim and Liu 2008; Liu et al. 2007; Zorzi and Rao 2003b]. Such contention does not arise in our case, since, due to low rate duty-cycling, the relays wake up sequentially in time rather than simultaneously. Further, in our case since the events are rare, with only one packet per event being generated, the possibility of contention between the forwarding nodes of two different alarm packets (e.g., in [Guo et al. 2009]) is also negligible.

Outline and Our Contributions: We will first fix the context by describing the mathematical model in Section 2, and then proceed to discuss the related work in Section 3. Sections 4 and 5 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 6 we will discuss the globally optimal, GLB-OPT, policy. Numerical and simulation results are presented in Section 7. Our main technical contributions are the following:

- We first characterize the optimal policy, RST-OPT, in terms of *stopping sets*, i.e., a subset of the state space in which the forwarder's optimal action is to stop and forward the packet. We prove that the stopping sets can be represented in terms of lower bound thresholds (Theorem 5.3).
- We further prove that the stopping sets are identical across the decision stages (Theorem 5.6 and 5.7). This result can be considered as a generalization of the *one-step-look ahead* rule (see the remark following Theorem 5.6).
- Through numerical work on the one-hop problem, we find that the performance of RST-OPT is close to that of GLB-OPT. This result is useful because, the sub-optimal RST-OPT is computationally much simpler than GLB-OPT. We have also conducted simulations to study the end-to-end performance of RST-OPT.

We will finally conclude in Section 8. For the sake of readability we have moved most of the proofs to the Appendix.

2. SYSTEM MODEL

We will describe the system model in the context of *geographical forwarding*, also known as location aware routing, [Akkaya and Younis 2005; Mauve et al. 2001]. 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. Since our objective is towards designing local forwarding rules, we assume that the forwarding region (see Fig. 2) of each node is nonempty (i.e., there are no *voids* in the network). This assumption can be justified by considering a sufficiently dense network so that the probability of void-occurrence is negligible. Thus, in this work we do not address the problem of routing around voids; algorithms such as GPRS (Greedy Perimeter Stateless Routing) [Karp and Kung 2000], GOAFR (Greedy Other Adaptive Face Routing) [Kuhn et al. 2008], etc., along with protocol proposals [Petrioli et al. 2014] are available in the literature addressing this issue.

Consider a forwarding node $\mathscr{F}$ located at v (see Fig. 2). The sink node is situated at v_0. Thus, the distance between $\mathscr{F}$ and the sink is $V = \parallel v - v_0 \parallel$ (we use $\parallel \cdot \parallel$ to denote the Euclidean norm). The *communication region* is the set of all locations where reliable exchange of *control messages* (transmitted using a low rate robust modulation technique on a separate control channel) can take place between $\mathscr{F}$ and a receiver, if any, at these locations. In Fig. 2 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 the *neighbors*.

Let $V_\ell = \parallel \ell - v_0 \parallel$ represent the distance of a location ℓ (which is a point in $\Re^2$) from the sink. Now define the *progress* of location ℓ as $Z_\ell = V - V_\ell$, which is simply the difference between the $\mathscr{F}$-to-sink and ℓ-to-sink distances. $\mathscr{F}$ is interested in forwarding the packet only to a neighbor within the *forwarding region* $\mathcal{L}$, which is defined as

$$\mathcal{L} = \Big\{ \ell \in \text{communication region} : Z_\ell \geq z_{min} \Big\} \tag{1}$$

where, $z_{min} > 0$ is the minimum progress constraint (see Fig. 2, where the hatched area is the forwarding region). The reason for using $z_{min} > 0$ in the definition of $\mathcal{L}$

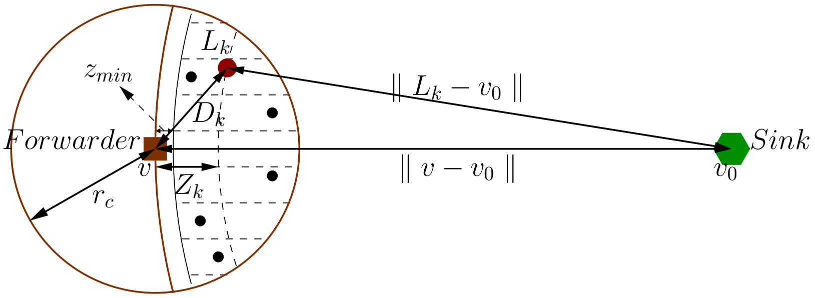


Fig. 2. The hatched area is the forwarding region $\mathcal{L}$. For $\ell \in \mathcal{L}$, the progress Z_ℓ is the difference between the forwarder-to-sink and ℓ-to-sink distances.

are: (1) practically this will ensure that a progress of at least z_{min} is made by the packet at each hop, and (2) mathematically this condition will allow us to bound the reward functions (to be defined sooner) to take values within an interval $[0, \overline{r}]$. Further, we assume that $\mathcal{L}$ is closed and bounded (the reason for imposing this condition will become clear in Section 5). Finally, we will refer to the nodes in the forwarding region as *relays*.

Sleep-Wake Process: Without loss of generality, we will assume that $\mathscr{F}$ receives an alarm packet at time 0 (from an upstream node; recall Fig. 1), which has to be forwarded to one of the relays. There are N relays that wake-up sequentially at the points of a Poisson process of rate $\frac{1}{\tau}$.[1] The wake-up times are denoted, $0 \leq W_1 \leq \cdots \leq W_N$. The relay waking up at the instant W_k is referred to as the k-th relay. Let $U_1 = W_1$ and $U_k = W_k - W_{k-1}$ ($k = 2, \cdots, N$) denote the *inter-wake-up time* between the k-th and the $(k-1)$-th relay. Thus, $\{U_k : k = 1, 2, \cdots, N\}$ are i.i.d. (independent and identically distributed) exponential random variables with mean τ.

Channel Model: Let $P_\ell(t)$ denote the transmission power required by $\mathscr{F}$ at time $t \geq 0$ to achieve an SNR (signal to noise ratio) constraint of Γ at some location ℓ, whose distance from $\mathscr{F}$ is more than d_{ref} (far-field reference distance beyond which the following expression will hold). We will consider the following standard model for $P_\ell(t)$ [Kumar et al. 2008; Tse and Viswanath 2005]:

$$P_\ell(t) = \frac{\Gamma N_0}{G_\ell(t)} \left(\frac{D_\ell}{d_{ref}} \right)^{\xi} \tag{2}$$

where, $D_\ell = \| \ell - v \|$ is the distance between $\mathscr{F}$ and ℓ, $G_\ell(t)$ is the random component of the channel gain between $\mathscr{F}$ and ℓ at time t, N_0 is the receiver noise variance, and ξ is the path-loss attenuation factor. We will assume that $d_{ref} \leq z_{min}$ so that $P_\ell(t)$ in (3) is the power required for any $\ell \in \mathcal{L}$. Also, for simplicity we will use Γ' to denote $\Gamma N_0 d_{ref}^{\xi}$.

Although $G_\ell(t)$ along with the path-loss, $(D_\ell/d_{ref})^{\xi}$, constitutes the gain of the channel at time t, for simplicity we will refer to $G_\ell(t)$ itself as the channel gain between $\mathscr{F}$ and the location ℓ. We will assume that the channel gain process $\{G_\ell(t) : t \geq 0\}$ is stationary and i.i.d. across ℓ. We will further assume that the channel coherence time is large so that the channels gains remain unchanged over the entire duration of the

[1] A practical approach for sleep-wake cycling is the *asynchronous periodic* process, where each relay i wakes up at the periodic instants $T_i + kT$ with $\{T_i\}$ being i.i.d. (independent and identically distributed) uniform on $[0, T]$ [Kim et al. 2011; Naveen and Kumar 2013]. Now, for large N if T scales with N such that $\frac{N}{T} \to \frac{1}{\tau}$, then the aggregate point process of relay wake-up instants converges to a Poisson process of rate $\frac{1}{\tau}$ [Cinlar 1975], thus justifying our Poisson process assumption.

decision process, i.e., in physical layer wireless terminology, we have a *slowly varying channel*. Thus, if G_ℓ denotes the random variable whose distribution is same as the marginal distribution of $\{G_\ell(t)\}$, then the marginal random variable of $\{P_\ell(t)\}$,

$$P_\ell = \frac{\Gamma'}{G_\ell} {D_\ell}^{\xi}, \tag{3}$$

is a representation of the power required to forward the packet to a relay at ℓ, irrespective of the time at which the relay was probed during the decision process. Hence, in the sequel we will remove the time variable from our notation and work only with the marginal random variables.

Remark: Regarding the channel gains being i.i.d., since the randomness in the channel is spatially correlated across relays [Agrawal and Patwari 2009], if two locations ℓ and u are very close then the corresponding gains, G_ℓ and G_u, will not be independent; a minimum separation between the receivers is required for the gains to be statistically independently. Thus, our assumption of independence between the channel gains across the relays requires that the relays should not be close to each other, or, equivalently, the relay density should not be large. We will assume that this physical property holds, and, thus, proceed with the technical assumption that the channel gains are i.i.d.

Reward Structure: Finally, combining progress, Z_ℓ, and power, P_ℓ, we define the reward associated with a location $\ell \in \mathcal{L}$ as,

$$R_\ell = \frac{Z_\ell^a}{P_\ell^{(1-a)}} = \frac{Z_\ell^a}{(\Gamma' D_\ell^{\xi})^{(1-a)}} G_\ell^{(1-a)}, \tag{4}$$

where $a \in [0,1]$ is used to trade-off between Z_ℓ and P_ℓ. The reward varying inversely with P_ℓ is clear because it is advantageous to use low power to get the packet across; R_ℓ increasing with Z_ℓ promotes progress towards the sink while choosing a relay for the next hop. The channel gains, $\{G_\ell\}$, are non-negative; we will further assume that they are bounded above by g_{max}. These conditions along with $Z_\ell \geq z_{min}$ (which implies that $D_\ell \geq z_{min}$) and $\mathcal{L}$ is bounded (so that $Z_\ell \leq z_{max}$ for all $\ell \in \mathcal{L}$) will provide the following upper bound for the reward functions $\{R_\ell : \ell \in \mathcal{L}\}$:

$$\overline{r} \;=\; \frac{z_{max}^a}{(\Gamma' z_{min}^{\xi})^{(1-a)}} g_{max}^{(1-a)}.$$

Thus, the reward values lie within the interval $[0, \overline{r}]$.

Let F_ℓ represent the c.d.f. (cumulative distribution function) of R_ℓ (or, strictly speaking the marginal distribution of $R_\ell(t)$), and

$$\mathcal{F} \;=\; \Big\{ F_\ell : \ell \in \mathcal{L} \Big\} \tag{5}$$

denote the collection of all possible reward distributions. From (4), note that, given a location ℓ it is only possible to know the reward distribution F_ℓ. To know the exact reward R_ℓ, $\mathscr{F}$ has to *transmit probe packets* to learn the channel gain G_ℓ (we will formalize probing very soon).

Remark: The motivation for using the particular reward in (4) comes from our prior work [Naveen and Kumar 2013] where we have observed that the solution to our local problem, obtained using the above reward structure, provides an end-to-end performance (in terms of end-to-end delay vs. total power) that is comparable with the performance of the globally optimal solution proposed by [Kim et al. 2011]. However, it is important to note that all our analysis in the subsequent sections will follow through for more general functions of the channel gain, as long as the corresponding distribution set $\mathcal{F}$ satisfies the *total stochastic ordering* property discussed below.

Definition 2.1 (*Stochastic Ordering*). Given two distributions F_ℓ and F_u, F_ℓ is stochastically greater than F_u, denoted as $F_\ell \geq_{st} F_u$, if $1 - F_\ell(r) \geq 1 - F_u(r)$, for all r. Equivalently [Stoyan 1983], $F_\ell \geq_{st} F_u$ if and only if for every non-decreasing function $f : \Re \rightarrow \Re$, $\mathbb{E}_\ell[f(R_\ell)] \geq \mathbb{E}_u[f(R_u)]$ where the distributions of R_ℓ and R_u are F_ℓ and F_u, respectively. □

Definition 2.2 (*Total Stochastic Ordering*). $\mathcal{F}$ is said to be *totally stochastically ordered* if any two distributions from $\mathcal{F}$ are stochastically ordered. Formally, for any $F_\ell, F_u \in \mathcal{F}$ either $F_\ell \geq_{st} F_u$ or $F_u \geq_{st} F_\ell$. Further, if there exists a distribution $F_m \in \mathcal{F}$ such that for every $F_\ell \in \mathcal{F}$ we have $F_\ell \geq_{st} F_m$ then we say that $\mathcal{F}$ is *totally stochastically ordered with a minimum distribution*. □

The following result will be useful in our analysis later.

LEMMA 2.3. *The set of reward distributions $\mathcal{F}$ in (5), is totally stochastically ordered with a minimum distribution.*

PROOF. The channel gains, $\{G_\ell : \ell \in \mathcal{L}\}$, being identically distributed will be essential to show that $\mathcal{F}$ is totally stochastically ordered. Existence of a minimum distribution will require the assumption we had made earlier (in Section 2) that $\mathcal{L}$ is compact (closed and bounded). The complete proof is available in Appendix A.3. □

Relay Locations: We will assume that each of the N relays is randomly and mutually independently located in the forwarding region $\mathcal{L}$. Formally, let $L_1, L_2, \cdots, L_N$ denote the random relay locations, that are i.i.d. uniform over the forwarding set $\mathcal{L}$ (this assumption holds if the nodes are deployed according to a spatial Poisson process). Let L denote the uniform distribution over $\mathcal{L}$ so that the distribution of L_k is L (for $k = 1, 2, \cdots, N$).

Remark: For the sake of motivating the model we assume that the location distribution L is uniform. However, our analysis holds good for any other distribution.

Sequential Decision Problem: At time 0, $\mathscr{F}$ only knows that there are N relays in its forwarding set $\mathcal{L}$, but *does not a-priori know their locations, L_k, nor their channel gains, G_{L_k}*. When the k-th relay wakes up, we assume that its location L_k is revealed[2], using which (in (4)) the distribution F_{L_k} of the reward R_{L_k} can be known (since the channel gain distribution is known). However, if $\mathscr{F}$ wishes to learn the exact reward value R_{L_k}, it has to estimate the channel gain G_{L_k}. This is accomplished by transmitting additional *probe packets*, incurring a power cost of $\delta \geq 0$ units. Thus, when the k-th relay wakes up (referred to as *stage k*), given the set of previously probed and unprobed relays (i.e., the history), the following actions are available to $\mathscr{F}$:

- s: *stop* and forward the packet to a relay with the maximum reward (*best relay*) among the probed relays; with this action the decision process ends.
- c: *continue* to wait for the next relay to wake-up (average waiting time is τ); with this action the decision process enters stage $k + 1$.
- p: probe a relay from the set of all unprobed relays (provided there is at least one unprobed relay). The probed relay's channel gain, and hence its reward value is then revealed, allowing $\mathscr{F}$ to update the best relay. *After probing, the decision process is still at stage k and $\mathscr{F}$ has to again decide upon an action.*

Remark: Note that, we are allowing $\mathscr{F}$ to forward the packet only to a probed relay. This is because, knowing the channel gain (since probing reveals the channel gain),

[2] which can be accomplished by including the location information L_k within a control packet (sent using a low rate robust modulation technique, and hence, assumed to be error free) transmitted by the k-th relay upon waking up

$\mathscr{F}$ can then choose an appropriate power level (using (2)) for its transmission. Although, using advanced adaptive coding techniques it may be possible to transmit to an unprobed relay, but for simplicity we do not consider this option. Moreover, implementing such coding algorithms at the memory-constrained wireless nodes would be difficult in practice. Further, for the sake of analysis, we neglect the time taken for the exchange of control packets and the time taken to probe a relay to learn its channel gain. We argue that 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 and for the exchange of control packets.

At stage k, let b_k denote the reward of the best relay, and $\mathcal{F}_k$ be the vector of reward distribution of the unprobed relays, i.e., formally,

$$b_k = \max\Big\{R_{L_i} : i \leq k, \text{ relay } i \text{ has been probed}\Big\}, \tag{6}$$

$$\mathcal{F}_k = \Big(F_{L_i} : i \leq k, \text{ relay } i \text{ is unprobed}\Big). \tag{7}$$

We will regard $(b_k, \mathcal{F}_k)$ to be the state of the system at stage k. Note that, it is possible that until stage k no relay has been probed, in which case $b_k = -\infty$, or all the relays are probed so that $\mathcal{F}_k$ is empty. Whenever $\mathcal{F}_k$ is empty we will represent the state as simply b_k. Now we can define a forwarding policy π as follows:

Definition 2.4. A policy π is a sequence of mappings $(\mu_1, \mu_2, \cdots, \mu_N)$ where,

- for $k = 1, 2, \cdots, N-1$, $\mu_k(b_k, \mathcal{F}_k) \in \{\mathsf{s}, \mathsf{c}, \mathsf{p}\}$ and $\mu_k(b_k) \in \{\mathsf{s}, \mathsf{c}\}$, and
- $\mu_N(b_N, \mathcal{F}_N) \in \{\mathsf{s}, \mathsf{p}\}$ and $\mu_N(b_N) \in \mathsf{s}$.

