Relay Selection for Geographical Forwarding in Sleep-Wake Cycling Wireless Sensor Networks

K. P. Naveen, *Student Member, IEEE* and Anurag Kumar, *Fellow, IEEE*

Abstract—Our work is motivated by geographical forwarding of sporadic alarm packets to a base station in a wireless sensor network (WSN), where the nodes are sleep-wake cycling periodically and asynchronously. We seek to develop local forwarding algorithms that can be tuned so as to trade-off the end-to-end delay against a *total cost*, such as the hop count or total energy. Our approach is to solve, at each *forwarding node* enroute to the sink, the *local forwarding problem* of minimizing one-hop waiting delay subject to a lower bound constraint on a suitable *reward* offered by the next-hop *relay*; the constraint serves to tune the trade-off. The reward metric used for the local problem is based on the end-to-end total cost objective (for instance, when the total cost is hop count, we choose to use the progress towards sink made by a relay as the reward). The forwarding node, to begin with, is uncertain about the number of relays, their wake-up times, and the reward values, but knows the probability distributions of these quantities. At each relay wake-up instant, when a relay reveals its reward value, the forwarding node's problem is to forward the packet or to wait for further relays to wake-up. In terms of the operations research literature, our work can be considered as a variant of the *asset selling problem*. We formulate our local forwarding problem as a partially observable Markov decision process (POMDP) and obtain *inner* and *outer* bounds for the optimal policy. Motivated by the computational complexity involved in the policies derived out of these bounds, we formulate an alternate *simplified model*, the optimal policy for which is a *simple* threshold rule. We provide simulation results to compare the performance of the inner and outer bound policies against the simple policy, and also against the optimal policy when the source knows the exact number of relays. Observing the good performance and the ease of implementation of the simple policy, we apply it to our motivating problem, i.e., local geographical routing of sporadic alarm packets in a large WSN. We compare the end-to-end performance (i.e., average total delay and average total cost) obtained by the simple policy, when used for local geographical forwarding, against that obtained by the globally optimal forwarding algorithm proposed by Kim et al. [1].

Index Terms—Relay selection, wireless sensor networks, sleep-wake cycling, geographical forwarding, asset selling problem, wireless networks with intermittent links, opportunistic forwarding.

1 Introduction

We are interested in the problem of packet forwarding in a class of wireless sensor networks (WSNs) in which local inferences based on sensor measurements could result in the generation of occasional "alarm" packets that need to be routed to a base-station, where some sort of action could be taken [1], [2], [3]. Such a situation could arise, for example, in a WSN for human intrusion detection or fire detection in a large region. Such WSNs often need to run on batteries or on harvested energy and, hence, must be energy conscious in all their operations. To conserve energy and also since the events are rare, it is best if the nodes are allowed to sleep-wake cycle, waking up only periodically to perform their tasks. In this work we consider *asynchronous* sleep-wake cycling [1], [4], where the sleep-wake process of each node is statistically independent of the sleep-wake process of any other node in the network.

Due to the asynchronous sleep-wake behaviour of the nodes, an alarm packet has to incur a random waiting delay at each hop enroute to the sink. The end-to-end performance metrics we are interested in are the average total delay and an average total cost (e.g., hop count, total power etc.,). To optimize the performance metrics one could use a distributed Bellman-Ford algorithm, e.g., the LOCAL-OPT algorthm proposed by Kim et al. [1]. However such a global solution requires a *pre-configuration phase* during which a globally optimal forwarding policy is obtained, and involves substantial control packets exchange. The focus of our research is, instead, towards designing *simple forwarding rules* based only on the *local information* available at a forwarding node (see Fig. 1). Towards this end the approach of *geographical forwarding* turns out to be useful. In geographical forwarding ([5], [6]) nodes know their own locations and that of the sink, and forward packets to neighbors that are closer to sink, i.e., to neighbors within the *forwarding region* (which is the hatched area in Fig. 1).

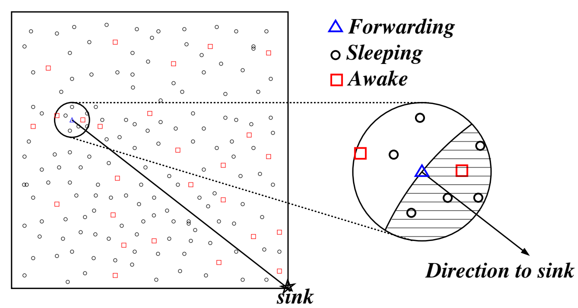


Fig. 1. Illustration of the local forwarding problem.

• *Both the authors are with the Dept. of Electrical Communication Engineering, Indian Institute of Science, Bangalore 560 012, India. E-mail:{naveenkp, anurag}@ece.iisc.ernet.in*

The local problem setting is the following. Somewhere in the network a node has just received a packet to forward (refer Fig. 1); for the *local problem* we refer to this forwarding node as the *source* and think of the time at which it gets the packet as 0. There is an *unknown number of relays* in the forwarding region of the source. In the geographical forwarding context, this lack of information on the number of relays could model the fact that the neighborhood of a forwarding node could vary over time due, for example, to node failures, variation in channel conditions, or (in a mobile network) the entry or exit of mobile relays. The source desires to forward the packet within the interval $(0, T)$, while knowing that the relays wake-up independently and uniformly over $(0, T)$. When a neighbor node wakes up, the source can evaluate it for its use as a relay, e.g., in terms of the progress it makes towards the destination node, the quality of the channel to the relay, the energy level of the relay, etc., (see [7], [8] for different routing metrics based on the above mentioned quantities). We think of this as a *reward* offered by the potential relay. Thus, at each relay wake-up instant, given the reward values of the relays that have woken up thus far, the source is faced with the sequential decision problem of whether to forward the packet or wait for further relays to wake-up.

By solving the local problem using a "suitable" reward metric, and then applying its solution at each hop towards the sink, we expect to capture the end-to-end problem of minimizing total delay subject to a constraint on an end-to-end total cost metric (e.g., hop count or total power). For instance, if the constraint is on hop count then it is reasonable to choose the local reward metric to be the progress, towards sink, made by a relay. Smaller end-to-end hop count can be achieved by using a larger progress constraint at each hop and vice versa. For total-power constraint we find that using a combination of one-hop power and progress as a reward for the local problem performs well for the end-to-end problem. We will formally introduce our local forwarding problem in Section 2 and discuss the end-to-end results in Section 7.2. Next we discuss related work and highlight our contributions.

1.1 Related Work

Although our work has been motivated by the geographical forwarding problem, outlined earlier, the local forwarding problem that we study also arises during channel selection in cognitive radio networks. Further, the local problem belongs to the class of *asset selling* problems, studied in the operations research literature. For completeness, we review the related literature from these areas as well.