Note that the action to continue is not available at the last stage N. Let Π denote the set of all policies. □

Remark: Thus, we are considering a scenario where the forwarder can base its decision by retaining (or recalling) the best probed relay (see (6)). This property will enable us to prove an additional structural result (in Section 5.2) that the optimal policy is characterized by *stage independent thresholds*, which is not possible if recalling is not allowed. However, for the latter case, the threshold property of the optimal policy (in Section 5.1) can still be deduced so that a threshold policy remains to be optimal, although it would be *stage dependent*. We will remark more on this in Section 5.2.

Now, for a policy $\pi \in \Pi$, the delay incurred, denoted D, is the time until a relay is chosen. Let R denote the reward offered by the chosen relay. Further, let M denote the total number of relays that were probed during the decision process. Then, recalling that δ is the probing cost, δM represents the total cost of probing. We would like to think of $(R - \delta M)$ as the *effective reward* achieved using policy π. Then, denoting $\mathbb{E}[\cdot]$ to be the expectation operator conditioned on using policy π, the problem we are interested in is the following:

$$\text{Minimize}_{\pi \in \Pi} \left(\mathbb{E}_\pi[D] - \eta\Big(\mathbb{E}_\pi[R] - \delta\mathbb{E}_\pi[M]\Big)\right), \tag{8}$$

where $\eta > 0$ is the coefficient used to tradeoff between delay and effective reward.

Note that, the coefficients η and δ in the above objective function will enable us to tradeoff between the various quantities (namely delay, reward and probing cost). For instance, a small value of η would result in an objective function which gives more weight to the delay term, $\mathbb{E}_\pi[D]$. Hence, the forwarding node, in view of minimizing delay, would simply probe and transmit to the relay that wakes up first, irrespective of its reward value. On the other hand, if η is large, the objective would be more in favor of minimize the effective reward, $(\mathbb{E}_\pi[R] - \delta\mathbb{E}_\pi[M])$. Thus, now the forwarder, targeting for a relay with a good reward value, would end up waiting for more relays to wake-up, while probing every relay if the probing cost δ is small, or cautiously probing only good relays if δ is large. Hence, a range of tradeoff can be obtained by varying η and δ, which is in general captured by the objective function in (8). We will discuss these tradeoffs in more detail while presenting the numerical results in Section 7.

Restricted Class $\overline{\Pi}$: Recall that the state at stage k is of the form $(b_k, \mathcal{F}_k)$ where $\mathcal{F}_k$ is the set of all unprobed relays. The size of $\mathcal{F}_k$ can vary from 0 (if all the k relays that have woken up thus far have been probed) to k (if none have been probed). Further, suppose the size of $\mathcal{F}_k$ is m $(0 < m \leq k)$ then $\mathcal{F}_k \in \mathcal{F}^m$ (the m times Cartesian product of $\mathcal{F}$) since the reward distribution of each unprobed relay can be any distribution from $\mathcal{F}$. Thus, the set of all possible states at stage k is large. Hence, for analytical tractability, we first consider (in Sections 4 and 5) solving the problem in (8) over a *restricted class* of policies, $\overline{\Pi} \subseteq \Pi$, where a policy is restricted to take decisions keeping only up to two relays awake $-$ one the best among all probed relays and other the best among the unprobed ones. Thus, the decision at stage k is based on (b_k, H_k) where H_k is the stochastically greatest distribution in $\mathcal{F}_k$. Later in Section 6 we will discuss the optimal policy within the *unrestricted class* of policies Π.

3. RELATED WORK

Although the motivation for our work comes from the context of geographical forwarding in WSNs, related literature on the local decision problem can be found from other topics as well, e.g., the problem of channel probing in wireless networks, and the asset selling problem studied by the operations research community. In this section we will discuss related work from all these topics.

Geographical forwarding and routing in wireless networks: The problem of choosing a next-hop relay usually arises in the context of *geographical forwarding*. As mentioned earlier, *geographical forwarding* [Akkaya and Younis 2005; Mauve et al. 2001] is a forwarding technique where the prerequisite is that the nodes know their respective locations as well as the sink's location. The method of geographical forwarding was already envisioned in the 80's in the context of routing in packet radio networks (PRNs) [Takagi and Kleinrock 1984; Hou and Li 1986]. One of the simplest geographical forwarding technique is the greedy algorithm where each node forwards to a neighbor in its communication region which makes maximum progress towards the sink. This greedy algorithm is referred to as the *MFR (Max Forward within Radius)* routing in [Takagi and Kleinrock 1984]. Akin to MFR is the *NFP (Nearest with Forward Progress)* proposed in [Hou and Li 1986] where a node with a positive progress, and closest to the transmitting node is chosen. A generalization of MFR and NFP routing is to randomly choose any neighbor which makes a positive progress towards the sink [Nelson and Kleinrock 1984].

More recently, there is work which considers applying geographical forwarding for routing in sleep-wake cycling networks [Hao et al. 2012] is a recent survey on this topic which includes some of the work we will discuss below (this survey paper includes one of our prior work on the topic [Naveen and Kumar 2010] in its list of references).

Authors in [Liu et al. 2007] propose a protocol named CMAC (Convergent MAC), using which a forwarding node chooses a relay whose normalized latency (which is the expected ratio of one-hop delay and progress) is more than a threshold r_0, where r_0 is chosen so as to minimize the expected latency. The Random Asynchronous Wakeup (RAW) protocol in [Paruchuri et al. 2004] also (heuristically) considers transmitting to the first node to wake-up that makes a progress of greater than a threshold. Interestingly, such a threshold policy is optimal for our basic model (see [Naveen and Kumar 2013, Section 6],[Naveen and Kumar 2010]).

Zorzi and Rao in [Zorzi and Rao 2003b] study a time slotted system where nodes follow geometric sleep-wake patterns, i.e., a node is active in a slot with probability p. For a greedy scheme, referred to as GeRaF (Geographical Random Forwarding), where a forwarding node chooses the awake neighbor closest to the sink, the authors obtain its multi-hop performance in terms of the average number of hops required as a function of distance to the sink. Energy and latency performance of GeRaF is studied by the same authors in [Zorzi and Rao 2003a]. In contrast to GeRaF, ExOR [Biswas and Morris 2005] uses a metric called ETX (Estimated Transmission Time), which is an estimate of the number of transmissions required to reach the destination, to choose a next-hop relay. Alternatively, the authors in [Ghadimi et al. 2014] propose an opportunistic routing algorithm (referred to as ORW) that uses EDC (Expected Number of Duty Cycled Wakeups) metric instead for forwarding; a version of ORW is studied in [So and Byun 2014] where in-network aggregation of packets is performed before forwarding. However, in all these work, including others [Ozen and Oktug 2014; Guo et al. 2009], the main focus is on the design of MAC for resolving contention, that could arise when multiple relays become active simultaneously. Such contentions do not arise in our model since we are assuming a low duty-cycle sleep-wake cycling network (so that the probability of more than one relay waking up simultaneously is very low, and hence can be safely neglected). Thus, ours is instead a problem of *resource (in particular, relay) allocation (or acquisition)* that arises when a collection of resources become available sequentially in time.

Application of control theory [Bertsekas 2005; Puterman 1994] for the problem of routing in sleep-wake cycling networks can also be found in the literature [Kim et al. 2011; Kim and Liu 2008]. However, as already mentioned in the introduction, the one of Kim et al. [Kim et al. 2011] is based on the Bellman-Ford algorithm, and hence requires a global pre-configuration phase for offline computation of an optimal forwarding policy. The algorithm in [Kim and Liu 2008] requires a central entity to choose a next-hop relay. Although the authors in [Kim and Liu 2008] propose a distributed implementation, but this requires a "priority update" phase where each node has to compute its priority to each of its neighboring node. In contrast to the above work, *our algorithm is completely online, with the forwarding node deciding, as and when the relays wake-up, whether or not to forward to a relay*. Also, we have incorporated an additional "probe" action into our formulation, which is not considered in any of the above work.

Channel probing in wireless networks: From practical standpoint, testbed experiments involving WSNs ([Kumar et al. 2010; Bhattacharya et al. 2013]) require estimating link quality measurements using known signals (probe packets), before the nodes can exchange any useful data. Thus, channel probing is an inherent feature of the wireless system. In wireless networks, models with channel probing are generally studied in the context of channel selection [Chaporkar and Proutiere 2008; Chang and Liu 2007]. For instance, the authors in [Chaporkar and Proutiere 2008] study the following problem: a transmitter, aiming to maximize its throughput, has to choose a channel for its transmissions, among several available ones. The transmitter, only knowing the channel gain distributions, has to send probe packets to learn the exact channel state information (CSI). Probing many channels yields a channel with a good gain but reduces the effective time for transmission within the channel coherence period. The problem is to obtain optimal strategies to decide when to stop probing and to transmit. An important difference with our work is that, in [Chaporkar and Proutiere 2008; Chang and Liu 2007], all the channel gain distributions are known a priori while in our present paper the reward distributions are revealed as and when the relays wake-up. We will discuss more about the work in [Chaporkar and Proutiere 2008] in Section 6.

Thejaswi et al. in [Thejaswi et al. 2010] consider a model where, initially only a coarse estimate of the channel gain is available to the transmitter, and the transmitter can choose to probe the channel a second time to get a finer estimate of the gain (and hence the rate at which it can transmit). The objective is to optimize the trade-off between the throughput gain obtained from the more accurate rate estimation and the resulting additional delay. The authors pose the problem as an optimal stopping problem and show that the optimal policy is characterized by two rate thresholds, such that it is optimal to probe if and only if the initial rate estimate lies between these thresholds. The thresholds are stage dependent, which is a consequence of the horizon length of their stopping problem being infinite. In general, for a finite horizon stopping problem the optimal policy would be stage dependent. For our problem, despite being a finite horizon one, we are able to show that certain stopping sets are identical across stages. This is due to the fact that we allow the best probed relay to stay awake.

Asset selling problem: Let us recall the objective in (8). Suppose the probing cost $\delta = 0$, then (8) will reduce to minimizing $(\mathbb{E}_\pi[D] - \eta\mathbb{E}_\pi[R])$. Further, when $\delta = 0$, since there is no advantage in not probing, an optimal policy is to always probe relays as they wake-up so that their reward value is immediately revealed to $\mathscr{F}$. Alternatively, if $\mathscr{F}$ is not allowed to exercise the option to not-probe a relay, then again the model reduces to the case where the relay rewards are immediately revealed as and when they wake-up. We have studied this particular case of our relay selection problem (which we will refer to as the *basic relay selection model*) in our earlier work [Naveen and Kumar 2013, Section 6],[Naveen and Kumar 2010], and this basic model can be shown to be equivalent to a basic version of the *asset selling problem* [Bertsekas 2005, Section 4.4], [Karlin 1962] studied in the operations research literature.

The basic asset selling problem comprises a seller (with some asset to sell) and a collection of buyers who are arriving sequentially in time. The offers made by the buyers are i.i.d. If the seller wishes to choose an early offer, then he can invest the funds received for a longer time period. On the other hand, waiting could yield a better offer, but with the loss of time during which the sale-proceeds could have been invested. The seller's objective is to choose an offer so as to maximize his final revenue (received at the end of the investment period). Thinking of the offer of a buyer as analogous to the reward of a relay, the seller's objective of maximizing revenue is equivalent to the forwarder's objective of minimizing a combination of delay and reward.

Over the years, several variants of the basic problem have been studied. For instance, Kang in [Kang 2005] has considered a model where a cost has to be paid to recall the previous best offer; further, the previous best offer can be lost at the next time instant with some probability. In [David and Levi 2004], David and Levi consider a model in which the offers arrive at the points of a renewal process. Variants with unknown offer (or reward) distribution, or one where a parameter of the offer distribution is unknown have been studied in [Albright 1977; Rosenfield et al. 1983]. However, in the above models, unlike in our case, the reward value is immediately revealed upon an offer arrival. Further, they do not incorporate an additional probe action like in our model.

One model that is close to ours is that of Stadje [Stadje 1997], 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. The optimal policy is characterized by stage independent thresholds, which is again due to, as in [Thejaswi et al. 2010], the problem horizon length being infinite. Recall that ours is a finite horizon problem.

In the present work we generalize the basic model in a different direction, by introducing an additional probe action and an associated (positive) probing cost (i.e., $\delta > 0$

case) into the model, so that a relay's reward value (equivalently, buyer's offer value) is now not revealed to the forwarder (equivalently, seller) for free. Instead the forwarder can choose to probe a relay to know its reward value after incurring an additional cost of δ. To the best of our knowledge, the particular model we study here is not available in the asset selling problem literature.

4. RESTRICTED CLASS $\overline{\Pi}$: AN MDP FORMULATION

Confining to the restricted class $\overline{\Pi}$, in this section we will formulate the problem in (8) as a Markov decision process. This will require us to first discuss the one-step cost functions and state transitions before proceeding to write the Bellman optimality equations.

4.1. One-Step Costs and State Transitions

The decision instants or the decision stages are the times at which the relays wake-up. Thus, there are N decision stages indexed by $k = 1, 2, \cdots, N$. Recall that for any policy in the restricted class $\overline{\Pi}$, the decision at stage k is based on (b_k, H_k), where b_k is the best reward so far and $H_k \in \mathcal{F}_k$ is the best reward distribution with $\mathcal{F}_k$ being the set of reward distributions of all the unprobed relays so far. As mentioned earlier, if no relay has been probed until stage k then $b_k = -\infty$. On the other hand, if all the relays have been probed, in which case $\mathcal{F}_k$ is empty, then we will denote the state as simple b_k. Hence, the state space can be written as,

$$\mathcal{X} \;=\; [0,\overline{r}] \cup \Big\{(b, F_\ell) : b \in \{-\infty\} \cup [0,\overline{r}], \ell \in \mathcal{L}\Big\} \cup \{\boldsymbol{t}\}$$

where $\boldsymbol{t}$ is the cost-free termination state. We will use (b, F_ℓ) to denote a generic state at stage k.

Now, at stage $k = 1, 2, \cdots, N-1$, given that the state is (b, F_ℓ), if $\mathscr{F}$'s decision is to stop then the decision process enters $\boldsymbol{t}$, with $\mathscr{F}$ incurring a termination cost of $-\eta b$ (recall from (8) that $\eta > 0$ is the trade-off parameter). On the other hand, if the action is to continue then $\mathscr{F}$ will first incur a waiting cost of U_{k+1} (the time until the next relay wakes up) and then, when the $(k+1)$-th relay wakes-up (whose reward distribution is $F_{L_{k+1}}$), $\mathscr{F}$ chooses between the two unprobed relays $-$ one the previous relay with reward distribution F_ℓ, and other the new one with distribution $F_{L_{k+1}}$ $-$ so that the state at stage $k+1$ will be either (b, F_ℓ) or $(b, F_{L_{k+1}})$. The best reward value continues to be b since no new relay has been probed during the state transition.

Alternatively, $\mathscr{F}$ could choose the action to probe the available unprobed relay (whose reward distribution is F_ℓ) incurring a cost of $\eta\delta$ (recall that δ is the probing cost). After probing, the decision process is still considered to be at stage k with the new state being $b' = \max\{b, R_\ell\}$, where R_ℓ is the reward value of the just probed relay (thus the distribution of R_ℓ is F_ℓ). $\mathscr{F}$ has to now further decide whether to stop (incurring a one-step cost of $-\eta b'$ and enter $\boldsymbol{t}$), or continue (in which case the one-step cost is U_{k+1} and the next state is $(b', F_{L_{k+1}})$).

Summarizing the above we can write the one-step cost, when the state at stage k is (b, F_ℓ), as

$$g_k\Big((b, F_\ell), a_k\Big) \;=\; \begin{cases} -\eta b & \text{if } a_k = \mathsf{s} \\ U_{k+1} & \text{if } a_k = \mathsf{c} \\ \eta\delta & \text{if } a_k = \mathsf{p}. \end{cases}$$

The next state, X', is given by

$$X' \;=\; \begin{cases} \boldsymbol{t} & \text{if } a_k = \mathsf{s} \\ (b, F_\ell) \text{ or } (b, F_{L_{k+1}}) & \text{if } a_k = \mathsf{c} \\ \max\{b, R_\ell\} & \text{if } a_k = \mathsf{p}. \end{cases}$$

We have used X' to denote the next state instead of X_{k+1} because, if $a_k = \mathsf{p}$ then the system is still at stage k. Only when the action is s or c the system transits to the stage $k+1$.

Next, if the state at stage k is b (states of this form occur after probing the available unprobed relay; recall the above expressions when $a_k = \mathsf{p}$), then

$$g_k(b, a_k) \;=\; \begin{cases} -\eta b & \text{if } a_k = \mathsf{s} \\ U_{k+1} & \text{if } a_k = \mathsf{c}, \end{cases}$$

and the next state is

$$X_{k+1} \;=\; \begin{cases} \boldsymbol{t} & \text{if } a_k = \mathsf{s} \\ (b, F_{L_{k+1}}) & \text{if } a_k = \mathsf{c}. \end{cases}$$

The action to probe is not available whenever the state is b.

At the last stage N, action c is not available, so that

$$g_N(b, F_\ell) \;=\; \begin{cases} -\eta b & \text{if } a_k = \mathsf{s} \\ \eta\delta & \text{if } a_k = \mathsf{p}, \end{cases}$$

with the system entering $\boldsymbol{t}$ if $a_k = \mathsf{s}$, otherwise (i.e., if $a_k = \mathsf{p}$) the state transits to $\max\{b, R_k\}$. Finally, $g_N(b) = -\eta b$. Note that for a policy π, the expected sum of all the one-step costs starting from stage 1, plus the average waiting time for the first relay, $\mathbb{E}[U_1] = \tau$,[3] will equal the total cost in (8).

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

Let J_k, $k = 1, 2, \cdots, N$, 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, $J_N(b) = -\eta b$, using which we obtain,

$$\begin{aligned} J_N(b, F_\ell) \;&=\; \min\Big\{ -\eta b, \eta\delta + \mathbb{E}_\ell\Big[J_N(\max\{b, R_\ell\})\Big]\Big\} \\ &=\; \min\Big\{ -\eta b, \eta\delta - \eta\mathbb{E}_\ell\Big[\max\{b, R_\ell\}\Big]\Big\}, \end{aligned} \tag{9}$$

where $\mathbb{E}_\ell[\cdot]$ denotes the expectation with respect to (w.r.t.) R_ℓ whose distribution is F_ℓ. The first term in the $\min$-expression above is the cost of stopping and the second term is the expected cost of probing and then stopping (recall that action c is not available at the last stage N). Next, for stages $k = 1, 2, \cdots, N-1$, denoting the expectation w.r.t. the distribution, L, of the location, L_{k+1}, of the next relay by $\mathbb{E}_L[\cdot]$, we have

$$J_k(b) \;=\; \min\Big\{ -\eta b, \tau + \mathbb{E}_L\Big[J_{k+1}(b, F_{L_{k+1}})\Big]\Big\}, \tag{10}$$

and

$$\begin{aligned} J_k(b, F_\ell) \;=\; \min\Big\{ &-\eta b, \eta\delta + \mathbb{E}_\ell\Big[J_k(\max\{b, R_\ell\})\Big], \\ &\tau + \mathbb{E}_L\Big[\min\{J_{k+1}(b, F_\ell), J_{k+1}(b, F_{L_{k+1}})\}\Big]\Big\}. \end{aligned} \tag{11}$$

The first term in both the min-expressions above is the cost of stopping. The middle term in (11) is the expected cost of probing, with $\eta\delta$ being the one-step cost and the remaining term being the future cost. The last term in both expressions is the expected cost of continuing, with τ representing the mean waiting time until the next relay

[3] Since invariably a relay has to be chosen, every policy has to wait for at least the first relay to wake-up, at which instant the decision process begins. Thus, U_1 need not be accounted for in the total cost incurred by any policy.

wakes up. The future cost-to-go in the last term of (11) can be understood as follows. When the state at stage $k = 1, 2, \cdots, N-1$ is (b, F_ℓ) and, if $\mathscr{F}$ 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 $\mathscr{F}$ is asked to retain one of them, then it is 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 $J_{k+1}(b, F_\ell) \leq J_{k+1}(b, F_{L_{k+1}})$, otherwise retain $F_{L_{k+1}}$.[4] Later in this section we will show that, given two distributions, F_ℓ and F_u, if F_ℓ is *stochastically greater than* F_u (recall Definition 2.1) then $J_{k+1}(b, F_\ell) \leq J_{k+1}(b, F_u)$ so that it is optimal to retain the stochastically greater distribution (Lemma 4.2-(i)).

First, for simplicity let us introduce the following notation. For $k = 1, 2, \cdots, N-1$, let C_k represent the cost of continuing:

$$C_k(b) \; = \; \tau + \mathbb{E}_L\Big[J_{k+1}(b, F_{L_{k+1}})\Big] \tag{12}$$

$$C_k(b, F_\ell) = \tau + \mathbb{E}_L\Big[\min\{J_{k+1}(b, F_\ell), J_{k+1}(b, F_{L_{k+1}})\}\Big]. \tag{13}$$

For $k = 1, 2, \cdots, N$, the cost of probing, P_k, is given by

$$P_k(b, F_\ell) \; = \; \eta\delta + \mathbb{E}_\ell\Big[J_k(\max\{b, R_\ell\})\Big]. \tag{14}$$

From (12) and (13) it is immediately clear that $C_k(b, F_\ell) \leq C_k(b)$ for any F_ℓ ($\ell \in \mathcal{L}$). This inequality should be intuitive as well, since $\mathscr{F}$ can expect to accrue a better cost if, in addition to a probed relay, it also possesses an unprobed relay. It will be useful to note this inequality as a lemma.

LEMMA 4.1. *For* $k = 1, 2, \cdots, N-1$ *and any* (b, F_ℓ) *we have* $C_k(b, F_\ell) \leq C_k(b)$.

PROOF. As discussed just before the Lemma statement, the inequality follows easily from the expressions of these costs; recall (12) and (13). □

Finally, using the above cost notation, the cost-to-go functions in (10) and (11) can be written as, for $k = 1, 2, \cdots, N-1$,

$$J_k(b) \; = \; \min\Big\{-\eta b, C_k(b)\Big\} \tag{15}$$

$$J_k(b, F_\ell) \; = \; \min\Big\{-\eta b, P_k(b, F_\ell), C_k(b, F_\ell)\Big\}. \tag{16}$$

4.3. Ordering Results for the Cost-to-go Functions

We will examine how the cost-to-go functions $J_k(b)$ and $J_k(b, F_\ell)$ behave as functions of F_ℓ and the stage index k. Consider two relays at locations ℓ and u. If the corresponding reward distributions, F_ℓ and F_u, are such that $F_\ell \geq_{st} F_u$ (recall Definition 2.1) then $\mathscr{F}$ can expect that probing the relay at ℓ would yield a better reward value than the relay at u. Thus, $\mathscr{F}$ would prefer the stochastically greater reward distribution F_ℓ, over F_u. Extending this observation, it is reasonable to expect that $\mathscr{F}$ can accrue lower expected costs (total, continuing and probing costs) if the unprobed reward distribution available at stage k is F_ℓ than if it is F_u. We will formally prove this result next. Also, we will show that, if the state remains the same, the expected cost at stage k is less than that at stage $k+1$, i.e., $J_k(x) \leq J_{k+1}(x)$ for any state x. This again should be intuitive because, starting from stage k, $\mathscr{F}$ has the option to observe an additional

[4] 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 (11).

relay than if it were to start from stage $k+1$; with more resources available, and with these being i.i.d., $\mathscr{F}$ is expected to achieve a lower cost. We will state these two results in the following lemma.

LEMMA 4.2.

(i) For $k = 1, 2, \cdots, N-1$, if $F_\ell \geq_{st} F_u$ then $C_k(b, F_\ell) \leq C_k(b, F_u)$, (and including $k = N$) $P_k(b, F_\ell) \leq P_k(b, F_u)$ and $J_k(b, F_\ell) \leq J_k(b, F_u)$.

(ii) For $k = 1, 2, \cdots, N-2$, $C_k(b) \leq C_{k+1}(b)$ and $C_k(b, F_\ell) \leq C_{k+1}(b, F_\ell)$, (and including $k = N-1$) $P_k(b, F_\ell) \leq P_{k+1}(b, F_\ell)$ and $J_k(b, F_\ell) \leq J_{k+1}(b, F_\ell)$.

PROOF. To prove (i) we first show that the various costs are non-increasing functions of b. We then complete the proof using the definition of stochastic ordering (Definition 2.1). Part (ii) follows from induction. Details of the proof is available in Appendix A.1. □

5. RESTRICTED CLASS $\overline{\Pi}$: STRUCTURAL RESULTS

We begin by defining, at stage $k = 1, 2, \cdots, N-1$, the *stopping set* $\mathcal{S}_k$ as

$$\mathcal{S}_k = \Big\{ b : -\eta b \leq C_k(b) \Big\}. \tag{17}$$

From (15) it follows that the stopping set $\mathcal{S}_k$ is the set of all states b (states of this form are obtained after probing at stage k) where it is better to stop than to continue.

Similarly, for a given distribution F_ℓ we define the stopping set $\mathcal{S}_k^\ell$ as, for $k = 1, 2, \cdots, N-1$,

$$\mathcal{S}_k^\ell = \Big\{ b : -\eta b \leq \min\{P_k(b, F_\ell), C_k(b, F_\ell)\} \Big\}. \tag{18}$$

Using (16) the set $\mathcal{S}_k^\ell$ has to be interpreted as, for a given distribution F_ℓ, the set of b such that whenever the state at stage k is (b, F_ℓ) it is better to stop than to either probe or continue. Note that when $b = -\infty$ it is never optimal to stop; hence, both these stopping sets are subsets of $[0, \overline{r}]$. Finally, stopping sets can also be defined for $k = N$ as, $\mathcal{S}_N = [0, \overline{r}]$ (since, at the last stage N, for any b the only action available is to stop), and

$$\mathcal{S}_N^\ell = \Big\{ b : -\eta b \leq P_k(b, F_\ell) \Big\}. \tag{19}$$

The following set inclusion properties easily follow from the definition of these sets and the properties of the cost functions in Lemma 4.1 and Lemma 4.2.

LEMMA 5.1.

(i) For $k = 1, 2, \cdots, N$ and for any F_ℓ we have $\mathcal{S}_k^\ell \subseteq \mathcal{S}_k$.

(ii) For $k = 1, 2, \cdots, N$, if $F_\ell \geq_{st} F_u$ then $\mathcal{S}_k^\ell \subseteq \mathcal{S}_k^u$.

(iii) For $k = 1, 2, \cdots, N-1$ we have $\mathcal{S}_k \subseteq \mathcal{S}_{k+1}$, and for any F_ℓ, $\mathcal{S}_k^\ell \subseteq \mathcal{S}_{k+1}^\ell$.

PROOF. Recall the definition of the stopping sets from (17) and (18). Part (i) follows from Lemma 4.1. Parts (ii) and (iii) are due to Parts (i) and (ii) of Lemma 4.2, respectively. □

Discussion: The above results can be understood as follows. Whenever an unprobed relay (say with reward distribution F_ℓ) is available, $\mathscr{F}$ can be more stringent about the best reward values, b, for which it chooses to stop. This is because, $\mathscr{F}$ can now additionally choose to probe F_ℓ possibly yielding a better reward than b. Thus, unless the best reward b is already good (so that there is no gain in probing F_ℓ), $\mathscr{F}$ will not choose to stop. Hence, we have $\mathcal{S}_k^\ell \subseteq \mathcal{S}_k$ (Part (i)). Next, if $F_\ell \geq_{st} F_u$ then since probing

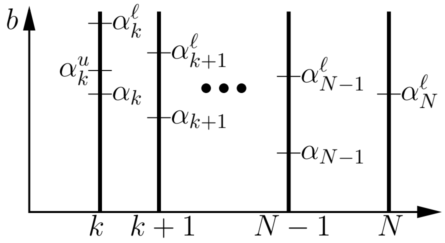


Fig. 3. Illustration of the threshold property: the vertical lines are the reward axis, with each line corresponding to a different stage. The stopping sets are represented by marking their thresholds on the respective vertical lines.

F_ℓ has a higher chance of yielding a better reward, the stopping condition is more stringent if the reward distribution of the available unprobed relay is F_ℓ than F_u. Hence, the corresponding stopping sets are ordered as in Part (ii) of the above lemma, i.e., $\mathcal{S}_k^\ell \subseteq \mathcal{S}_k^u$. Finally, whenever there are more stages to-go, $\mathscr{F}$ can be more cautious about stopping since it has the option to observe more relays. This suggests that $\mathcal{S}_k \subseteq \mathcal{S}_{k+1}$ and $\mathcal{S}_k^\ell \subseteq \mathcal{S}_{k+1}^\ell$ (Part (iii)).

From our above discussion, the phrase "$\mathscr{F}$ being more stringent about stopping," suggests that it may be better to stop for larger values of b. Equivalently, this would mean that the stopping sets are characterized by *thresholds*, beyond which it is optimal to stop. This is exactly our first main result (Theorem 5.3). Later we will prove a more interesting result (Theorem 5.6 and 5.7) where we show that the stopping sets are *stage independent*, i.e., $\mathcal{S}_k = \mathcal{S}_{k+1}$ and $\mathcal{S}_k^\ell = \mathcal{S}_{k+1}^\ell$. In the following sub-sections we will work the details of these two results.

5.1. Stopping Sets: Threshold Property

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 increments in the cost of stopping.

LEMMA 5.2. *For $k = 1, 2, \cdots, N-1$ (for Part (ii), $k = 1, 2, \cdots, N$), for any F_ℓ, and for $b_2 > b_1$ we have*

(i) $C_k(b_1) - C_k(b_2) \leq \eta(b_2 - b_1)$,
(ii) $P_k(b_1, F_\ell) - P_k(b_2, F_\ell) \leq \eta(b_2 - b_1)$
(iii) $C_k(b_1, F_\ell) - C_k(b_2, F_\ell) \leq \eta(b_2 - b_1)$.

PROOF. Available in Appendix A.2. □

THEOREM 5.3. *For $k = 1, 2, \cdots, N$ 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. Since $\mathcal{S}_N = [0, \overline{r}]$, Part (i) trivially holds for $k = N$. Next, for $k = 1, 2, \cdots, N-1$, using Lemma 5.2-(i) we can write,

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

Since $b_1 \in \mathcal{S}_k$, from (17) we know that $-\eta b_1 \leq C_k(b_1)$, using which in the above expression we obtain $-\eta b_2 \leq C_k(b_2)$ implying that $b_2 \in \mathcal{S}_k$. Part (ii) can be similarly completed using Parts (ii) and (iii) of Lemma 5.2. □

Discussion: Thus, the stopping sets $\mathcal{S}_k$ and $\mathcal{S}_k^\ell$ can be characterized in terms of lower bounds α_k and α_k^ℓ, respectively, as illustrated in Fig. 3 (see the vertical line corresponding to the stage index k). Also shown in Fig. 3 is the threshold, α_k^u, corresponding to a distribution $F_u \leq_{st} F_\ell$. From Lemma 5.1-(i) and 5.1-(ii) it follows that these thresholds are ordered, $\alpha_k \leq \alpha_k^u \leq \alpha_k^\ell$. Further, in Fig. 3 we have depicted these thresholds to be decreasing with the stage index k (vertical lines from left to right); this is due to Lemma 5.1-(iii) from where we know that the stopping sets are increasing with k. Our main result in the next section (Theorem 5.6 and 5.7) is to show that these thresholds are, in fact, equal (i.e., $\alpha_k = \alpha_{k+1}$ and $\alpha_k^\ell = \alpha_{k+1}^\ell$). Finally, note that in Fig. 3 we have not shown the threshold α_N corresponding to the stopping set $\mathcal{S}_N$; this is simply because $\alpha_N = 0$ (since $\mathcal{S}_N = [0, \overline{r}]$).

5.2. Stopping Sets: Stage Independence Property

From Lemma 5.1-(iii) we already know 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, thus leading to the result that the stopping sets are identical across the stages. We will begin by defining the sets $\mathcal{Q}_k^\ell$ as, for $k = 1, 2, \cdots, N-1$,

$$\mathcal{Q}_k^\ell \;=\; \Big\{ b : \min\{-\eta b, P_k(b, F_\ell)\} \leq C_k(b, F_\ell) \Big\}. \tag{20}$$

From (16) it follows that $\mathcal{Q}_k^\ell$ is, for a given distribution F_ℓ, the set of all b such that whenever the state at stage k is (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 (18) and (20), respectively) it immediately follows that $\mathcal{S}_k^\ell \subseteq \mathcal{Q}_k^\ell$. Also from Lemma 5.1-(i) we already 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. Using the total stochastic ordering property of $\mathcal{F} = \{F_\ell : \ell \in \mathcal{L}\}$ (Lemma 2.3), we will show that $\mathcal{S}_k \subseteq \mathcal{Q}_k^\ell$ (Lemma 5.5). This result will be essential for proving our main theorems.

Remark: We again recall that our subsequent results are not simply limited to the $\mathcal{F}$ in (5) which is the distribution set arising from the particular reward structure, R_ℓ, we had assumed in (4). All our subsequent results will hold for any other collection of bounded reward random variables $\{R_\ell\}$, as long as the corresponding $\mathcal{F}$ is totally stochastically ordered with a minimum distribution.