Geographical forwarding problems: In our prior work [4] we have considered a simple model where the number of relays is a constant which is known to the source. There the reward is simply the progress made by a relay node towards the sink. In the current work we have generalized our earlier model by allowing the number of relays to be *not known* to the source. Also, here we allow a general reward structure.

There has been other work in the context of geographical forwarding and anycast routing, where the problem of choosing one among several neighboring nodes arises. Zorzi and Rao [9] propose a distributed relaying algorithm called GeRaF (Geographical Random Forwarding) whose objective is to carry a packet to its destination in as few hops as possible, by making as large progress as possible at each relaying stage. These authors do not consider the trade-off between the relay selection delay and the reward gained by selecting a relay, which is a major contribution of our work. Liu et al. [10] propose a relay selection approach as a part of CMAC, a protocol for geographical packet forwarding. Under CMAC, node i chooses an r_0 that minimizes the expected normalized latency (which is the average ratio of one-hop delay and progress). The Random Asynchronous Wakeup (RAW) protocol [11] also considers transmitting to the first node to wake-up that makes a progress of greater than a threshold. Interestingly, this is the structure of the optimal policy for our simplified model in Section 6. Thus we have provided analytical support for using such a threshold policy.

For a sleep-wake cycling network, Kim et al. in [1] have considered the problem of minimizing average end-to-end delay as a stochastic shortest path problem and have developed a distributed Bellman-Ford algorithm (referred to as the LOCAL-OPT) which yields optimal forwarding strategies for each node. However a major drawback is that a *pre-configuration phase* is required to run the LOCAL-OPT algorithm. We will discuss the work of Kim et al. [1] in detail in Section 7.2.

Channel selection problems: Akin to the relay selection problem is the problem of channel selection. The authors in [12], [13] consider a model where there are several channels available to choose from. The transmitter has to probe the channels to learn their quality. Probing many channels may yield one 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 start transmitting. Here the number of channels is known and all the channels are available at the very beginning of the decision process. In our problem the number of relays is not known, and they become available at random times.

Asset selling problems: The basic asset selling problem [14], [15], comprises N offers that arrive sequentially over discrete time slots. The offers are independent and identically distributed (iid). As the offers arrive, the seller has to decide whether to take an offer or wait for future ones. The seller has to pay a cost to observe the next offer. Previous offers cannot be recalled. The decision process ends with the seller choosing an offer. Over the years, several variants of the basic problem have been studied, both with and without recalling the previous offers. Recently Kang [16] 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. See [16] for further references to literature on models with uncertain recall. In [17], the authors consider a model in which the offers arrive at the points of a renewal process. In these models, either the number of potential offers is known or is infinite. In [18], a variant is studied in which the asset selling process can reach a deadline in the next slot with some fixed probability, provided that the process has proceeded upto the present slot.

In our work the number of offers (i.e., relays) is not known. Also the successive instants at which the offers arrive are the order statistics of an unknown number of iid uniform random variables over an interval $(0, T)$. After observing a relay, the probability that there are no more relays to go (which is the probability that the present stage is the last one) is not fixed. This probability has to be updated depending on the previous such probabilities and the inter wake-up times between the sucessive relays. Although our problem falls in the class of asset selling problems, to the best of our knowledge the particular setting we have considered in this paper has not been studied before.

1.2 Outline and Our Contributions

In Section 2 we formally describe our local forwarding problem of choosing a relay when the number of relays in the forwarding region is *unknown*. We then formulate it as a POMDP in Section 3. Before analysing the POMDP case, in Section 4 we recall, from our earlier work [4], the solution for the COMDP (Completely Observable MDP) version of the problem where the number of relays in the forwarding region is *known* to the source. For the POMDP, the optimal policy is characterized in terms of optimum stopping sets (Section 5). The main technical contributions are,

1) We prove that the the optimum stopping sets are convex (Section 5.1), and provide *inner* (*subset*, Section 5.2) and *outer bounds* (*superset*, Section 5.3) for it.

2) The computational complexity of the above bounds motivates us to consider a *simplified* model (Section 6). We prove that the optimal policy for this simplified model is a *simple* threshold rule.

3) We first perform one-hop simulations (Section 7.1) to compare the performance of the various policies derived out of the analysis. The performance of the simple policy turns out to be close to optimal.

4) Finally, we simulate a large WSN with sleep-wake cycling nodes and apply our simple policy at each hop enroute to the sink (Section 7.2). We compare the average total delay and average total cost obtained by the simple policy with that obtained by a distributed Bellman-Ford algorithm proposed by Kim et al. [1].

For the ease of presentation, we do not provide any proofs here. An interested reader can refer to our technical report [19].

2 Local Forwarding Problem

Recall that our local problem is motivated by the end-to-end problem of minimizing total delay subject to a total cost constraint. The total (end-to-end) cost is translated into a reward metric for the local problem. Formally, the local forwarding problem we consider is the following. A node in the network has just received a packet to forward. Abusing terminology, we call this node the "source" and the nodes that it could potentially forward the packet to are called "relays." The *local problem* is taken to start at time 0, and some of the associated processes are depicted in Fig. 2.

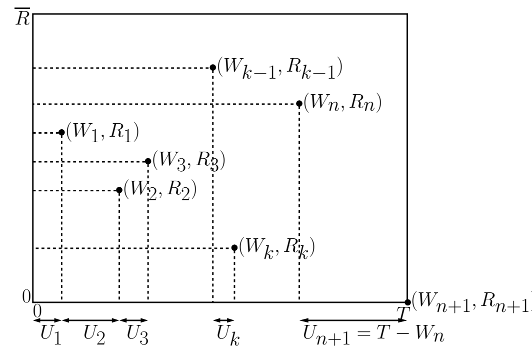


Fig. 2. When there are $N = n$ relays, then, for $k = 1, 2, \cdots, n$, (W_k, R_k) represents the wake-up instant and reward repectively, of the kth relay. These are shown as points in $[0, T] \times [0, \overline{R}]$.

There is a *nonempty* set of N relay nodes, labeled by the indices $1, 2, \cdots, N$. N is a random variable bounded above by K, a system parameter that is known to the source node, i.e., the support of N is $\{1, 2, \cdots, K\}$. The source does not know N, but knows the bound K, and a probability mass function (pmf) p_0 on $\{1, 2, \cdots, K\}$, which is the initial pmf of N. A relay node i, $1 \le i \le N$, becomes available to the source at the instant T_i. The source knows that the instants $\{T_i\}$ are iid uniformly distributed on $(0, T)$[1]. We call T_i the *wake-up instant* of relay i. Given that $N = n$ (throughout this discussion we will focus on the event $(N = n)$), let $W_1, W_2, \cdots, W_n$ represent the order statistics of $T_1, T_2, \cdots, T_n$, i.e., the $\{W_k\}$ sequence is the $\{T_i\}$ sequence sorted in the increasing order. Let $W_0 = 0$ and define $U_k = W_k - W_{k-1}$ for $k = 1, 2, \cdots, n$. U_k are the *inter-wake-up* time instants between the consecutive nodes (see Fig. 2).