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

LEMMA 5.4. *Suppose $\mathcal{S}_k \subseteq \mathcal{Q}_k^u$, for some F_u, and some $k \in \{1, 2, \cdots, N-1\}$. Then for every $b \in \mathcal{S}_k$ we have $J_k(b, F_u) = J_N(b, F_u)$.*

PROOF. Fix a $b \in \mathcal{S}_k \subseteq \mathcal{Q}_k^u$. Then,

$$\begin{aligned}
J_k(b, F_u) \;&=\; \min\Big\{ -\eta b, P_k(b, F_u), C_k(b, F_u) \Big\} \\
&\overset{*}{=}\; \min\Big\{ -\eta b, P_k(b, F_u) \Big\} \\
&\overset{o}{=}\; \min\Big\{ -\eta b, \eta\delta + \mathbb{E}_u\Big[J_k(\max\{b, R_u\}) \Big] \Big\} \\
&\overset{\dagger}{=}\; \min\Big\{ -\eta b, \eta\delta - \eta\mathbb{E}_u\Big[\max\{b, R_u\} \Big] \Big\} \\
&=\; J_N(b, F_u).
\end{aligned}$$

In the above derivation, $*$ is because, b being in $\mathcal{Q}_k^u$, at (b, F_u) it is optimal to either stop or probe (recall (20)); o is simply obtained by substituting for $P_k(b, F_u)$ from (14). Further, after probing (since retaining the best relay is allowed) the new state, $\max\{b, R_u\} \geq b$, is also in $\mathcal{S}_k$ (recall Theorem 5.3) so that it is optimal to stop after probing; this observation yields †. Finally, the last equality is obtained by recalling the expression of $J_N(b, F_u)$ from (9). □

Remark: The proof of the above lemma crucially uses the fact that retaining (or *recalling*) the best relay is allowed. Thus, if recalling is not allowed, it is not possible to show the stage independence property (Theorem 5.6 and 5.7). However, for the latter case, the threshold property (Theorem 5.3) still holds so that the optimal policy is characterized by stage dependent thresholds as in Fig. 3 (for more details, refer to the problem of asset selling without recall [Bertsekas 2005, Section 4.4]).

Next we show that the hypothesis in the above lemma indeed holds for every $F_\ell \in \mathcal{F}$.

LEMMA 5.5. *For $k = 1, 2, \cdots, N-1$ and for any $F_\ell \in \mathcal{F}$ we have $\mathcal{S}_k \subseteq \mathcal{Q}_k^\ell$.*

PROOF. The proof involves two steps:

1) First we show that if there exists an F_u such that, for $k = 1, 2, \cdots, N-1$, $\mathcal{S}_k \subseteq \mathcal{Q}_k^u$ (thus satisfying the hypothesis in Lemma 5.4), then for every $F_\ell \geq_{st} F_u$ we have $\mathcal{S}_k \subseteq \mathcal{Q}_k^\ell$. Lemma 5.4 and the total stochastic ordering of $\mathcal{F}$ are required for this part.

2) Next we show that a minimum distribution F_m satisfies the hypothesis in Lemma 5.4, i.e., for every $k = 1, 2, \cdots, N-1$, $\mathcal{S}_k \subseteq \mathcal{Q}_k^m$. 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 existence of a minimum distribution F_m (recall Lemma 2.3) is essential here.

Formal proofs of both steps are available in Appendix A.4. □

The following are the main theorems of this section:

THEOREM 5.6. *For $k = 1, 2, \cdots, N-2$, $\mathcal{S}_k = \mathcal{S}_{k+1}$.*

PROOF. From Lemma 5.1-(iii) 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 5.5 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. Now, applying Lemma 5.4 we can write, $J_{k+1}(b, F_\ell) = J_{k+2}(b, F_\ell) = J_N(b, F_\ell)$. Thus,

$$\begin{aligned}
C_{k+1}(b) &= \tau + \mathbb{E}_L\Big[J_{k+2}(b, F_{L_{k+2}})\Big] \\
&= \tau + \mathbb{E}_L\Big[J_{k+1}(b, F_{L_{k+2}})\Big] \\
&\overset{*}{=} \tau + \mathbb{E}_L\Big[J_{k+1}(b, F_{L_{k+1}})\Big] \\
&= C_k(b),
\end{aligned}$$

where $*$ is obtained by replacing L_{k+2} by L_{k+1} since these are identically distributed. Finally, since $b \in \mathcal{S}_{k+1}$ we have $-\eta b \leq C_{k+1}(b) = C_k(b)$ which implies that $b \in \mathcal{S}_k$. □

THEOREM 5.7. *For $k = 1, 2, \cdots, N-1$ and any F_ℓ, $\mathcal{S}_k^\ell = \mathcal{S}_{k+1}^\ell$.*

PROOF. Similar to the proof of Theorem 5.6, here we need to show that the probing and continuing costs satisfy analogous equalities, i.e., for $b \in \mathcal{S}_k^\ell$ we need to show that $P_{k+1}(b, F_\ell) = P_k(b, F_\ell)$ and $C_{k+1}(b, F_\ell) = C_k(b, F_\ell)$. Formal proof is available in Appendix A.5. □

Discussion: It will be interesting to compare the above results with the solution to the basic relay selection model (i.e., $\delta = 0$ case; recall Section 3) or equivalently the basic asset selling problem. Towards this end, it will be useful to recall some definitions first. The basic version of the problem comprises only the stop and continue actions (where, in general, there can be more than one type of continue action). A problem is said to be *monotone* if the stopping sets $\mathcal{S}_k$ are *absorbing*, i.e., if $X_k \in \mathcal{S}_k$ and suppose the process is allowed to continue, then the next state $X_{k+1} \in \mathcal{S}_{k+1}$ so that it is optimal to stop at the next stage as well. For a monotone problem, it is known that the *1-step-look-ahead (OSLA)* rule is optimal at any stage, implying that stopping sets are

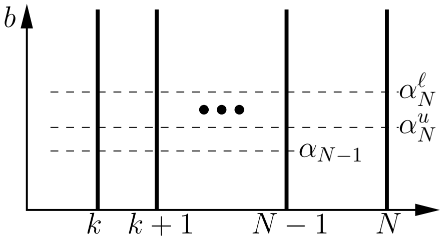


Fig. 4. Illustration of the stage independence property: only the thresholds corresponding to the last stage (and stage $N-1$ for $\mathcal{S}_k$) are shown, since these are sufficient to characterize the stopping sets for any k.

identical across the stages [Bertsekas 2005, Section 4.4]. Finally, for convenience, let us recall the definition of the OSLA rule: A policy is said to be OSLA if, at any stage, it chooses to stop if and only if (iff) the "cost of stopping" is less than the "cost of continuing for one-more step and then stopping".

In contrast to the basic setting, our formulation includes an additional probe action, due to which the above definitions will not directly apply. For instance, if $b \in \mathcal{S}_k$ and suppose we continue, then it is possible that the next state (b, F_ℓ) is such that $b \notin \mathcal{S}_{k+1}^\ell$, so that it is not optimal to stop at the next stage. Thus, our problem is not monotone from the sense of the above standard definition. Similarly, the standard OSLA rule is not optimal for our case, since the "cost of stopping $(-\eta b)$" is always less than the "cost of continuing for one-more step and then stopping $(\tau - \eta b)$", implying that it is optimal to stop at any b or (b, F_ℓ) which is not true. However, owing to Lemma 5.5, our setting satisfies the following modified definition of monotonicity: if $b \in \mathcal{S}_k$ then the next state (b, F_ℓ) is such that it is either optimal to "stop" or "probe and stop". The following modified definition of the OSLA rule is optimal for our case: A policy is said to be OSLA, if at any stage,

- for states of the form b, it chooses to stop iff the "cost of stopping" is less than the "cost of continuing for one more step and then choosing optimally between stopping or probing-and-stopping"
- for states of the form (b, F_ℓ), it chooses to stop iff the "cost of stopping" is less than the "cost of probing and stopping".

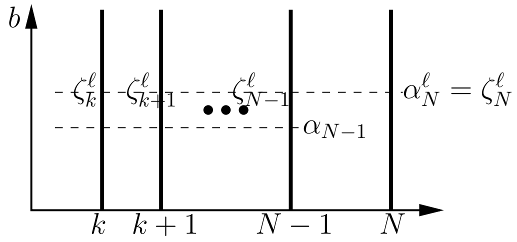
Now, for the case where the probing cost $\delta = 0$, the probe action is always exercised so that the above definitions reduce to the standard ones; the decision problem effectively simplifies to the basic setting of choosing between the stop and continue actions. Thus, our formulation can be thought of as generalizing the basic setting ($\delta = 0$ case) by incorporating an additional probe action ($\delta > 0$ case) into the existing set of stop and continue actions.

Finally, owing to Theorem 5.6 and 5.7, we can now modify the illustration in Fig. 3 to Fig. 4 where we show only a single threshold corresponding to each stopping set. Thus, to characterize the stopping set $\mathcal{S}_k^\ell$ for any k, it is sufficient to compute only the threshold α_N^ℓ corresponding to the last stage. Similarly, the stopping set $\mathcal{S}_k$ is characterized by the threshold α_{N-1} computed for stage $N-1$ (recall that $\alpha_N = 0$).

5.3. Probing Sets

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

$$\mathcal{P}_k^\ell = \Big\{ b : P_k(b, F_\ell) \leq \min\{-\eta b, C_k(b, F_\ell)\} \Big\}. \tag{21}$$



(a) (b)

Fig. 5. Structure of the probing sets if Conjecture 5.8 is true. (a) Probing sets corresponding to a distribution F_ℓ such that $\alpha_N^\ell > \alpha_{N-1}$, (b) Probing sets corresponding to an F_u such that $\alpha_N^u = \alpha_{N-1}$

Note that $\mathcal{P}_k^\ell$ is simply the difference of the sets $\mathcal{Q}_k^\ell$ and $\mathcal{S}_k^\ell$, i.e., $\mathcal{P}_k^\ell = \mathcal{Q}_k^\ell \setminus \mathcal{S}_k^\ell$.

From our numerical work we have observed that, similar to the stopping sets, the probing sets $\mathcal{P}_k^\ell$ are characterized by upper bounds ζ_k^ℓ (see Fig. 5). The intuition for this is as follows. Let (b, F_ℓ) be the state at stage $N-1$. If the value of b is very small, then it is better to probe than to continue, because probing will give an opportunity to probe an additional relay at stage N in case the process continues after probing at stage $N-1$, while continuing without probing will deprive $\mathscr{F}$ of this opportunity. This argument can be extended to any stage k to conclude that it may be better to probe for small values of b. However, as b increases, probing may not yield a better reward than the existing b; hence probing might not be worth the cost, so that it may be better to simply continue.

To formally show the threshold property of the probing set $\mathcal{P}_k^\ell$, the following is sufficient: for any $b_2 > b_1$,

$$P_k(b_1, F_\ell) - P_k(b_2, F_\ell) \ \leq \ C_k(b_1, F_\ell) - C_k(b_2, F_\ell).$$

This is because, if $b_2 \notin \mathcal{S}_k^\ell$ (so that stopping is not optimal) is such that $b_2 \in \mathcal{P}_k^\ell$ (i.e., $P_k(b_2, F_\ell) \leq C_k(b_2, F_\ell)$) then from the above inequality we obtain $P_k(b_1, F_\ell) \leq C_k(b_1, F_\ell)$, implying that it is optimal to probe at b_1 as well so that probing sets are characterized by upper bounds. However, we have not yet been able to prove or disprove such a result, but we strongly believe that it is true and make the following conjecture.

CONJECTURE 5.8. *For $k = 1, 2, \cdots, N-1$, for any F_ℓ, if $b_2 \in \mathcal{P}_k^\ell$ then for any $b_1 < b_2$ we have $b_1 \in \mathcal{P}_k^\ell$.* $\square$

Discussion: If the above conjecture is true, then some additional structural results can be deduced. For instance, suppose for some F_ℓ, $\alpha_k^\ell > \alpha_k$, or equivalently, $\alpha_N^\ell > \alpha_{N-1}$ (refer to Fig. 5(a)). Then, since $\mathcal{S}_k \subseteq \mathcal{Q}_k^\ell$ (from Lemma 5.5), for any (b, F_ℓ) such that $\alpha_{N-1} < b < \alpha_N^\ell$, it should be optimal to probe. Now, invoking Conjecture 5.8 we can conclude that it is optimal to probe for any $b < \alpha_N^\ell$, so that $\zeta_k^\ell = \alpha_N^\ell$ for all k. Thus, for such "good" distributions, F_ℓ, (i.e., F_ℓ such that $\alpha_N^\ell > \alpha_{N-1}$) the policy corresponding to it is completely characterized by a single threshold α_N^ℓ. Next, for distributions F_u such that $\alpha_k^u = \alpha_k$ (equivalently, $\alpha_N^\ell = \alpha_{N-1}$; see Fig. 5(b)), there is a window between ζ_k^u and α_N^u where, for any (b, F_ℓ) such that $\zeta_N^u \leq b < \alpha_N^u$, it is optimal to continue. Unlike α_k^u, the thresholds ζ_k^u are stage dependent. In fact, from our numerical work, we observe that ζ_k^u are increasing with k. Finally, as depicted in Fig. 5, for any distribution F_ℓ, at the last stage we invariably should have $\alpha_N^\ell = \zeta_N^\ell$ since the action to continue is not available at stage N.

ALGORITHM 1: RST-OPT (ReSTricted-OPTimal Forwarding Policy)

```
Input: Thresholds $\alpha_{N-1}, \alpha_N^\ell, \zeta_k^\ell$ for all $\ell \in \mathcal{L}, k = 1, 2, \cdots, N$; /* These thresholds have to
      be computed offline by the forwarder $\mathscr{F}$ via backward value iteration */
Initialize: $\alpha_k = \alpha_{N-1}, k = 1, 2, \cdots, N-1$; $\alpha_N = 0$;
    $b \leftarrow -\infty$; /* best reward value is initially set to $-\infty$ */
    $F_\ell = F_m$; /* best distribution is initially set to the minimum distribution $F_m$ */
    $PROBED = 0$; /* Index of best-probed relay */
    $UNPROBED = 0$; /* Index of best-unprobed relay */
    $STATE = 1$; /* 0-1 flag to indicate whether the state is $b$ or $(b, F_\ell)$ */
for $k = 1, 2, \cdots, N$ do
    Wait until the $k$-th relay wakes-up;
    Receive $L_k$; /* location of the $k$-th relay */
    Compute $F_{L_k}$; /* reward distribution of the $k$-th relay */
    if $STATE == 1$ then
        /* State is of the form $(b, F_\ell)$ */
        if $F_{L_k} \geq_{st} F_\ell$ then
            $F_\ell \leftarrow F_{L_k}$; /* update best distribution */
            $UNPROBED = k$; /* update index of best-unprobed relay */
        end
    else
        /* State is of the form $b$ */
        $F_\ell \leftarrow F_{L_k}$;
        $UNPROBED = k$;
        $STATE = 1$;
    end
    if $b \geq \alpha_N^\ell$ then
        /* Optimal to stop */
        Forward the packet to $PROBED$;
        break;
    else if $b < \zeta_k^\ell$ then
        /* Optimal to probe */
        Probe $UNPROBED$;
        Receive $R_k$; /* Reward value of $UNPROBED$ */
        if $R_k > b$ then
            $b \leftarrow R_k$; /* Update best reward value */
            $PROBED = UNPROBED$;
            $STATE = 0$;
        end
    end
    if $STATE == 0$ then
        if $b \geq \alpha_k$ then
            Forward the packet to $PROBED$;
            break;
        end
    end
end
```

5.4. RST-OPT (ReSTricted-OPTimal) Policy

Recall from Theorem 5.3 that the stopping sets $\mathcal{S}_k$ and $\mathcal{S}_k^\ell$ ($\ell \in \mathcal{L}, k = 1, 2, \cdots, N$) are characterized by lower bounds α_k and α_k^ℓ. Also, in Theorem 5.6 and 5.7 we proved that these thresholds are stage independent. Hence, it is sufficient to compute only α_{N-1} and α_N^ℓ, thus simplifying the overall computation of the optimal policy (which is optimal within the restricted class; recall the discussion following (8)). Further, if Conjecture 5.8 is true, then the upper bounds ζ_k^ℓ are sufficient to characterize the probing sets $\mathcal{P}_k^\ell$. The various thresholds can be computed by the forwarding node by solving the Bellman equation in in (10) and (11) via backward value iteration, starting with the initial conditions, $J_N(b) = -\eta b$ and $J_N(b, F_\ell)$ given by (9) (for all b and F_ℓ).

Now, $\mathscr{F}$ after computing these thresholds, operates as follows: At stage $k = 1, 2, \cdots, N-1$, whenever the state is (b, F_ℓ), **(1)** if $b \geq \alpha_N^\ell$ then stop and forward the packet to the probed relay, **(2)** if $b \leq \zeta_k^\ell$ then probe the unprobed relay and update the best reward to $b' = \max\{b, R_\ell\}$. Now, if $b' \geq \alpha_{N-1}$ stop, otherwise continue to wait for the next relay, **(3)** otherwise (i.e., if $\zeta_k^\ell < b < \alpha_N^\ell$), 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 and if the state is (b, F_ℓ) then if $b \geq \alpha_N^\ell$ stop, otherwise probe (continue is not available). Finally, if the state at stage N is b then stop irrespective of its value.

We summarize the above discussion in the form of Algorithm 1 (labeled RST-OPT). Thus, in a large sleep-wake cycling wireless network, routing of an alarm packet to the sink node can be accomplished by successively using RST-OPT at each node along the path of the packet (recall Fig. 1); the thresholds required for implementing RST-OPT can be computed and stored offline, independently by each node during the *normal operational phase* (i.e., the phase where no event of interest has occurred yet so that the nodes are sleep-wake cycling without any disruption). In Section 7 we will study the end-to-end performance achieved by RST-OPT.

6. UNRESTRICTED CLASS Π: AN INFORMAL DISCUSSION

In this section, based on the insights we have obtained from the analysis in the previous sections, we will informally discuss the possible structure of the optimal policy within the unrestricted class of policies, Π.

Recall that a policy within Π, at stage k, is in general allowed to base its decision on $(b_k, \mathcal{F}_k)$ where b_k is the reward of the best probed relay ($b_k = -\infty$ if no relay has been probed yet) and $\mathcal{F}_k$ is the set of unprobed relays ($\mathcal{F}_k = \{\}$ if all the relays have been probed). Thus, the state space at stage k can be written as

$$\mathcal{X}_k = \Big\{(b, \mathcal{H}) : b \in \{-\infty\} \cup [0, \overline{r}], \mathcal{H} \in \mathcal{F}^j, 0 \leq j \leq k\Big\}. \tag{22}$$

Again the actions available are stop, probe, and continue. If the action is to probe then $\mathscr{F}$ has to further decide which relay to probe among the several ones available at stage k. When there are no unprobed relays (i.e., $\mathcal{H} = \{\ \}$) we will represent the state as simply b. We now proceed to write the recursive Bellman optimality equation for this more general unrestricted problem. Although these equations are more involved than the ones in Section 4 (recall (9) through (11)), these can be understood similarly and hence we do not provide an explanation. The sole purpose for writing these equations here is because we will require these (in Section 7) to perform value iteration and numerically compute an optimal policy for the unrestricted problem. Hence these equations can be omitted without affecting the readability of the remainder of this section.

Let J_k, $k = 1, 2, \cdots, N$, represent the optimal cost-to-go at stage k (for simplicity we are again using J_k), then, $J_N(b) = -\eta b$, and

$$J_N(b, \mathcal{H}) \ = \ \min\Big\{ -\eta b, \eta\delta + \min_{F_\ell \in \mathcal{H}} \mathbb{E}_\ell\Big[J_N(\max\{b, R_\ell\}, \mathcal{H} \setminus \{F_\ell\})\Big]\Big\}. \tag{23}$$

For stage $k = 1, 2, \cdots, N-1$ we have

$$J_k(b) \ = \ \min\Big\{ -\eta b, \tau + \mathbb{E}_L\Big[J_k(b, \{F_{L_{k+1}}\})\Big]\Big\}, \tag{24}$$

$$J_k(b,\mathcal{H}) \;=\; \min\Big\{-\eta b, \eta\delta + \min_{F_\ell \in \mathcal{H}} \mathbb{E}_\ell\Big[J_k(\max\{b,R_\ell\}, \mathcal{H}\setminus\{F_\ell\})\Big], \\ \tau + \mathbb{E}_L\Big[J_{k+1}(b,\mathcal{H}\cup\{F_{L_{k+1}}\})\Big]\Big\}. \quad (25)$$

In view of the complexity of the problem, we do not pursue the formal analysis of characterizing the structure of the optimal policy within the unrestricted class. However, based on our results from the previous sections and a related work by Chaporkar and Proutiere [Chaporkar and Proutiere 2008], we will discuss the possible structure of the unrestricted-optimal policy.

6.1. Discussion on the Last Stage N

Suppose the decision process enters the last stage N. Now, given the best reward value among the probed relays, b, and the set $\mathcal{H}$ of reward distributions of the unprobed relays, $\mathscr{F}$ has to decide whether to stop, or probe a relay (note that continue action is not available at the last stage). Suppose the action is to probe then, after probing and updating the best reward value, if still there are some unprobed relays left, $\mathscr{F}$ has to again decide to stop or probe. This decision problem is similar to the one studied by Chaporkar and Proutiere in [Chaporkar and Proutiere 2008], but from the context of channel selection. In the following, we will briefly describe the problem in [Chaporkar and Proutiere 2008].

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 reducing the effective transmission time within the channel 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.

The authors in [Chaporkar and Proutiere 2008], for their channel probing problem, prove that the 1-step-look-ahead (OSLA) rule is optimal: 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 expected throughput obtained by probing any unprobed channel and then stopping (by transmitting on the new-best channel). Further, they prove that if the set of channel gain distributions is totally stochastically ordered (recall Definition 2.2), then it is optimal to probe the channel whose distribution is stochastically largest among all the unprobed channels. However, in their problem maximizing throughput involves optimizing a product of the channel gain and the remaining transmission time, unlike in our problem where (at the last stage) we optimize a linear combination of reward and the probing cost. But, from our numerical work we have seen that a similar OSLA rule is optimal once our decision process enters the last stage N: given a state $(b,\mathcal{H})$ at stage N, it is optimal to stop if the cost of stopping is less than the cost of probing any distribution from $\mathcal{H}$ and then stopping; otherwise it is optimal to probe the stochastically largest distribution from $\mathcal{H}$.

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

For the other stages $k = 1, 2, \cdots, N-1$, one can begin by defining the stopping sets $\mathcal{S}_k$ and $\mathcal{S}_k^{\mathcal{H}}$, and the sets $\mathcal{Q}_k^{\mathcal{H}}$, analogous to the ones in (17), (18) and (20). Note that, here we need to define $\mathcal{S}_k^{\mathcal{H}}$ and $\mathcal{Q}_k^{\mathcal{H}}$ for a set of distributions $\mathcal{H}$ unlike in the earlier case where we had defined these sets only for a given distribution F_ℓ. We expect that the results analogous to the ones in Section 5, namely Theorems 5.6 and 5.7 where we prove that the stopping sets are stage independent, hold true for this more general setting as well. Further, similar to that at stage N, for any stage k we expect that if it is optimal to probe at some state $(b, \mathcal{H})$ then it is better to probe the stochastically largest distribution from $\mathcal{H}$. Again, we have seen that these observations hold in our numerical work.

7. NUMERICAL AND SIMULATION RESULTS

7.1. One-Hop Study

We begin by listing the various parameter values that we have used in our numerical work. The forwarder and the sink are separated by a distance of $V = 1000$ meters (m); recall Fig. 2. The radius of the communication region is 50 m. We set $z_{min} = 5$ m. 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 have discretized the forwarding region $\mathcal{L}$ into a grid of 20 uniformly spaced points within $\mathcal{L}$ and then map the location of each relay to a grid point closest to it. Since the grid is symmetric about the line joining $\mathscr{F}$ and the sink (with 4 points lying on the line so that these do not have symmetric pairs), we have in total $(\frac{20-4}{2} + 4 =)$ 12 different possible D_ℓ values, giving rise to 12 different reward distributions constituting the set $\mathcal{F}$.

Next, recall the reward expression from (4); we have fixed, $d_{ref} = 5$ m, $\xi = 2.5$, and $a = 0.5$. For ΓN_0, which is referred to as the *receiver sensitivity*, we use a value of 10^{-9} mW (equivalently -90 dBm) specified for the Crossbow TelosB wireless mote [Crossbow 2006]. To ensure that the transmit power of a relay from any grid location is within the range of 1 mW to 0.003 mW (equivalently 0 dBm to -24 dBm; again from TelosB datasheet [Crossbow 2006]),[5] we allow for four different channel gain values: $0.4 \times 10^{-3}, 0.6 \times 10^{-3}, 0.8 \times 10^{-3}$, and 1×10^{-3}, each occurring with equal probability. Since channel probing is usually performed using the maximum allowable transmit power, we set the probing cost δ to be 1 mW. Finally, the inter-wake-up times $\{U_k\}$ are exponentially distributed random variables with mean $\tau = 20$ milliseconds (ms).

One-Hop Policies: The following is the description of the policies that we will study:

- RST-OPT (ReSTricted OPTimal): The optimal policy within the restricted class (Sections 4 and 5) where $\mathscr{F}$ is allowed to keep at most two relays awake – the best probed and the best unprobed relay; recall the summary in Section 5.4.
- GLB-OPT (GLoBal OPTimal): The optimal policy within the unrestricted class of policies where $\mathscr{F}$ operates by keeping all the unprobed relays awake. We obtain GLB-OPT by numerically solving the optimality equations in (23), (25) and (24).
- BAS-OPT (BASic OPTtimal): The optimal policy for the basic relay selection model where $\mathscr{F}$ is not allowed to exercise the option of *not-probing* a relay (recall discussion of the basic model from related work). Thus, each time a relay wakes up, it is immediately probed (incurring a cost of $\eta\delta$) and its reward value is revealed to $\mathscr{F}$. By incorporating $\eta\delta$ into the term τ (so that the inter-wake-up time is modified to $\tau + \eta\delta$), the solution to this model can be characterized (see our prior work [Naveen and Kumar 2013, Section 6]) in terms of a single threshold α as follows: at any stage $k = 1, 2, \cdots, N-1$, stop if and only if the best reward value $b_k \geq \alpha$; at stage N stop for any b_N. Note that the threshold α depends on η.

Discussion: In Fig. 6 we have plotted the total cost (i.e., the objective in (8)) incurred by each of the above policies as a function of the coefficient η. GLB-OPT being the

[5] Although practically only a finite set of transmit power levels will be allowed, for our numerical work we assume that the relays can transmit using any power within the specified range.

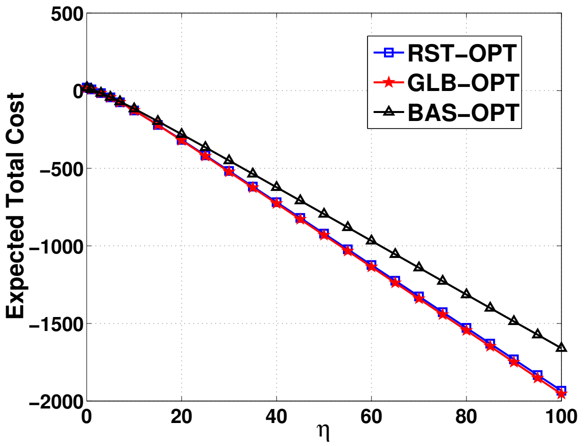


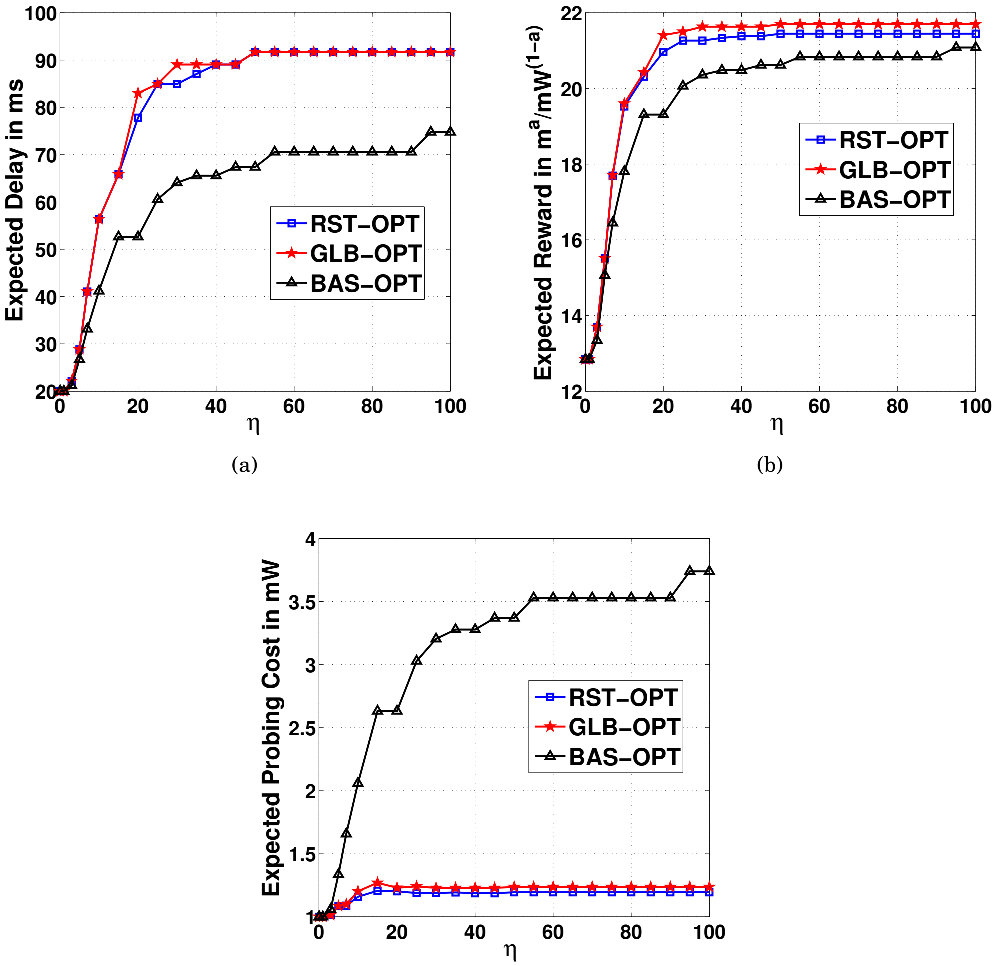
Fig. 6. Expected total cost as a function of the trade-off coefficient η; see (8). Recall that a large η implies less emphasis on expected delay.

globally optimal policy achieves the minimum cost. However, interestingly we observe that the total cost obtained by RST-OPT is very close to that of GLB-OPT. While the performance of BAS-OPT is good for small values of η, the performance degrades as η increases illustrating that it is not wise to naively probe every relay as and when they wake-up.

In Fig. 7 we have shown the individual components of the total cost (namely delay, reward, and probing cost) as functions of η. As η decreases to 0 we see (from Fig. 7(a)) that the expected delay incurred by all the policies converges to 20 ms which is the mean time, τ, until the first relay wakes up. Similarly, the expected rewards (in Fig. 7(b)) converge to reward of the first relay, and the probing costs (in Fig. 7(c)) converge to the cost of probing a single relay, i.e., $\delta = 1$ mW. This is because, for small values of η, since delay is valued more (recall the total cost expression from (8)), all the policies essentially end up probing the first relay and then forwarding the packet to it. This also explains as to why similar total cost (recall Fig. 6) is incurred by all the policies in the low η regime (e.g., $\eta \le 20$).

Next, as η increases we see that the delay incurred and the reward achieved by all the policies increases (see Fig. 7(a) and 7(b), respectively). While the probing cost of BAS-OPT naively increases (see Fig. 7(c)), probing costs incurred by RST-OPT and GLB-OPT saturate beyond $\eta = 20$. This is because, whenever η is large, RST-OPT and GLB-OPT are aware that the gain in reward value obtained by probing more relays is negated by the cost term, $\eta\delta$, which is added to the total cost each time a new relay is probed; BAS-OPT, not allowed to not-probe, ends up probing all the relays until the best reward exceeds the threshold α. Thus, although BAS-OPT incurs a smaller delay than the other two policies, but suffers both in terms of reward and probing cost, leading to an higher total cost. On the other hand, RST-OPT and GLB-OPT wait for more relays and then probe only the relays with good reward distribution to accrue a better total cost.

Finally, the marginal improvement in performance obtained by GLB-OPT over RST-OPT can be understood as follows. Although the delay incurred by these two policies is almost identical, for large η values, GLB-OPT achieves a better reward than RST-OPT by incurring a slightly higher probing cost. Thus, whenever the reward offered by the relay with the best distribution is not good enough, GLB-OPT probes an additional



(a)

(b)

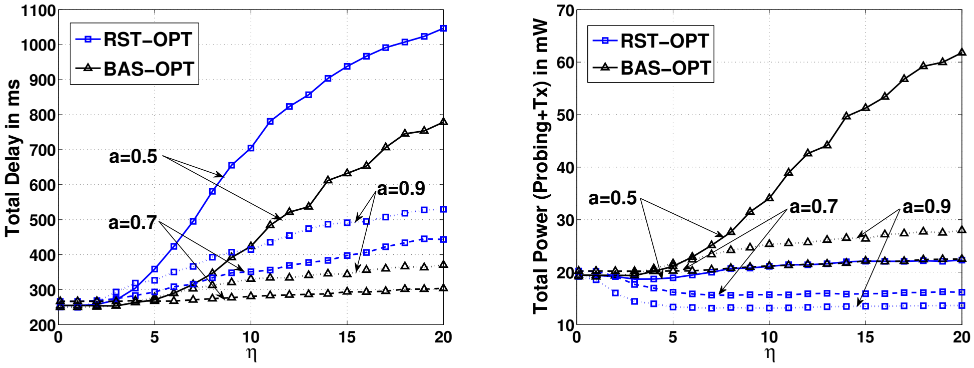
(c)

Fig. 7. Individual components of the total cost in Fig. 6 as functions of η: (a) Delay (b) Reward and (c) Probing Cost.

relay to improve the reward; such improvement is not possible by RST-OPT since it is restricted to keep only one unprobed relay awake.