Definition 1: For Simplicity we will use the following notations to denote the conditional pdf of U_{k+1} ($k = 0, 1, \cdots, n-1$) and the conditional expectation (both conditioned on W_k and N),

$$f_k(u|w, n) \;:=\; f_{U_{k+1}|W_k, N}(u|w, n),$$
$$\mathbb{E}_k[\cdot|w, n] \;:=\; \mathbb{E}[\cdot|W_k = w, N = n].$$

1. Such a model would arise if each node wakes up periodically with period T, and the point processes of the wake up instants across the nodes are stationary and independent versions of the wake-up process with period T.

If the source forwards the packet to the relay i, then a *reward* of R_i is accrued. The rewards $R_i, i = 1, 2, \cdots, n$, are iid random variables with probability density function (pdf) f_R. The support of f_R is $[0, \overline{R}]$. The source knows this statistical characterisation of the rewards, and also that the $\{R_i\}$ are independent of the wake-up instants $\{T_i\}$. When a relay wakes up at T_i and reveals its reward R_i, the source has to decide whether to transmit to relay i or to wait for further relays. *If the source decides to wait, then it instructs the relay with the best reward to stay awake, while letting the rest go back to sleep*. This way the source can always forward to a relay with the best reward among those that have woken up so far.

Since the reward sequence $R_1, R_2, \cdots, R_n$ is iid and independent of the wake-up instants $T_1, T_2, \cdots, T_n$, we write (W_k, R_k) as the pairs of ordered wake-up instants and the corresponding rewards. Evidently, $f_{R_{k+1}|W_k,N}(r|w,n) = f_R(r)$ for $k = 0, 1, \cdots, n-1$. Further we define (when $N = n$) $W_{n+1} := T$, $U_{n+1} := (T - W_n)$ and $R_{n+1} := 0$. Also $\mathbb{E}_n[U_{n+1}|w, n] := T - w$. All these variables are depicted in Fig. 2.

Decision Instants and Actions: We assume that the time instants at which the relays wake-up, i.e., $W_1, W_2, \cdots$, constitute the *decision instants or stages*[2]. At each decision instant, there are two possible actions available to the source, denoted 0 and 1, where

- 0 represents the action to *continue* waiting for more relays to wake-up, and
- 1 represents the action to *stop* and forward the packet to the relay that provides the best reward among those that have woken up thus far.

Since there can be at most K relays, the total number of possible decision instants is K.

Stopping Rules and the Optimization Problem: If the source has not forwarded the packet until stage $k-1$ then define, $I_k := (p_0, (W_1, R_1), \cdots, (W_k, R_k))$ to be the *information vector* (or the history) available at the source when the k-th relay (provided $N \geq k$) wakes up. Now, the source's decision whether to stop or continue should be based on I_k. Formally, define a *stopping rule* or a *policy*, π, as a sequence of mappings $(\mu_1, \cdots, \mu_K)$ where each μ_k maps I_k to the action set $\{0, 1\}$. Let Π represent the set of all policies. The delay, D_π, incurred using policy π is the instant at which the source forwards the packet. It could be either one of the W_k, or the instant T (which is possible if, even at W_N, the source decides to continue, not knowing that it has seen all the relays). The reward R_π is the reward associated with the relay to which the packet is forwarded. It is possible for the source to maximize the reward by waiting until time T (incurring maximum delay) and then choosing the best relay. On the other hand, the source can minimize the delay by forwarding to the first relay, irrespective of its reward value, but at the expense of depriving itself of an opportunity to choose a better relay that could wake-up later. Thus, there is a trade-off between these two quantities which we are interested in studying. Formally, the problem we are interested in is the following,

$$\begin{aligned} &\min_{\pi \in \Pi} \quad \mathbb{E}D_\pi \\ &\text{Subject to} \quad \mathbb{E}R_\pi \geq \gamma. \end{aligned} \tag{1}$$

We introduce a Lagrange multiplier $\eta > 0$ and consider solving the following unconstrained problem,

$$\min_{\pi \in \Pi} \Big(\mathbb{E}D_\pi - \eta \mathbb{E}R_\pi \Big). \tag{2}$$

The following lemma relates both the problems,

Lemma 1: Let π^* be an optimal policy for the unconstrained problem in (2). Suppose the chosen η is such that $\mathbb{E}R_{\pi^*} = \gamma$, then π^* is optimal for the main problem in (1) as well.

Proof: See [19, Lemma 1]. □

3 POMDP Formulation

With the number of relays being unknown, the natural approach is to formulate the problem as a partially observed Markov decision process (POMDP). A POMDP is a generalization of an MDP, where at each stage the actual internal state of the system is not available to the controller. Instead, the controller can observe a value from an observation space. The observation probabilistically depends on the current actual state and the previous action. In some cases, a POMDP can be converted to an equivalent MDP by regarding a belief (i.e., a probability distribution) on the state space as the state of the equivalent MDP. For a survey of POMDPs see [20], [21, Chapter 5]. In this section we will set up the unconstrained problem in (2) as a POMDP.

3.1 Belief State and Belief State Transition

Since the source does not know the actual number of relays N, the state is only partially observable. As mentioned earlier the source's decision should be based on the entire history of the information vector $I_k = (p_0, (w_1, b_1), \cdots, (w_k, b_k))$ where $w_1, \cdots, w_k$ represents the wake-up instants of relays waking up at stages $1, \cdots, k$, and $b_1, \cdots, b_k$ are the corresponding best rewards. Define p_k to be the *belief state* about N at stage k given the information vector I_k, i.e., $p_k(n) = \mathbb{P}(N = n | I_k)$ for $n = k, k+1, \cdots, K$ (note that $p_k(k)$ is the probability that the k th relay is the last one). *Thus, p_k is a pmf in the $K-k$ dimensional probability simplex*. Let us denote this simplex as $\mathcal{P}_k$.