Computational Complexity: Finally on the computational complexity of these policies. To obtain GLB-OPT we had to recursively solve the Bellman equation (referred to as the *value iteration*) in (23), (25) and (24), for every stage k and every possible state at stage k. The total number of all possible states at stage k, i.e., the cardinality of the state space $\mathcal{X}_k$ in (22), 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.

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 only linearly with the cardinality of $\mathcal{F}$. Also, the size of the state space does not grow with k. Furthermore, from our analysis in Section 5 we know that the stopping sets are threshold based, and



(a) (b)

Fig. 8. End-to-end performance of RST-OPT and BAS-OPT as functions of η for different values of a: (a) Total delay, and (b) Total power.

moreover the thresholds, α_k and $\{\alpha_k^\ell : F_\ell \in \mathcal{F}\}$, are stage independent. Hence, these thresholds have to be computed only once (for stage $N-1$ and N, respectively), thus further reducing the complexity of RST-OPT. BAS-OPT, being a single-threshold based policy, is much simpler to implement but is not a good choice whenever η is large.

7.2. End-to-End Study

The good one-hop performance of RST-OPT and its computational simplicity motivates us to apply RST-OPT to route packets in an asynchronously sleep-wake cycling WSN and study its end-to-end performance. We will also obtain the end-to-end performance of the naive BAS-OPT policy.

First we will describe the setting that we have considered for our end-to-end simulation study. We construct a network by randomly placing 500 nodes in a square region of side 500 m. The sink node is placed at the location $(500, 0)$. The network nodes are asynchronously and periodically sleep-wake cycling, i.e., a node i wakes up at the periodic instants, $\{T_i + kT : k \geq 0\}$, where $\{T_i\}$ are i.i.d. uniform on $[0, T]$ with T being the sleep-wake cycling period (recall our justification for the periodic sleep-wake cycling from footnote 1 in page 5). We fix $T = 100$ ms. A source node is randomly chosen, which generates an alarm packet at time 0. This alarm packet has to be routed to the sink node.

Here, in addition to varying η, we will also vary the coefficient a and study the end-to-end performance. Recall from (4) that a is the coefficient used to trade-off between progress and power in the reward expression; a larger value of a implies more emphasis on progress. The values of all the other parameters, e.g., r_c, δ, ΓN_0, channel gains, etc., remain as in our one-hop study.

Now, for a given η and a, each node computes the corresponding RST-OPT and BAS-OPT policies assuming a mean inter-wake-up time of $\frac{T}{N_i}$ ms, where N_i is the number of nodes in the forwarding region of node i. In Fig. 8, for three different values of a (namely 0.5, 0.7, and 0.9) we have plotted, as functions of η, the total delay and the total power (which is the sum of the probing and the transmission powers incurred at each hop) incurred, by applying RST-OPT and BAS-OPT policies at each hop en-route to the sink node. Each data point in Fig. 8 is obtained by averaging the respective quantities over 1000 alarm packets.

Discussion: First, note that both total delay and total power incurred by BAS-OPT are increasing with η for each a. Hence, no favorable trade-off between delay and power can be obtained using BAS-OPT; it is better to operate BAS-OPT at a low value of η, where the total delay incurred is (approximately) 250 ms while the total power expended is about 20 mW. In fact, as η decreases to 0, we see that the performance of all the policies (i.e., RST-OPT and BAS-OPT for different values of a) converge to these values. This is simply because, whenever η is small, since (one-hop) delay is valued more, all the policies, at each hop, essentially forward the packet to the first relay that wakes up.

For RST-OPT, while only a marginal trade-off between delay and power can be achieved for $a = 0.5$ (see from Fig. 8(b) that the corresponding total power decreases only marginally as η increases from 1 to 4), but as we increase the value of a to 0.7 and then to 0.9, we see that the total power sharply decreases with η. For instance, for $a = 0.9$, from Fig. 8(b) we see that the total power decrease from 20 mW to 13 mW as η goes from 0 to 7. However, over this range of η, total delay increases from 250 ms to 360 ms (see the plot corresponding to RST-OPT, $a = 0.9$, from Fig. 8(a)). Thus, for these higher values of a, trade-off between delay and power can be achieved using RST-OPT.

Next, for any fixed η, from Fig. 8(b) observe that the total power incurred by RST-OPT is improving (i.e., decreasing) with a. This can be understood as follows: since a larger a gives less emphasis on power and more emphasis on progress in the reward expression (recall (4)), then, although the one-hop transmissions may be of higher power, but there are fewer hops and hence fewer transmissions, thus resulting in a lower total power. This observation would suggest that it is advantageous to use RST-OPT by setting $a = 0.9$ rather than $a = 0.5$ or 0.7. However, from Fig. 8(a) we see that the total delay is not decreasing with a. In fact, delay incurred by RST-OPT first decreases as a increases from 0.5 to 0.7, and then increases as a is further increased to 0.9. Similar is the case for the plots corresponding to BAS-OPT in Fig. 8(a). This observation can be understood as follows. When $a = 0.5$, since (one-hop) power is valued more, the respective forwarding nodes at each hop will end up spending more time waiting for a relay which require strictly lesser transmission power. Similarly, when $a = 0.9$, larger delay is incurred at each hop since the forwarding nodes now have to wait for relays whose progress value is more (however, since $a = 0.9$ results in a fewer hops we see that the delay incurred in this case is considerably less than the $a = 0.5$ case). On the other hand, when $a = 0.7$, since a relatively fair trade-off between progress and power exists, the waiting time at each hop is reduced because now any relay with a moderate progress and a moderate transmission power would suffice.

The above argument is precisely the reason as to why the total power incurred by BAS-OPT behaves as in Fig. 8(b): when $a = 0.5$ or 0.9, each forwarder, in the process of waiting for a relay whose transmission power requirement is low or progress is large, respectively, will end up probing more relays. RST-OPT benefits over BAS-OPT here by probing only good relays at each hop, thus yielding a lower total power.

Finally, summarizing our end-to-end results, we see that no trade-off between delay and power can be achieved by the naive BAS-OPT policy, while RST-OPT achieves such a trade-off (by varying η) for $a = 0.7$ or 0.9. Further, for a fixed η, favorable trade-off between delay and power can be obtained by varying a. For instance, from Fig. 8 we see that when $\eta = 7$, moving from $a = 0.7$ to 0.9 will result in a power saving of about 3 mW while increasing the end-to-end delay by 130 ms. Thus, depending on the application requirement (i.e., delay or power sensitive application) one has to appropriately choose the values of η and a.

8. CONCLUSION

Motivated by the problem of end-to-end geographical forwarding in a sleep-wake cycling wireless sensor network, we formulated a decision problem of choosing a next-hop relay node when a set of potential relay neighbors are sequentially waking up in time. A power cost is incurred for probing a relay to learn its channel gain. We first studied a restricted class of policies where a policy's decision is based only on, in addition to the best probed relay, the best unprobed relays (instead of all the unprobed relays). We characterized the optimal policy in terms of stopping sets. Our first main result (Theorem 5.3) was to show that the stopping sets are threshold based. Then we proved that the stopping sets are stage independent (Theorem 5.6 and 5.7). A discussion on the more general unrestricted class of policies was provided. We conducted numerical work to compare the performances of the restricted optimal (RST-OPT) and the global optimal (GLB-OPT) policies. We observed that the performance of RST-OPT is close to that of GLB-OPT. We also conducted simulation experiments to study the end-to-end performance of RST-OPT. Finally, it is worth noting that our work being a variant of the asset selling problem, can, in general, find application wherever the problem of resource-selection occurs, when a collection of resources are sequentially arriving.

APPENDIX

For convenience, we will recall the respective Lemma/Theorem statement before providing its proof.

A.1. Proof of Lemma 4.2

Before proceeding to the proof of Lemma 4.2, we will require the following result first.

LEMMA A.1. *For $k = 1, 2, \cdots, N$, $J_k(b)$ and $J_k(b, F_\ell)$ are decreasing in b.*

PROOF. Proof is by induction. For stage N we know that $J_N(b) = -\eta b$, and hence is decreasing in b. Also, recalling $J_N(b, F_\ell)$ from (9):

$$J_N(b, F_\ell) = \min\Big\{ -\eta b, \eta\delta - \eta\mathbb{E}_\ell\Big[\max\{b, R_\ell\}\Big]\Big\},$$

it is easy to see that $J_N(b, F_\ell)$ is also decreasing in b. Thus, the monotonicity properties holds for stage N. Now, suppose $J_{k+1}(b)$ and $J_{k+1}(b, F_\ell)$ (for all F_ℓ) are decreasing in b for some $k+1 = 2, 3, \cdots, N$, then we will show that the result holds for stage k as well.

First, recall the expressions of $J_k(b)$ and $J_k(b, F_\ell)$ (from (15) and (16) respectively): $J_k(b) = \min\Big\{ -\eta b, C_k(b)\Big\}$ and $J_k(b, F_\ell) = \min\Big\{ -\eta b, P_k(b, F_\ell), C_k(b, F_\ell)\Big\}$. Thus to complete the proof it is sufficient to show that $C_k(b)$, $P_k(b, F_\ell)$ and $C_k(b, F_\ell)$ are decreasing in b. From the induction hypothesis, it is easy to see that $C_k(b)$ (in (12)) is decreasing in b, so that we obtain $J_k(b)$ is decreasing in b. Now that we have established $J_k(b)$ is decreasing in b, it will immediately follow that the probing cost $P_k(b, F_\ell)$ (in (14)) is decreasing in b. Finally, again using the induction argument, observe that $\min\Big\{J_{k+1}(b, F_\ell), J_{k+1}(b, F_{L_{k+1}})\Big\}$ is decreasing in b so that the continuing cost $C_k(b, F_\ell)$ (in (13)) is also decreasing. □

We are now ready to prove Lemma 4.2.
Lemma 4.2:

(i) For $k = 1, 2, \cdots, N-1$, if $F_\ell \geq_{st} F_u$ then $C_k(b, F_\ell) \leq C_k(b, F_u)$, (including $k = N$) $P_k(b, F_\ell) \leq P_k(b, F_u)$ and $J_k(b, F_\ell) \leq J_k(b, F_u)$.

(ii) For $k = 1, 2, \cdots, N-2$, $C_k(b) \leq C_{k+1}(b)$ and $C_k(b, F_\ell) \leq C_{k+1}(b, F_\ell)$, (including $k = N-1$) $P_k(b, F_\ell) \leq P_{k+1}(b, F_\ell)$ and $J_k(b, F_\ell) \leq J_{k+1}(b, F_\ell)$.

PROOF OF PART-(I). Consider stage N and recall the expression for the optimal cost-to-go function $J_N(b, F_\ell)$ from (9):

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

Since the function $f(r) = \max\{b, r\}$ is increasing in r, using the definition of stochastic ordering (Definition 2.1) we can write

$$\mathbb{E}_\ell\Big[\max\{b, R_\ell\}\Big] \geq \mathbb{E}_u\Big[\max\{b, R_u\}\Big],$$

so that we have $P_N(b, F_\ell) \leq P_N(b, F_u)$ and $J_N(b, F_\ell) \leq J_N(b, F_u)$. Thus, the result holds for stage N.

Now suppose the result is true for some $k+1 = 2, 3, \cdots, N$. From Lemma A.1 we know that $J_k(b)$ is decreasing in b, which would imply that, for any b, the function $f(r) = J_k(\max\{b, r\})$ is decreasing in r. Again, using the definition of stochastic ordering (in Definition 2.1) we can conclude that

$$\mathbb{E}_\ell\Big[J_k(\max\{b, R_\ell\})\Big] \leq \mathbb{E}_u\Big[J_k(\max\{b, R_u\})\Big],$$

so that $P_k(b, F_\ell) \leq P_k(b, F_u)$ (see (14)). Next, from the induction argument we know that $J_{k+1}(b, F_\ell) \leq J_{k+1}(b, F_u)$ so that

$$\min\Big\{J_{k+1}(b, F_\ell), J_{k+1}(b, F_{L_{k+1}})\Big\} \leq \min\Big\{J_{k+1}(b, F_u), J_{k+1}(b, F_{L_{k+1}})\Big\}.$$

Thus, we also have $C_k(b, F_\ell) \leq C_k(b, F_u)$ (see (13)). The proof can now be easily completed by recalling (from (16)) that $J_k(b, F_\ell) = \min\Big\{ -\eta b, P_k(b, F_\ell), C_k(b, F_\ell)\Big\}$. □

PROOF OF PART-(II). This result is very intuitive, since with more number of stages to go, one is expected to accrue a lower cost. However, we prove it here for completeness. Again the proof is by induction. For stage $N-1$ we easily have,

$$\begin{aligned} J_{N-1}(b) &= \min\Big\{ -\eta b, C_k(b)\Big\} \\ &\leq -\eta b \\ &= J_N(b). \end{aligned}$$

Next, consider a state of the form (b, F_ℓ). The cost of probing $P_{N-1}(b, F_\ell)$ can be bounded as follows:

$$\begin{aligned} P_{N-1}(b, F_\ell) &= \eta\delta + \mathbb{E}_\ell\Big[J_{N-1}(\max\{b, R_\ell\})\Big] \\ &\overset{*}{\leq} \eta\delta + \mathbb{E}_\ell\Big[J_N(\max\{b, R_\ell\})\Big] \\ &\overset{o}{=} \eta\delta - \eta\mathbb{E}_\ell\Big[\max\{b, R_\ell\}\Big] \\ &\overset{\dagger}{=} P_N(b, F_\ell), \end{aligned}$$

where, to obtain $*$ we have used, $J_{N-1}(b) \leq J_N(b)$ (which we had just proved), o is because $J_N(b) = -\eta b$ for all b, and † is simply obtained by recalling the expression for

$P_N(b, F_\ell)$. Using the above inequality in the following, we obtain

$$\begin{aligned} J_{N-1}(b, F_\ell) &= \min\Big\{-\eta b, P_{N-1}(b, F_\ell), C_{N-1}(b, F_\ell)\Big\} \\ &\leq \min\Big\{-\eta b, P_{N-1}(b, F_\ell)\Big\} \\ &\leq \min\Big\{-\eta b, \eta\delta - \eta\mathbb{E}_\ell\Big[\max\{b, R_\ell\}\Big]\Big\} \\ &= J_N(b, F_\ell). \end{aligned}$$

Thus we have shown the result for stage $N-1$.

Suppose the result is true for some stage $k+1 = 2, 3, \cdots, N-1$. i.e., $J_{k+1}(b) \leq J_{k+2}(b)$ and $J_{k+1}(b, F_\ell) \leq J_{k+2}(b, F_\ell)$ (for all F_ℓ), then, using the induction hypothesis, the cost of continuing, $C_k(b)$, can be bounded as

$$\begin{aligned} C_k(b) &= \tau + \mathbb{E}_L\Big[J_{k+1}(b, F_{k+1})\Big] \\ &\leq \tau + \mathbb{E}_L\Big[J_{k+2}(b, F_{k+2})\Big] \\ &= C_{k+1}(b). \end{aligned}$$

Thus, we have $J_k(b) \leq J_{k+1}(b)$ (see (15)). Next, consider the probing cost,

$$\begin{aligned} P_k(b, F_\ell) &= \eta\delta + \mathbb{E}_\ell\Big[J_k(\max\{b, R_\ell\})\Big] \\ &\overset{*}{\leq} \eta\delta + \mathbb{E}_\ell\Big[J_{k+1}(\max\{b, R_\ell\})\Big] \\ &= P_{k+1}(b, F_\ell) \end{aligned}$$

where, to obtain $*$ we have used $J_k(b) \leq J_{k+1}(b)$ which we have already shown. The cost of continuing can be similarly bounded:

$$\begin{aligned} C_k(b, F_\ell) &= \tau + \mathbb{E}_L\Big[\min\{J_{k+1}(b, F_\ell), J_{k+1}(b, F_{L_{k+1}})\}\Big] \\ &\overset{*}{\leq} \tau + \mathbb{E}_L\Big[\min\{J_{k+2}(b, F_\ell), J_{k+2}(b, F_{L_{k+2}})\}\Big] \\ &= C_{k+1}(b, F_\ell), \end{aligned}$$

where $*$ is due to the induction hypothesis and the fact that location random variables, L_{k+1} and L_{k+2}, are identically distributed. Finally, using the above inequalities in the expression of $J_k(b, F_\ell)$ $\Big($recall (16); $J_k(b, F_\ell) = \min\Big\{-\eta b, P_k(b, F_\ell), C_k(b, F_\ell)\Big\}\Big)$, we obtain $J_k(b, F_\ell) \leq J_{k+1}(b, F_\ell)$, thus completing the proof. □

A.2. Proof of Lemma 5.2

The following simple property about the min-operator will be useful while proving Lemma 5.2.