Definition 2: For $k = 1, 2, \cdots, K$, let $\mathcal{P}_k$:= set of all pmfs on the set $\{k, k+1, \cdots, K\}$. $\mathcal{P}_k$ is the $K-k$ dimensional probability simplex in $\Re^{K-k+1}$. □

The "observation" (w_k, b_k) at stage k is a part of the actual state (N, w_k, b_k). For a general POMDP problem the

2. A better choice for the decision instants may be to allow the source to take decision at any time $t \in (0, T]$. When N is known to the source it can be argued that it is optimal to take decisions only at relay wake-up instances. However, this may not hold for our case where N is unknown. In this paper we proceed with our restriction on the decision instants and consider the general case as a topic for future work.

$$c_{k,n}(p,w,b) = \mathbb{E}_k\left[U_{k+1} + J_{k+1}\Big(\tau_{k+1}(p,w,U_{k+1}), w+U_{k+1}, \max\{b, R_{k+1}\}\Big)\Big| w, n\right] \quad (3)$$

observation can belong to a completely different space than the actual state space. Moreover the distribution of the observation at any stage can in general depend on all the previous states, observations, actions and disturbances. Suppose this distribution depends only on the state, action and disturbance of the immediately preceding stage, then a belief on the actual state given the entire history turns out to be sufficient for taking decisions [21, Chapter 5]. For our case, this condition is met and hence at stage k, (p_k, w_k, b_k) is a *sufficient statistic* to take decision. Hence the state space at stage $k = 1, 2\cdots, K$, for our POMDP problem is

$$\mathcal{S}_k = \Big\{(p,w,b) : p \in \mathcal{P}_k, w \in (0,T], b \in [0,\overline{R}]\Big\} \cup \{\psi\}, \quad (4)$$

with the initial state (i.e., state at stage 0) being $(p_0, 0, 0)$. Here ψ is the *terminating state*. The state at stage k will be ψ, if the source has already forwarded the packet at an earlier stage. The decision process ends once the system enters ψ.

After seeing k relays, suppose the source chooses not to forward the packet, then upon the next relay waking up (if any), the source needs to update its belief about the number of relays. Formally, if $(p,w,b) \in \mathcal{S}_k$ is the state at stage k and $(w+u)$ is the wake-up instant of the next relay then, using Bayes rule, the next belief state can be obtained via the following *belief state transition function* which yields a pmf in $\mathcal{P}_{k+1}$,

$$\tau_{k+1}(p,w,u)(n) = \frac{p(n)f_k(u|w,n)}{\sum_{\ell=k+1}^{K} p(\ell)f_k(u|w,\ell)} \quad (5)$$

for $n = k+1, \cdots, K$. Note that this function does not depend on b. Thus, if at stage $k \in \{0,1,\cdots,K-1\}$, the state is $(p,w,b) \in \mathcal{S}_k$, then the next state is

$$s_{k+1} = \begin{cases} \psi \text{ if } w = T \text{ and/or } a_k = 1 \\ \Big(\tau_{k+1}(p,w,U_{k+1}), w+U_{k+1}, \max\{b, R_{k+1}\}\Big) \\ \text{otherwise,} \end{cases} \quad (6)$$

When the action $a_k = 1$ the source enters ψ. Further the source decides to stop (i.e., enter ψ) even when $w = T$. This is because it knows that the wake-up time of each relay is strictly less than T. Such a situation can arise when, at stage k, the actual number of relays happened to be k and the source decides to continue (possible because the source does not know the actual number). Then the source will end up waiting until time T and then transmit to the relay with the best reward.

3.2 One-Step Costs

The objective in (2) can be seen as accumulating additively over each step. If the decision at a stage is to continue then the delay until the next relay wakes up (or until T) gets added to the cost. On the other hand if the decision is to stop then the source collects the reward offered by the relay to which it forwards the packet and the decision process enters the state ψ. The cost in state ψ is 0. Suppose (p,w,b) is the state at stage k. Then the one-step-cost function is, for $k = 0, 1, \cdots, K-1$,

$$g_k\Big((p,w,b), a_k\Big) = \begin{cases} -\eta b & \text{if } w = T \text{ and/or } a_k = 1 \\ U_{k+1} & \text{otherwise.} \end{cases} \quad (7)$$

The cost of termination is $g_K(p,w,b) = -\eta b$. Also note that for $k = 0$, the possible state is $(p_0, 0, 0)$ and the only possible action is $a_0 = 1$, so that $g_0\Big((p,0,0), a_0\Big) = U_1$. Note that, for a given policy π if s_k represent the state at stage k then, $\sum_{k=0}^{K} g_k(s_k, \mu_k(s_k)) = D_\pi - \eta R_\pi$.

3.3 Optimal Cost-to-go Functions

For $k = 1, 2, \cdots, K$, let $J_k(\cdot)$ represent the *optimal cost-to-go function* at stage k. For any state $s_k \in \mathcal{S}_k$, $J_k(s_k)$ can be written as,

$$J_k(s_k) = \min\{\textit{stopping cost}, \textit{continuing cost}\}, \quad (8)$$

where *stopping cost* (*continuing cost*) represents the average cost incurred, if the source, at the current stage decides to stop (continue), and takes optimal action at the subsequent stages. For the termination state, since the one step cost is zero and since the system remains in ψ in all the subsequent stages, we have $J_k(\psi) = 0$. For a state $(p,w,b) \in \mathcal{S}_k$, we next evaluate the two costs.

First let us obtain the stopping cost. Suppose that there were K relay nodes and the source has seen them all. In such a case if $(p,w,b) \in \mathcal{S}_K$ (note that p will just be a point mass on K) is the state at stage K then the optimal cost is simply the cost of termination, i.e., $J_K(p,w,b) = g_K(p,w,b) = -\eta b$. For $k = 1, 2, \cdots, K-1$, if the action is to stop then the one step cost is $-\eta b$ and the next state is ψ so that the further cost is $J_{k+1}(\psi) = 0$. Therefore, the stopping cost at any stage is simply $-\eta b$.

On the other hand the cost for continuing, when the state at stage k is $(p,w,b) \in \mathcal{S}_k$, using the total expectation law, can be written as,

$$c_k(p,w,b) = p(k)\Big(T - w - \eta b\Big) + \sum_{n=k+1}^{K} p(n)c_{k,n}(p,w,b) \quad (9)$$

where $c_{k,n}(p,w,b)$ is the average cost to continue conditioned on the event $(N = n)$ (see (3)), the probability of which is $p(n)$. Referring to (3), U_{k+1} is the (random) time until the next relay wakes up (U_{k+1} is the one step cost) and $J_{k+1}(\cdot)$ is the optimal cost-to-go from the next stage onwards ($J_{k+1}(\cdot)$ constitutes the future cost). The next state is obtained via the state transition equation (6). The term $(T - w - \eta b)$ in (9) associated with $p(k)$ is the cost of continuing when the number of relays happen to be k, i.e., $(N = k)$ and there are no more relays to go. Recall that we had defined (in Section 2)

$$\phi_\ell(w,b) = \mathbb{E}_{K-\ell}\left[\max\left\{b, R, \phi_{\ell-1}\Big(w+U, \max\{b,R\}\Big)\right\} - \frac{U}{\eta}\middle| w, K\right] \tag{10}$$

$U_{k+1} = T - w$ and $R_{k+1} = 0$ when the actual number of relays is $N = k$. Therefore $T - w$ is the one step cost when $N = k$. Also $w + U_{k+1} = T$ and $\max\{b, R_{k+1}\} = b$ so that at the next stage (which occurs at T) the process will terminate (enter ψ) with a cost of $-\eta b$ (see (6) and (7)), which represents the future cost.