LEMMA A.2. *If* $x_1, x_2, \cdots, x_j$ *and* $y_1, y_2, \cdots, y_j$ *in* $\Re$*, are such that,* $x_i - y_i \leq x_1 - y_1$ *for all* $i = 1, 2, \cdots, j$*, then*

$$\min\{x_1, x_2, \cdots, x_j\} - \min\{y_1, y_2, \cdots, y_j\} \leq x_1 - y_1 \tag{26}$$

PROOF. Suppose $\min\{y_1, y_2, \cdots, y_j\} = y_i$, for some $1 \leq i \leq j$, then the LHS of (26) can be written as,

$$LHS = \min\{x_1, x_2, \cdots, x_j\} - y_i \leq x_i - y_i.$$

The proof is complete by recalling that we are given, $x_i - y_i \leq x_1 - y_1$. □

Lemma 5.2*:* For $k = 1, 2, \cdots, N-1$ (for part (ii), $k = 1, 2, \cdots, N$), for any F_ℓ, and for $b_2 > b_1$ we have

(i) $C_k(b_1) - C_k(b_2) \leq \eta(b_2 - b_1)$,
(ii) $P_k(b_1, F_\ell) - P_k(b_2, F_\ell) \leq \eta(b_2 - b_1)$
(iii) $C_k(b_1, F_\ell) - C_k(b_2, F_\ell) \leq \eta(b_2 - b_1)$.

PROOF. Since $J_N(b)$ is $-\eta b$ we already have, for stage N, $J_N(b_1) - J_N(b_2) = \eta(b_2 - b_1)$. Also, for a given distribution F_ℓ and for $b_2 > b_1$,

$$\begin{aligned} P_N(b_1, F_\ell) - P_N(b_2, F_\ell) &= \eta \mathbb{E}_\ell \Big[\max\{b_2, R_\ell\} - \max\{b_1, R_\ell\} \Big] \\ &\overset{*}{\leq} \eta(b_2 - b_1), \end{aligned}$$

where to obtain $*$, first consider all the three cases that are possible: (1) $R_\ell \leq b_1 < b_2$, (2) $b_1 < R_\ell < b_2$, and (3) $b_1 < b_2 \leq R_\ell$, and then note that in all these cases, $\Big(\max\{b_2, R_\ell\} - \max\{b_1, R_\ell\}\Big)$, is bounded above by $b_2 - b_1$. Now, since $J_N(b, F_\ell) = \min\Big\{-\eta b, P_N(b, F_\ell)\Big\}$, the above inequality along with Lemma A.2 will yield, $J_N(b_1, F_\ell) - J_N(b_2, F_\ell) \leq \eta(b_2 - b_1)$.

Suppose for some stage $k+1 = 1, 2, \cdots, N$ we have $J_{k+1}(b_1) - J_{k+1}(b_2) \leq \eta(b_2 - b_1)$ and $J_{k+1}(b_1, F_\ell) - J_{k+1}(b_2, F_\ell) \leq \eta(b_2 - b_1)$ for all $b_2 > b_1$, and for all F_ℓ. Then we will show that all the inequalities listed in the lemma will hold for stage k as well. First, a simple application of the induction hypothesis will yield,

$$\begin{aligned} C_k(b_1) - C_k(b_2) &= \mathbb{E}_L \Big[J_{k+1}(b_1, F_{L_k}) - J_{k+1}(b_2, F_{L_k}) \Big] \\ &\leq \eta(b_2 - b_1). \end{aligned}$$

Since $J_k(b) = \min\Big\{-\eta b, C_k(b)\Big\}$, the above inequality along with Lemma A.2 gives, $J_k(b_1) - J_k(b_2) \leq \eta(b_2 - b_1)$, for any $b_2 > b_1$. Using this we can write

$$\begin{aligned} P_k(b_1, F_\ell) - P_k(b_2, F_\ell) &= \mathbb{E}_\ell \Big[J_k(\max\{b_1, R_\ell\}) - J_k(\max\{b_2, R_\ell\}) \Big] \\ &\leq \mathbb{E}_\ell \Big[\eta \Big(\max\{b_2, R_\ell\} - \max\{b_1, R_\ell\} \Big) \Big] \\ &\leq \eta(b_2 - b_1), \end{aligned} \tag{27}$$

where the last inequality is again by considering all the three regions where R_ℓ can lie.

To show part (iii), define $\mathcal{L}_\ell$ as the set of all distributions that are stochastically greater than ℓ, i.e., $\mathcal{L}_\ell = \Big\{F_t \in \mathcal{F} : F_t \geq_{st} F_\ell\Big\}$. Let $\mathcal{L}_\ell^c$ denote the set of all the remaining distributions, i.e., $\mathcal{L}_\ell^c = \mathcal{F} \backslash \mathcal{L}_\ell$. From Lemma 2.3, where we have shown that $\mathcal{F}$ is totally stochastically ordered (see Definition 2.2), it follows that $\mathcal{L}_\ell^c$ contains all distributions in $\mathcal{F}$ which are stochastically smaller than F_ℓ. Recalling the expression for $C_k(b, F_\ell)$

from (13), the difference in the cost of continuing can now be bounded as follows:

$$\begin{aligned}C_k(b_1,F_\ell)-C_k(b_2,F_\ell) &= \int_{\mathcal{F}}\Big(\min\{J_{k+1}(b_1,F_\ell),J_{k+1}(b_1,F_t)\}\\ &\qquad -\min\{J_{k+1}(b_2,F_\ell),J_{k+1}(b_2,F_t)\}\Big)dL(t)\\ &\overset{*}{=} \int_{\mathcal{L}_\ell}(J_{k+1}(b_1,F_t)-J_{k+1}(b_2,F_t))dL(t)\\ &\qquad + \int_{\mathcal{L}_\ell^c}(J_{k+1}(b_1,F_\ell)-J_{k+1}(b_2,F_\ell))dL(t).\\ &\overset{o}{\leq} \eta(b_2-b_1). \qquad (28)\end{aligned}$$

In the above derivation, $*$ is obtained by using Lemma 4.2-(i), and o is simply by applying the induction argument. Since $J_k(b,F_\ell)=\min\Big\{-\eta b, P_k(b,F_\ell), C_k(b,F_\ell)\Big\}$, using (27) and (28) along with Lemma A.2, we obtain, $J_k(b_1,F_\ell)-J_k(b_2,F_\ell)\leq\eta(b_2-b_1)$, thus completing the induction argument. □

A.3. Proof of Lemma 2.3

Lemma 2.3*:* The set of reward distributions $\mathcal{F}$ in (5), is totally stochastically ordered with a minimum distribution.

PROOF. Recall the reward expression from (4),

$$R_\ell=\frac{Z_\ell^a}{P_\ell^{(1-a)}}=\frac{Z_\ell^a}{(\Gamma' D_\ell^\xi)^{(1-a)}}G_\ell^{(1-a)}.$$

The distribution, F_ℓ, of R_ℓ can be written as,

$$\begin{aligned}F_\ell(r) &= \mathbb{P}(R_\ell\leq r)\\ &= \mathbb{P}\left(\frac{Z_\ell^a}{(\Gamma' D_\ell^\xi)^{(1-a)}}G_\ell^{(1-a)}\leq r\right)\\ &= \mathbb{P}\left(G_\ell^{(1-a)}\leq\kappa_\ell r\right), \qquad (29)\end{aligned}$$

where $\kappa_\ell=\frac{(\Gamma' D_\ell^\xi)^{(1-a)}}{Z_\ell^a}$.

Let ℓ,u be any two locations in $\mathcal{L}$. Since the rewards are non-negative, we have $F_\ell(r)=F_u(r)=0$ for $r<0$. Hence, we only need to consider the case $r\geq 0$. Now, given $\ell,u\in\mathcal{L}$, either $\kappa_\ell\leq\kappa_u$ or $\kappa_\ell>\kappa_u$. Thus we have, either $\kappa_\ell r\leq\kappa_u r$ or $\kappa_\ell r\geq\kappa_u r$, for every $r\geq 0$. Finally, since G_ℓ and G_u are identically distributed, we have, either $F_\ell(r)\leq F_u(r)$ or $F_\ell(r)\geq F_u(r)$, for all r, so that F_ℓ and F_u are stochastically ordered (recall Definition 2.1).

To show that there exists a minimum distribution, first note that κ_ℓ as a function of $\ell\in\mathcal{L}$ is continuous. Then, since we had assumed that $\mathcal{L}$ is compact (closed and bounded), there exists an $m\in\mathcal{L}$ where the maximum is achieved, i.e., $\kappa_\ell\leq\kappa_m$ for all $\ell\in\mathcal{L}$. Again, since the gains G_ℓ and G_m are identically distributed, from (29) it follows that $F_\ell\geq_{st}F_m$ for all $\ell\in\mathcal{L}$, so that F_m is the minimum distribution. □

A.4. Proof of Lemma 5.5

As discussed in the outline of the proof of Lemma 5.5, the result immediately follows once we show *Step 1* and *Step 2*. First we will formally state and prove *Step 1*.

LEMMA A.3. *Suppose F_u is a distribution such that for all $k = 1, 2, \cdots, N-1$, $\mathcal{S}_k \subseteq \mathcal{Q}^u_k$. Then for any distribution $F_\ell \geq_{st} F_u$ we have $\mathcal{S}_k \subseteq \mathcal{Q}^\ell_k$.*

PROOF. We will first show that $\mathcal{S}_{N-1} \subseteq \mathcal{Q}^\ell_{N-1}$. Fix a $b \in \mathcal{S}_{N-1}$. Then $b \in \mathcal{Q}^u_{N-1}$ (because it is given that $\mathcal{S}_{N-1} \subseteq \mathcal{Q}^u_{N-1}$), so that using the definition of the set $\mathcal{Q}^u_{N-1}$ (from (20)) we can write

$$\min\Big\{-\eta b, P_{N-1}(b, F_u)\Big\} \leq C_{N-1}(b, F_u). \tag{30}$$

For any generic distribution F_s, whenever $b \in \mathcal{S}_{N-1}$, the minimum of the cost of stopping and the cost of probing can be simplified as follows:

$$\begin{aligned}\min\Big\{-\eta b, P_{N-1}(b, F_s)\Big\} &\overset{*}{=} \min\Big\{-\eta b, \eta\delta + \mathbb{E}_s\Big[J_{N-1}(\max\{b, R_s\})\Big]\Big\} \\ &\overset{o}{=} \min\Big\{-\eta b, \eta\delta - \eta\mathbb{E}_s\Big[\max\{b, R_s\}\Big]\Big\} \\ &\overset{\dagger}{=} J_N(b, F_s).\end{aligned} \tag{31}$$

In the above, $*$ is obtained by recalling the expression for the probing cost from (14). o is because, after probing we are still at stage $N-1$ with the new state $\max\{b, R_s\}$ also in $\mathcal{S}_{N-1}$ (Lemma 5.3); in $\mathcal{S}_{N-1}$ we know that it is optimal to stop, so that $J_{N-1}(\max\{b, R_s\}) = -\eta \max\{b, R_s\}$. Finally, to obtain $\dagger$, recall the expression for $J_N(b, F_s)$ from (9).

Now using (31) in (30) we see that, the hypothesis $b \in \mathcal{S}_{N-1}$ implies, $J_N(b, F_u) \leq C_{N-1}(b, F_u)$. Also, from Lemma 4.2-(i) we have, $J_N(b, F_\ell) \leq J_N(b, F_u)$ for any $F_\ell \geq_{st} F_u$. Combining these we can write

$$J_N(b, F_\ell) \leq J_N(b, F_u) \leq C_{N-1}(b, F_u). \tag{32}$$

To conclude that $b \in \mathcal{Q}^\ell_{N-1}$, we need to show

$$\min\Big\{-\eta b, P_{N-1}(b, F_\ell)\Big\} \leq C_{N-1}(b, F_\ell),$$

or, alternatively, recalling (31), it is sufficient to show,

$$J_N(b, F_\ell) \leq C_{N-1}(b, F_\ell). \tag{33}$$

Now for any generic distribution $F_s \in \mathcal{F}$ define $\mathcal{L}_s = \Big\{t \in \mathcal{L} : F_t \geq_{st} F_s\Big\}$ i.e., $\mathcal{L}_s$ is the set of all distributions in $\mathcal{F}$ that are stochastically greater than F_s. Let $\mathcal{L}^c_\ell$ denote the set of all the remaining distributions, i.e., $\mathcal{L}^c_\ell = \mathcal{F} \setminus \mathcal{L}_\ell$. Since $\mathcal{F}$ is totally stochastically ordered (Lemma 2.3), $\mathcal{L}^c_s$ contains all distributions in $\mathcal{F}$ that are stochastically smaller than F_s. Further, for $F_\ell \geq_{st} F_u$ we have $\mathcal{L}_\ell \subseteq \mathcal{L}_u$. Then, recalling the expression for $C_{N-1}(b, F_u)$ from (13) we can write

$$\begin{aligned}C_{N-1}(b, F_u) &= \tau + \mathbb{E}_L\Big[\min\{J_N(b, F_u), J_N(b, F_{L_N})\}\Big] \\ &\overset{*}{=} \tau + \int_{\mathcal{L}_u} J_N(b, F_t)\, dL(t) + \int_{\mathcal{L}^c_u} J_N(b, F_u)\, dL(t) \\ &\overset{o}{=} \tau + \int_{\mathcal{L}_\ell} J_N(b, F_t)\, dL(t) + \int_{\mathcal{L}_u \setminus \mathcal{L}_\ell} J_N(b, F_t)\, dL(t) + \int_{\mathcal{L}^c_u} J_N(b, F_u)\, dL(t),\end{aligned}$$

where, $*$ is obtained by using Lemma 4.2-(i) and the definition of $\mathcal{L}_u$, and to obtain o we have split the integral over $\mathcal{L}_u$ (first integral in $*$) into two integrals – one over $\mathcal{L}_\ell$ and the other over $\mathcal{L}_u \setminus \mathcal{L}_\ell$. Now, for any $F_t \in \mathcal{L}_u \setminus \mathcal{L}_\ell$ we know that $F_t \geq_{st} F_u$ so that $J_N(b, F_t) \leq J_N(b, F_u)$ (again from Lemma 4.2-(i)). Thus, in the above expression, replacing $J_N(b, F_t)$ by $J_N(b, F_u)$ in the middle integral, and then combining it with the last integral, we obtain

$$C_{N-1}(b, F_u) \ \le \ \tau + \int_{\mathcal{L}_\ell} J_N(b, F_t) \ dL(t) + \left(\int_{\mathcal{L}_\ell^c} dL(t) \right) \ J_N(b, F_u) \tag{34}$$

From (32) and (34) we see that we have an inequality of the following form

$$J_N(b, F_\ell) \le J_N(b, F_u) \le c + pJ_N(b, F_u), \tag{35}$$

where $c = \tau + \int_{\mathcal{L}_\ell} J_N(b, F_t) \ dL(t)$ and $p = \int_{\mathcal{L}_\ell^c} dL(t)$. Since $p \in [0, 1]$ we can write

$$J_N(b, F_\ell)(1-p) \le J_N(b, F_u)(1-p),$$

rearranging which we obtain,

$$\begin{aligned} J_N(b, F_\ell) \ &\le \ pJ_N(b, F_\ell) + J_N(b, F_u) - pJ_N(b, F_u) \\ &\overset{*}{\le} \ pJ_N(b, F_\ell) + c + pJ_N(b, F_u) - pJ_N(b, F_u) \\ &= \ c + pJ_N(b, F_\ell) \end{aligned}$$

where, to obtain $*$ we have used (35). Finally, note that

$$\begin{aligned} c + pJ_N(b, F_\ell) \ &= \ \tau + \int_{\mathcal{L}_\ell} J_N(b, F_t) \ dL(t) + \left(\int_{\mathcal{L}_\ell^c} dL(t) \right) J_N(b, F_\ell) \\ &= \ \tau + \mathbb{E}_L \Big[\min\{ J_N(b, F_\ell), J_N(b, F_{L_N}) \} \Big] \\ &= \ C_{N-1}(b, F_\ell). \end{aligned}$$

Thus, as desired we have shown $J_N(b, F_\ell) \le C_{N-1}(b, F_\ell)$ (recall the discussion leading to (33)).

Suppose that for some $k+1 = 2, 3, \cdots, N-1$ we have $\mathcal{S}_{k+1} \subseteq \mathcal{Q}_{k+1}^\ell$. We will have to show that the same holds for stage k. Fix any $b \in \mathcal{S}_k$, then for any generic distribution F_s, exactly as in (31) we have

$$\begin{aligned} \min\Big\{ -\eta b, P_k(b, F_s) \Big\} \ &= \ \min\Big\{ -\eta b, \eta\delta + \mathbb{E}_s \Big[J_k(\max\{b, R_s\}) \Big] \Big\} \\ &= \ \min\Big\{ -\eta b, \eta\delta - \eta \mathbb{E}_s \Big[\max\{b, R_s\} \Big] \Big\} \\ &= \ J_N(b, F_s). \end{aligned} \tag{36}$$

Thus the hypothesis $\mathcal{S}_k \subseteq \mathcal{Q}_k^u$ implies $J_N(b, F_u) \le C_k(b, F_u)$, and to show $\mathcal{S}_k \subseteq \mathcal{Q}_k^\ell$ it is sufficient to obtain $J_N(b, F_\ell) \le C_k(b, F_\ell)$. Proceeding as before (recall how (34) was obtained) we can write

$$C_k(b, F_u) \ \le \ \tau + \int_{\mathcal{L}_\ell} J_{k+1}(b, F_t) \ dL(t) + \left(\int_{\mathcal{L}_\ell^c} dL(t) \right) \ J_{k+1}(b, F_u).$$

Now using Lemma 5.4, we conclude

$$C_k(b, F_u) \ \le \ \tau + \int_{\mathcal{L}_\ell} J_{k+1}(b, F_t) \ dL(t) + \int_{\mathcal{L}_\ell^c} dL(t) \ J_N(b, F_u).$$

Note that the conditions required to apply Lemma 5.4 hold i.e., $b \in \mathcal{S}_{k+1}$ (since $\mathcal{S}_k \subseteq \mathcal{S}_{k+1}$ from Lemma 5.1-(iii)) and $\mathcal{S}_{k+1} \subseteq \mathcal{Q}_{k+1}^u$ (this is given).

Thus, again we have an inequality of the form $J_N(b, F_\ell) \le J_N(b, F_u) \le c' + pJ_N(b, F_u)$ (where $c' = \tau + \int_{\mathcal{L}_\ell} J_{k+1}(b, F_t) \ dL(t)$). As before we can show that $J_N(b, F_\ell) \le c' +$

$pJ_N(b, F_\ell)$. Finally the proof is complete by showing that $c' + pJ_N(b, F_\ell) = C_k(b, F_\ell)$ as follows:

$$\begin{aligned} C_k(b, F_\ell) &= \tau + \int_{\mathcal{L}_\ell} J_{k+1}(b, F_t)\, dL(t) + \int_{\mathcal{L}_\ell^c} J_{k+1}(b, F_\ell)\, dL(t) \\ &= c' + pJ_N(b, F_\ell), \end{aligned} \tag{37}$$

where to replace $J_{k+1}(b, F_\ell)$ by $J_N(b, F_\ell)$ we have to again apply Lemma 5.4. However this time $\mathcal{S}_{k+1} \subseteq \mathcal{Q}_{k+1}^{\ell}$, is by the induction hypothesis. □

We still require a distribution F_u satisfying $\mathcal{S}_k \subseteq \mathcal{Q}_k^u$, for every k. The minimum distribution F_m turns out to be useful in this context. The following lemma thus constitutes *Step 2* of the proof of Lemma 5.5.

LEMMA A.4. *For every $k = 1, 2, \cdots, N-1$, the set $\mathcal{Q}_k^m$ corresponding to the minimum distribution F_m satisfies, $\mathcal{S}_k \subseteq \mathcal{Q}_k^m$.*

PROOF. First note that the existence of a minimum distribution F_m follows from Lemma 2.3. Now, F_m being minimum we have $F_\ell \geq_{st} F_m$ for all F_ℓ. Then, using Lemma 4.2-(i) we can write

$$J_{k+1}(b, F_{L_{k+1}}) \leq J_{k+1}(b, F_m).$$

Using the above expression in (13) and then recalling (12), we obtain $C_k(b, F_m) = C_k(b)$. Finally, the result follows from the definition of the sets $\mathcal{Q}_k^m$ and $\mathcal{S}_k$. □

A.5. Proof of Theorem 5.7

Theorem 5.7*:* For $k = 1, 2, \cdots, N-1$ and for any F_ℓ, $\mathcal{S}_k^\ell = \mathcal{S}_{k+1}^\ell$.

PROOF. Recalling the definition of the set $\mathcal{S}_k^\ell$ (from (18)), for any $b \in \mathcal{S}_{k+1}^\ell$ we have (if $k+1 = N$, note that the following expression will not contain the continuing cost),

$$-\eta b \leq \min\Big\{P_{k+1}(b, F_\ell), C_{k+1}(b, F_\ell)\Big\}.$$

Suppose, as in Theorem 5.6, we can show that for any $b \in \mathcal{S}_{k+1}^\ell$, the various costs at stages k and $k+1$ are same, i.e., $P_k(b, F_\ell) = P_{k+1}(b, F_\ell)$ and $C_k(b, F_\ell) = C_{k+1}(b, F_\ell)$, then the above inequality would imply, $\mathcal{S}_k^\ell \supseteq \mathcal{S}_{k+1}^\ell$. The proof is complete by recalling that we already have $\mathcal{S}_k^\ell \subseteq \mathcal{S}_{k+1}^\ell$ (from Lemma 5.1-(iii)).

Fix a $b \in \mathcal{S}_{k+1}^\ell$. To show that $P_k(b, F_\ell) = P_{k+1}(b, F_\ell)$, first using Lemma 5.1-(i) and Theorem 5.6, note that $\mathcal{S}_{k+1}^\ell \subseteq \mathcal{S}_{k+1} = \mathcal{S}_k$. Since $b \in \mathcal{S}_{k+1}$ the cost of probing is

$$\begin{aligned} P_{k+1}(b, F_\ell) &= \eta\delta + \mathbb{E}_\ell\Big[J_{k+1}(\max\{b, R_\ell\})\Big] \\ &= \eta\delta - \eta\mathbb{E}_\ell\Big[\max\{b, R_\ell\}\Big] \end{aligned}$$

where, to obtain the second equality, note that $\max\{b, R_\ell\} \in \mathcal{S}_k$ (from Theorem 5.3) and hence at $\max\{b, R_\ell\}$ it is optimal to stop, so that $J_{k+1}(\max\{b, R_\ell\}) = -\eta \max\{b, R_\ell\}$. Similarly, since b is also in $\mathcal{S}_k$ the cost of probing at stage k, $P_k(b, F_\ell)$, is again $\eta\delta - \eta\mathbb{E}_\ell\Big[\max\{b, R_\ell\}\Big]$. Finally, following the same procedure used to show $C_k(b) = C_{k+1}(b)$ in Theorem 5.6, we can obtain $C_k(b, F_\ell) = C_{k+1}(b, F_\ell)$, thus completing the proof. □

REFERENCES

Andrea Abrardo, Lapo Balucanti, and Alessandro Mecocci. 2013. A Game Theory Distributed Approach for Energy Optimization in WSNs. *ACM Trans. Sen. Netw.* 9, 4, Article 44 (July 2013), 22 pages.

P. Agrawal and N. Patwari. 2009. Correlated Link Shadow Fading in Multi-Hop Wireless Networks. *IEEE Transactions on Wireless Communications* 8, 8 (2009), 4024–4036.

Kemal Akkaya and Mohamed Younis. 2005. A Survey on Routing Protocols for Wireless Sensor Networks. *Ad Hoc Networks* 3 (2005), 325–349.

S. Christian Albright. 1977. A Bayesian Approach to a Generalized House Selling Problem. *Management Science* 24, 4 (1977), 432–440.

Dimitri P. Bertsekas. 2005. *Dynamic Programming and Optimal Control, Vol. I*. Athena Scientific.

Dimitri P. Bertsekas and John N. Tsitsiklis. 1991. An Analysis of Stochastic Shortest Path Problems. *Mathematics of Operations Research* 16 (1991).

A. Bhattacharya, S.M. Ladwa, R. Srivastava, A. Mallya, a. Rao, D.G.R. Sahib, S.V.R. Anand, and A. Kumar. 2013. SmartConnect: A system for the design and deployment of wireless sensor networks. In *COMSNETS 13': Fifth International Conference on Communication Systems and Networks*.

Sanjit Biswas and Robert Morris. 2005. ExOR: Opportunistic Multi-hop Routing for Wireless Networks. *SIGCOMM Comput. Commun. Rev.* 35, 4 (August 2005), 133–144.

Ricardo C. Carrano, Diego Passos, Luiz C.S. Magalhes, and Clio V.N. Albuquerque. 2014. A Comprehensive Analysis on the use of Schedule-Based Asynchronous Duty Cycling in Wireless Sensor Networks. *Ad Hoc Networks* 16 (2014), 142 – 164.

Nicholas B. Chang and Mingyan Liu. 2007. Optimal Channel Probing and Transmission Scheduling for Opportunistic Spectrum Access. In *MobiCom '07: Proceedings of the 13th annual ACM international conference on Mobile computing and networking*. 27–38.

P. Chaporkar and A. Proutiere. 2008. Optimal Joint Probing and Transmission Strategy for Maximizing Throughput in Wireless Systems. *IEEE Journal on Selected Areas in Communications* 26, 8 (October 2008), 1546–1555.

Erhan Cinlar. 1975. *Introduction to Stochastic Processes*. Prentice-Hall.

Crossbow. 2006. TelosB Mote Platform. (2006). www.willow.co.uk/TelosB_Datasheet.pdf

Israel David and Ofer Levi. 2004. A New Algorithm for the Multi-item Exponentially Discounted Optimal Selection Problem. *European Journal of Operational Research* 153, 3 (2004), 782 – 789.

Euhanna Ghadimi, Olaf Landsiedel, Pablo Soldati, Simon Duquennoy, and Mikael Johansson. 2014. Opportunistic Routing in Low Duty-Cycle Wireless Sensor Networks. *ACM Transactions on Sensor Networks* 10, 4, Article 67 (June 2014), 39 pages.

Shuo Guo, Yu Gu, Bo Jiang, and Tian He. 2009. Opportunistic Flooding in Low-duty-cycle Wireless Sensor Networks with Unreliable Links. In *Proceedings of the 15th Annual International Conference on Mobile Computing and Networking (MobiCom '09)*. ACM, New York, NY, USA, 133–144.

Jie Hao, Baoxian Zhang, and H.T. Mouftah. 2012. Routing Protocols for Duty Cycled Wireless Sensor Networks: A Survey. *IEEE Communications Magazine* 50, 12 (December 2012), 116–123.

Ting Chao Hou and Victor Li. 1986. Transmission Range Control in Multihop Packet Radio Networks. *IEEE Transactions on Communications* 34, 1 (1986), 38–44.

Byung Kook Kang. 2005. Optimal Stopping Problem with Double Reservation Value Property. *European Journal of Operational Research* 165, 3 (2005), 765 – 785.

Samuel Karlin. 1962. *Stochastic Models and Optimal Policy for Selling an Asset*. Stanford University Press, Stanford. 148–158 pages.

Brad Karp and H. T. Kung. 2000. GPSR: Greedy Perimeter Stateless Routing for Wireless Networks. In *MobiCom '00: Proceedings of the 6th annual international conference on Mobile computing and networking*. ACM Press, New York, NY, USA, 243–254.

Dongsook Kim and Mingyan Liu. 2008. Optimal Stochastic Routing in Low Duty-cycled Wireless Sensor Networks. In *Proceedings of the 4th Annual International Conference on Wireless Internet (WICON '08)*. ICST (Institute for Computer Sciences, Social-Informatics and Telecommunications Engineering), ICST, Brussels, Belgium, Belgium, Article 56, 9 pages.

Joohwan Kim, Xiaojun Lin, and N.B. Shroff. 2011. Optimal Anycast Technique for Delay-Sensitive Energy-Constrained Asynchronous Sensor Networks. *IEEE/ACM Transactions on Networking* (April 2011).

Fabian Kuhn, Roger Wattenhofer, and Aaron Zollinger. 2008. An Algorithmic Approach to Geographic Routing in Ad Hoc and Sensor Networks. *IEEE/ACM Trans. Netw.* 16, 1 (2008), 51–62.

Anurag Kumar, D. Manjunath, and Joy Kuri. 2008. *Wireless Networking*. Morgan Kaufmann Publishers Inc., San Francisco, CA, USA.

Kumar et al. 2010. Wireless Sensor Networks for Human Intruder Detection. *Journal of the Indian Institute of Science, Special Issue on Advances in Electrical Science* 90, 3 (2010).

Zhenjiang Li, Mo Li, and Yunhao Liu. 2014. Towards Energy-Fairness in Asynchronous Duty-Cycling Sensor Networks. *ACM Transactions on Sensor Networks* 10, 3, Article 38 (May 2014), 26 pages.

Sha Liu, Kai Wei Fan, and P. Sinha. 2007. CMAC: An Energy Efficient MAC Layer Protocol Using Convergent Packet Forwarding for Wireless Sensor Networks. In *SECON '07: 4th Annual IEEE Communications Society Conference on Sensor, Mesh and Ad Hoc Communications and Networks*. 11–20.

Martin Mauve, Jrg Widmer, and Hannes Hartenstein. 2001. A Survey on Position-Based Routing in Mobile Ad-Hoc Networks. *IEEE Network* 15 (2001), 30–39.

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

K. P. Naveen and A. Kumar. 2012. Relay selection with Channel Probing for Geographical Forwarding in WSNs. In *10th International Symposium on Modeling and Optimization in Mobile, Ad Hoc and Wireless Networks (WiOpt 2012)*.

K. P. Naveen and A. Kumar. 2013. Relay Selection for Geographical Forwarding in Sleep-Wake Cycling Wireless Sensor Networks. *IEEE Transactions on Mobile Computing* 12, 3 (2013), 475–488.

R. Nelson and L. Kleinrock. 1984. The Spatial Capacity of a Slotted ALOHA Multihop Packet Radio Network with Capture. *IEEE Transactions on Communications* 32, 6 (1984), 684–694.

S. Ozen and S. Oktug. 2014. Forwarder Set Based Dynamic Duty Cycling in Asynchronous Wireless Sensor Networks. In *IEEE Wireless Communications and Networking Conference (WCNC)*. 2432–2437.

Vamsi Paruchuri, Shivakumar Basavaraju, Arjan Durresi, Rajgopal Kannan, and S. S. Iyengar. 2004. Random Asynchronous Wakeup Protocol for Sensor Networks. *International Conference on Broadband Networks* 0 (2004), 710–717.

C. Petrioli, M. Nati, P. Casari, M. Zorzi, and S. Basagni. 2014. ALBA-R: Load-Balancing Geographic Routing Around Connectivity Holes in Wireless Sensor Networks. *IEEE Transactions on Parallel and Distributed Systems* 25, 3 (March 2014), 529–539.

Martin L. Puterman. 1994. *Markov Decision Processes: Discrete Stochastic Dynamic Programming* (1st ed.). John Wiley & Sons, Inc., New York, NY, USA.

Theodore Rappaport. 2001. *Wireless Communications: Principles and Practice* (2nd ed.). Prentice Hall PTR, Upper Saddle River, NJ, USA.

Donald B. Rosenfield, Roy D. Shapiro, and David A. Butler. 1983. Optimal Strategies for Selling an Asset. *Management Science* 29, 9 (1983), 1051–1061.

Jungmin So and Heejung Byun. 2014. Opportunistic Routing with In-Network Aggregation for Asynchronous Duty-Cycled Wireless Sensor Networks. *Wireless Networks* 20, 5 (2014), 833–846.

Wolfgang Stadje. 1997. An Optimal Stopping Problem with Two Levels of Incomplete Information. *Mathematical Methods of Operations Research* 45 (1997), 119–131. Issue 1.

Dietrich Stoyan. 1983. *Comparison Methods for Queues and other Stochastic Models*. John Wiley & Sons, New York.

H. Takagi and L. Kleinrock. 1984. Optimal Transmission Ranges for Randomly Distributed Packet Radio Terminals. *IEEE Transactions on Communications [legacy, pre - 1988]* 32, 3 (1984), 246–257.

P. S. C. Thejaswi, Junshan Zhang, Man On Pun, H. V. Poor, and Dong Zheng. 2010. Distributed Opportunistic Scheduling with Two-Level Probing. *IEEE/ACM Transactions on Networking* 18, 5 (October 2010).

O.K. Tonguz, N. Wisitpongphan, J.S. Parikh, Fan Bai, P. Mudalige, and V.K. Sadekar. 2006. On the Broadcast Storm Problem in Ad hoc Wireless Networks. In *3rd International Conference on Broadband Communications, Networks and Systems*.

David Tse and Pramod Viswanath. 2005. *Fundamentals of wireless communication*. Cambridge University Press, New York, NY, USA.

Yu-Chee Tseng, Sze-Yao Ni, Yuh-Shyan Chen, and Jang-Ping Sheu. 2002. The Broadcast Storm Problem in a Mobile Ad Hoc Network. *Wireless Networks* 8, 2-3 (2002), 153–167.

M. Zorzi and R.R. Rao. 2003a. Geographic Random Forwarding (GeRaF) for Ad Hoc and Sensor Networks: Energy and Latency Performance. *IEEE Transactions on Mobile Computing* 2, 4 (2003), 349–365.

Michele Zorzi and Ramesh R. Rao. 2003b. Geographic Random Forwarding (GeRaF) for Ad Hoc and Sensor Networks: Multihop Performance. *IEEE Transactions on Mobile Computing* 2 (2003), 337–348.