Thus the optimal cost-to-go function (8) at stage $k = 1, 2, \cdots, K-1$, can be written as,

$$J_k(p,w,b) = \min\Big\{-\eta b, c_k(p,w,b)\Big\}. \tag{11}$$

From (11) it is clear that at stage k when the state is (p,w,b), the source has to compare the stopping cost, $-\eta b$, with the cost of continuing, $c_k(p,w,b)$, and stop iff $-\eta b \leq c_k(p,w,b)$. Later in Section 5, we will use this condition ($-\eta b \leq c_k(p,w,b)$) and define, the *optimum stopping set*. We will prove that the continuing cost, $c_k(p,w,b)$, is concave in p, leading to the result that the optimum stopping set is convex. (9) and (11) are extensively used in the subsequent development.

4 Relationship with the Case Where N is Known (the COMDP Version)

In the previous section (Section 3) we detailed our problem formulation as a POMDP. The state is partially observable because the source does not know the exact number of relays. It is interesting to first consider the simpler case where this number is known, which is the contribution of our earlier work in [4]. Hence, in this section, we will consider the case when the initial pmf, p_0, has all the mass only on some n, i.e., $p_0(n) = 1$. We call this, the COMDP version of the problem.

First we define a sequence of threshold functions which will be useful in the subsequent proofs. These are the same threshold functions that characterize the optimal policy for our model in [4].

Definition 3: For $(w,b) \in (0,T) \times [0,\overline{R}]$, define $\{\phi_\ell : \ell = 0, 1, \cdots, K-1\}$ inductively as follows: for all (w,b) $\phi_0(w,b) = 0$ and for $\ell = 1, 2, \cdots, K-1$ (recall Definition 1), $\phi_\ell(w,b)$ as in (10). In (10) we have suppressed the subscript $K - \ell + 1$ for R and U for simplicity. The pdf used to take the expectation in the above expression is $f_R(\cdot)f_{K-\ell}(\cdot|w,K)$ (again recall Definition 1). □

We will need the following simple property of the threshold functions in a later section.

Lemma 2: For $\ell = 1, 2, \cdots, K-1$, $-\eta\phi_\ell(w,b) \leq (T - w - \eta b)$.

Proof: See [19, Appendix I.A]. □

Next we state the main lemma of this section. We call this the *One-point* Lemma, because it gives the optimal cost, $J_k(p_k,w,b)$, at stage k when the belief state $p_k \in \mathcal{P}_k$ is such that it has all the mass on some $n \geq k$.

Lemma 3 (One-point): Fix some $n \in \{1, 2, \cdots, K\}$ and $(w,b) \in (0,T) \times [0,\overline{R}]$. For any $k = 1, 2, \cdots, n$, if $p_k \in \mathcal{P}_k$ is such that $p_k(n) = 1$ then,

$$J_k(p_k,w,b) = \min\Big\{-\eta b, -\eta\phi_{n-k}(w,b)\Big\}.$$

Proof: The proof is by induction. We make use of the fact that if at some stage $k < n$ the belief state p_k is such that $p_k(n) = 1$ then the next belief state $p_{k+1} (\in \mathcal{P}_{k+1})$, obtained by using the belief transition equation (5), is also of the form $p_{k+1}(n) = 1$. We complete the proof by using Definition 3 and the induction hypothesis. For a complete proof, see [19, Appendix I.B]. □

Discussion of Lemma 3: At stage k if the state is (p_k,w,b), where p_k is such that $p_k(n) = 1$ for some $n \geq k$, then from the One-point Lemma it follows that the optimal policy is to stop and transmit iff $b \geq \phi_{n-k}(w,b)$. The subscript $n-k$ of the function ϕ_{n-k} signifies the number of more relays to go. For instance, if we know that there are exactly 4 more relays to go then the threshold to be used is ϕ_4. Suppose at stage k if it was optimal to continue, then from (5) it follows that the next belief state $p_{k+1} \in \mathcal{P}_{k+1}$ also has mass only on $(N = n)$ and hence at this stage it is optimal to use the threshold function $\phi_{n-(k+1)}$. Therefore, if we begin with an intial belief $p_0 \in \mathcal{P}_1$ such that $p_0(n) = 1$ for some n, then the optimal policy is to stop at the first stage k such that $b \geq \phi_{n-k}(w,b)$ where $W_k = w$ is the wake-up instant of the k th relay and $B_k = \max\{R_1, \cdots, R_k\} = b$. Note that, since at stage n the threshold to be used is $\phi_0(w,b) = 0$ (see Definition 3), we invariably have to stop at stage n if we have not terminated earlier. This is exactly the same as our optimal policy in [4], where the number of relays is known to the source (instead of knowing the number wp1, as in our One-point Lemma here). □

5 POMDP Version: Bounds on the Optimum Stopping Set

In this section we will consider the general case developed in Section 3 where the number of relays N is not known to the source. There are several algorithms available to exactly obtain the optimal policy for a POMDP problem when the actual state space is finite [22], starting from the seminal work of Smallwood and Sondik [23]. However when the number of states is large, these algorithms are computationally intensive. In general, it is not easy to obtain an optimal policy for a POMDP. In this section, we have characterized the optimal policy in terms of *optimum stopping sets* and prove *inner* and *outer* bounds for this set.

Definition 4 (Optimum stopping set): For $1 \leq k \leq K-1$, let $\mathcal{C}_k(w,b) = \Big\{p \in \mathcal{P}_k : -\eta b \leq c_k(p,w,b)\Big\}$. Referring to (11) it follows that, for a given (w,b), $\mathcal{C}_k(w,b)$ represents the set of all beliefs $p \in \mathcal{P}_k$ at stage k at which it is optimal to stop. We call $\mathcal{C}_k(w,b)$ the *optimum stopping set* at stage k when the delay (W_k) and best reward (B_k) values are w and b, respectively. □

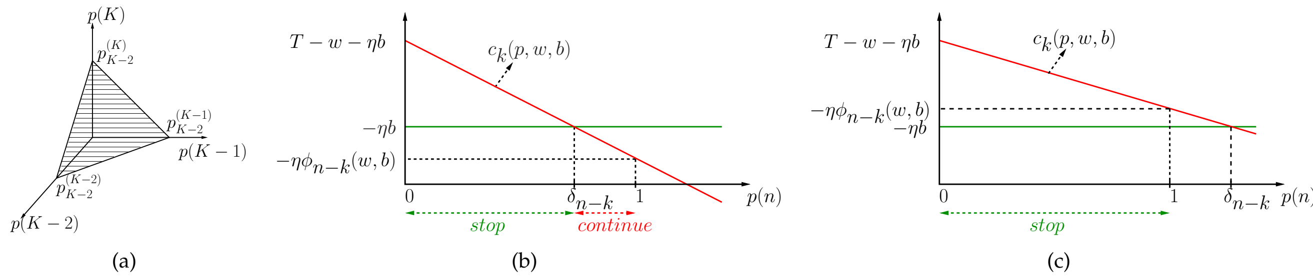


Fig. 3. (a) Probability simplex, $\mathcal{P}_{K-2}$ at stage $K-2$. A belief at stage $K-2$ is a pmf on the points $K-2$, $K-1$ and K (i.e., no-more, one-more and two-more relays to go, respectively). Thus $\mathcal{P}_{K-2}$ is a two dimensional simplex in $\Re^3$. (b) and (c): $c_k(p,w,b)$ in (13) is plotted as a function of $p(n)$. Also shown is the constant function $-\eta b$ which is the stopping cost. δ_{n-k} is the point of intersection of these two functions. (b) When $b \leq \phi_{n-k}$. (c) When $b > \phi_{n-k}(w,b)$.

5.1 Convexity of the Optimum Stopping Sets

We will prove (in Lemma 4) that the continuing cost, $c_k(p,w,b)$, in (3) is concave in $p \in \mathcal{P}_k$. From the form of the stopping set $\mathcal{C}_k(w,b)$, a simple consequence of this lemma will be that the optimum stopping set is convex. We further extend the concavity result of $c_k(p,w,b)$ for $p \in \overline{\mathcal{P}}_k$, where $\overline{\mathcal{P}}_k$ is the affine set containing $\mathcal{P}_k$ (to be defined shortly in this section).

Lemma 4: For $k = 1,2,\cdots,K-1$, and any given (w,b), the cost of continuing (defined in (3)), $c_k(\cdot,w,b)$, is concave on $\mathcal{P}_k$.

Proof: The essence of the proof is same as that in [24, Lemma 1]. Formal proof is available in our technical report [19, Appendix II.A]. □

The following corollary is a straight forward application of the above lemma.

Corollary 1: For $k = 1,2,\cdots,K-1$, and any given (w,b), $\mathcal{C}_k(w,b) (\subseteq \mathcal{P}_k)$ is a convex set.

Proof: From Lemma 4 we know that $c_k(p,w,b)$ is a concave function of $p \in \mathcal{P}_k$. Hence $\mathcal{C}_k(w,b)$ (see Definition 4), being a *super level set* of a concave function, is convex [25]. □

In the next section while proving an inner bound for the stopping set $\mathcal{C}_k(w,b)$, we will identify a set of points that could lie outside the probability simplex $\mathcal{P}_k$. We can obtain a better inner bound if we extend the concavity result to the affine set, $\overline{\mathcal{P}}_k = \left\{p \in \Re^{K-k+1} : \langle p, \mathbf{1}\rangle = 1\right\}$, where $\langle p, \mathbf{1}\rangle = \sum_{n=k}^{K} p(n)$, i.e., in $\overline{\mathcal{P}}_k$ the vectors sum to one, but we do not require non-negativity of the vectors. This can be done as follows. Define $\overline{\tau}_{k+1}(p,w,u)$ using (5) for every $p \in \overline{\mathcal{P}}_k$. Then $\overline{\tau}_{k+1}(.,w,u)$ as a function of p, is the extension of $\tau_{k+1}(.,w,u)$ from $\mathcal{P}_k$ to $\overline{\mathcal{P}}_k$. Similarly, for every $p \in \overline{\mathcal{P}}_k$, define $\overline{c}_k(p,w,b)$ and $\overline{J}_k(p,w,b)$ using (3) and (11). These are the extensions of $c_k(\cdot,w,b)$ and $J_k(\cdot,w,b)$ respectively. Then again, using the proof technique same as that in Lemma 4, we can obtain the following corollary,

Corollary 2: For $k = 1,2,\cdots,K-1$, and any given (w,b), $\overline{c}_k(\cdot,w,b)$ is concave on the affine set $\overline{\mathcal{P}}_k$. □

Using the above corollary, $\mathcal{C}_k(w,b)$ can be written as,

$$\mathcal{C}_k(w,b) = \mathcal{P}_k \cap \Big\{p \in \Re^{K-k+1} : \langle p, \mathbf{1}\rangle = 1, -\eta b \leq \overline{c}_k(p,w,b)\Big\}. \quad (12)$$

5.2 Inner Bound on the Optimum Stopping Set

We have showed that the optimum stopping set is convex. In this section, we will identify points that lie along certain edges of the simplex $\mathcal{P}_k$. A convex hull of these points will yield an inner bound to the optimum stopping set. This will first require us to prove the following lemma, referred to as the *Two-points* Lemma, and is a generalization of the One-point Lemma (Lemma 3). It gives the optimal cost, $J_k(p,w,b)$, at stage k when $p \in \mathcal{P}_k$ is such that it places all its mass on k and on some $n > k$, i.e., $p(k)+p(n) = 1$. Throughout this and the next section (on an outer bound) $(W_k, B_k) = (w,b)$ is fixed and hence, for the ease of presentation (and readability), we drop (w,b) from the notations $\delta_\ell(w,b)$, $a_k^\ell(w,b)$ and $b_k^\ell(w,b)$ (to appear in these sections later). However it is understood that these thresholds are, in general, functions of (w,b).

Lemma 5 (Two-points): For $k = 1,2,\cdots,K-1$, if $p \in \mathcal{P}_k$ is such that $p(k)+p(n) = 1$, where $k < n \leq K$ then,

$$J_k(p,w,b) = \min\Big\{-\eta b, p(k)\Big(T-w-\eta b\Big) + p(n)\Big(-\eta\phi_{n-k}(w,b)\Big)\Big\}.$$

Proof: Available in our technical report [19, Lemma 5]. □

Discussion of Lemma 5: The Two-points Lemma (Lemma 5) can be used to obtain certain threshold points in the following way. When $p \in \mathcal{P}_k$ has mass only on k and on some n, $k < n \leq K$, then using Lemma 5, the continuing cost can be written as a function of $p(n)$ as,

$$c_k(p,w,b) = \Big(T-w-\eta b\Big) - p(n)\Big(T-w-\eta\Big(b-\phi_{n-k}(w,b)\Big)\Big) \quad (13)$$

From Lemma 2, it follows that $c_k(p,w,b)$ in (13) is a decreasing function of $p(n)$. Let $p_k^{(k)}$ and $p_k^{(n)}$ be pmfs in $\mathcal{P}_k$ with mass only on $N = k$ and $N = n$ respectively. These are two of the corner points of the simplex $\mathcal{P}_k$ (as an example, Fig. 3(a) illustrates the simplex and the corner points for stage $k = K-2$. With at most two more nodes to go, $\mathcal{P}_{K-2}$ is a two dimensional simplex in $\Re^3$. $p_{K-2}^{(K-2)}$, $p_{K-2}^{(K-1)}$ and $p_{K-2}^{(K)}$ are the corner points of this simplex).

At stage k as we move along the line joining the points $p_k^{(k)}$ and $p_k^{(n)}$ (Fig. 3(b) and 3(c) illustrates this as $p(n)$ going from 0 to 1), the cost of continuing in (13) decreases and there is a threshold below which it is optimal to transmit and beyond which it is optimal to continue. The value of this threshold is that value of $p(n)$ in (13) at which the continuing cost becomes equal to $-\eta b$. Let δ_{n-k} denote this threshold value, then

$$\delta_{n-k} = \frac{T-w}{T-w-\eta\Big(b-\phi_{n-k}(w,b)\Big)}.$$

The cost of continuing in (13) as a function of $p(n)$ along with the stopping cost, $-\eta b$, is shown in Fig. 3(b) and 3(c). The threshold δ_{n-k} is the point of intersection of these two cost functions. The value of the continuing cost $c_k(p,w,b)$ at $p(n)=1$ is $-\eta\phi_{n-k}(w,b)$. Note that in the case when $b>\phi_{n-k}(w,b)$ the threshold δ_{n-k} will be greater than 1 in which case it is optimal to stop for any p on the line joining $p_k^{(k)}$ and $p_k^{(n)}$. □

There are similar thresholds along each edge of the simplex $\mathcal{P}_k$ starting from the corner point $p_k^{(k)}$. In general, let us define for $\ell=1,2,\cdots,K$,

$$\delta_\ell = \frac{T-w}{T-w-\eta\Big(b-\phi_\ell(w,b)\Big)}. \quad (14)$$

Remark: Note that (13) will also hold for the extended function $\overline{c}_k(p,w,b)$, where now $p\in\overline{\mathcal{P}}_k$. In terms of the extended function, δ_{n-k} represents the value of $p(n)$ (in (13) with c_k replaced by $\overline{c}_k$) at which $\overline{c}_k(p,w,b)=-\eta b$.

Recall that (from Lemma 5) the above discussion began with a $p\in\mathcal{P}_k$ such that $p(k)+p(n)=1$. At the threshold of interest we have $p(n)=\delta_{n-k}$ and hence $p(k)=1-\delta_{n-k}$, and the rest of the components are zero. We denote this vector as a_k^{n-k}. For instance in Fig. 4, where the face of the two dimensional simplex $\mathcal{P}_{K-2}$ is shown, the threshold along the lower edge of the simplex is $a_{K-2}^1=[1-\delta_1,\delta_1,0]$ and that along the other edge is $a_{K-2}^2=[1-\delta_2,0,\delta_2]$. Since it is possible for $\delta_{n-k}>1$, therefore the vector threshold a_k^{n-k} is not restricted to lie in the simplex $\mathcal{P}_k$, however it always stays in the affine set $\overline{\mathcal{P}}_k$. We formally define these thresholds next.

Definition 5: For a given $k\in\{1,2,\cdots,K-1\}$, for each $\ell=1,2,\cdots,K-k$ define a_k^ℓ as a $K-k+1$ dimensional point with the first and the $\ell+1$ th components equal to $1-\delta_\ell$ and δ_ℓ respectively, the rest of the components are zeros. As mentioned before, a_k^ℓ lies on the line joining $p_k^{(k)}$ and $p_k^{(k+\ell)}$. At stage k there are $K-k$ such points, one corresponding to each edge in $\mathcal{P}_k$ emanating from the corner point $p_k^{(k)}$. For an illustration of these points see Fig. 4 for the case $k=K-2$. □

Referring to Fig. 4(a) (which depicts the case, $k=K-2$), suppose all the vector thresholds, a_k^l, lie within the simplex $\mathcal{P}_k$ then, since at these points the stopping cost ($-\eta b$) is equal to the continuing cost ($c_k(a_k^l,w,b)$), all these points lie in the optimum stopping set $\mathcal{C}_k(w,b)$. Note that the corner point $p_k^{(k)}$ (belief with all the mass on no-more relays to go) also lies in $\mathcal{C}_k(w,b)$. Since we have already shown that $\mathcal{C}_k(w,b)$ is convex, the convex hull of these points will yield an inner bound. However as mentioned earlier (and as depicted in Fig. 4(b) and 4(c)) it is possible for some or all the thresholds a_k^l to lie outside the simplex (and hence these thresholds do not belong to $\mathcal{C}_k(w,b)$). This is where we will use Corollary 2, where the concavity result of the continuing cost, $c_k(p,w,b)$, is extended to the affine set $\overline{\mathcal{P}}_k$. We next state this *inner bound* theorem:

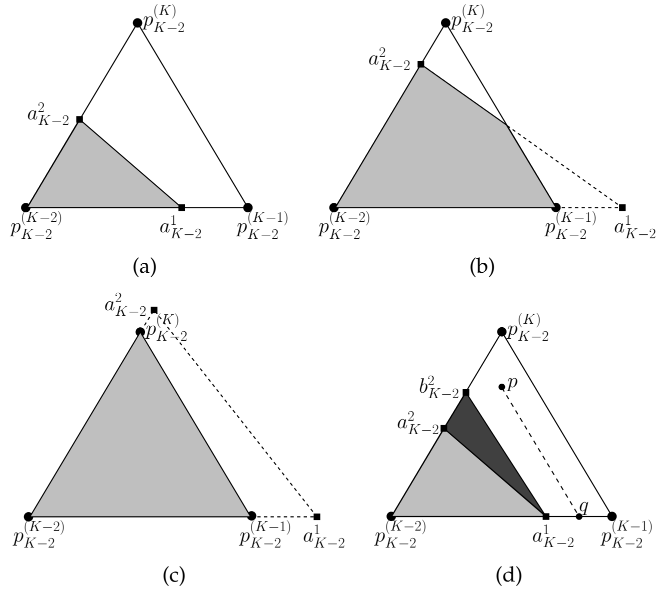


Fig. 4. Illustration of the inner and outer bounds. In all the above figures, only the face of the simplex, $\mathcal{P}_{K-2}$ (in Fig. 3(a)) is shown. The shaded regions in Fig. (a), (b), and (c) are the inner bound when (a) δ_1 and δ_2 are both less than 1 (b) $\delta_1>1$ and $\delta_2<1$ and (c) $\delta_1>1$ and $\delta_2>1$, respectively. The outer bound is the union of the light and the dark shaded regions in Fig. (d).

Theorem 1 (Inner bound): For $k=1,2,\cdots,K-1$, Recalling that $p_k^{(k)}$ is the pmf in $\mathcal{P}_k$ with point mass on k, define

$$\underline{\mathcal{C}}_k(w,b) := \mathcal{P}_k \cap conv\Big\{p_k^{(k)}, a_k^1,\cdots,a_k^{K-k}\Big\},$$

where *conv* denotes the convex hull of the given points. Then $\underline{\mathcal{C}}_k(w,b)\subseteq\mathcal{C}_k(w,b)$.

Proof: See our technical report [19, Theorem 1]. □

In Fig. 4, for stage $k=K-2$, we illustrate the various cases that can arise. In each of the figures the shaded region is the inner bound. In Fig. 4(a) all the thresholds lie within the simplex and simply the convex hull of these points gives the inner bound. When some or all the thresholds lie outside the simplex, as in Fig. 4(b) and 4(c), then the inner bound is obtained by intersecting the convex hull of the thresholds with the simplex. In Fig. 4(c), where all the thresholds lie outside the simplex, the inner bound is the entire simplex, $\mathcal{P}_{K-2}$, so that at stage $K-2$ with $(W_{K-2},B_{K-2})=(w,b)$ it is optimal to stop for any belief state.

5.3 Outer Bound on the Optimum Stopping Set

In this section we will obtain an outer bound (a superset) for the optimum stopping set. Again, as in the case of the inner bound, we will identify certain threshold points whose convex hull will contain the optimum stopping set. This will require us to first prove a monotonicity result which compares the cost of continuing at two belief states $p, q \in \mathcal{P}_k$ which are ordered, for instance for $k = K-2$, as in Fig 4(d). q in Fig. 4(d) is such that $q(K-2) = p(K-2)$ (i.e., the probability that there is no-more relays to go is same in both p and q) and $q(K-1) = 1-p(K-2)$ (i.e., all the remaining probability in q is on the event that there is one-more relay to go, while in p it can be on one-more or two-more relays to go). Thus q lies on the lower edge of the simplex. We will show that the cost of continuing at p is less than that at q.

Lemma 6: Given $p \in \mathcal{P}_k$ for $k = 1, 2, \cdots, K-1$, define $q(k) = p(k)$ and $q(k+1) = 1 - p(k)$, then $c_k(p, w, b) \leq c_k(q, w, b)$ for any (w, b).

Proof: See [19, Appendix II.B]. □

Discussion of Lemma 6: This lemma proves the intuitive result that the continuing cost with a pmf p that gives mass on a larger number of relays should be smaller than with a pmf q that concentrates all such mass in p on just one more relay to go. With more relays, the cost of continuing is expected to decrease. □

Similar to the thresholds a_k^ℓ we define the thresholds b_k^ℓ that lie along certain edges of the simplex. We will identify the threshold a_k^ℓ that is at a maximum distance from the corner point $p_k^{(k)}$ (in Fig. 4(d), this point is $a_{K-2}^1 = [1-\delta_1, \delta_1, 0]$). Next we define the thresholds b_k^ℓ to be the points on the edges emanating from $p_k^{(k)}$, which are at this same distance. Thus in Fig. 4(d), $b_{K-2}^1 = a_{K-2}^1$ and $b_{K-2}^2 = [1-\delta_1, 0, \delta_1]$.

Definition 6: For a given $k \in \{1, 2, \cdots, K-1\}$, let $\ell_{max} = \arg\max_{\ell=1,2,\cdots,K-k} \delta_\ell$. Now for $\ell = 1, 2, \cdots, K-k$ define b_k^ℓ as a $K-k+1$ dimensional point with the first and the $\ell+1$ th components equal to $1-\delta_{\ell_{max}}$ and $\delta_{\ell_{max}}$ respectively, the rest of the components are zeros. Each of the b_k^ℓ are at equal distance from $p_k^{(k)}$ but on a different edge starting from $p_k^{(k)}$. □

Using Lemma 6, we show that the convex hull of the thresholds b_k^l along with the corner point $p_k^{(k)}$ constitutes an outer bound for the optimum stopping set. The idea of the proof can be illustrated using Fig. 4(d). p in Fig. 4(d) is outside the convex hull and q is obtained from p as in Lemma 6. At q it is optimal to continue since it is beyond the threshold a_{K-2}^1 and hence the continuing cost at q, $c_k(q, w, b)$, is less than the stopping cost $-\eta b$. From Lemma 6 it follows that the continuing cost at p, $c_k(p, w, b)$, is also less than $-\eta b$ so that it is optimal to continue at p as well, proving that p does not belong to the optimum stopping set. Thus the convex hull contains the optimum stopping set. We formally state and prove this *outer bound* theorem next.

Theorem 2 (Outer bound): For $k = 1, 2, ..., K-1$, define

$$\overline{\mathcal{C}}_k(w, b) = \mathcal{P}_k \cap conv\Big\{p_k^{(k)}, b_k^1, \cdots, b_k^{K-k}\Big\}.$$

Then $\mathcal{C}_k(w, b) \subseteq \overline{\mathcal{C}}_k(w, b)$.

Proof: See [19, Theorem 2]. □

The outer bound for $k = K-2$ is illustrated in Fig. 4(d). The light shaded region is the inner bound. The outer bound is the union of the light and the dark shaded regions. The boundary of the optimum stopping set falls within the dark shaded region. For any p within the inner bound we know that it is optimal to stop and for any p outside the outer bound it is optimal to continue. We are uncertain about the optimal action for belief states within the dark shaded region.

6 Optimum Relay Selection in a Simplified Model

The bounds obtained in the previous section require us to compute the thresholds $\{\phi_\ell : \ell = 0, 1, \cdots, K-1\}$ (see Definition 3) recursively. These are computationally very intensive to obtain. Hence, in this section we simplify the exact model and extract a simple selection rule. Our aim is to apply this simple rule to the exact model and compare its performance with the other policies.

6.1 The Simplified Model

Now we describe our *simplified model*. There are $\tilde{N}$ relays. Here, $\tilde{N}$ is a constant and is known to the source. The key simplification in this model is that here the relay nodes wake-up *at the first $\tilde{N}$ points of a Poisson process of rate $\frac{\tilde{N}}{T}$*. The following are the motivations for considering such a simplification. Note that in our actual model (Section 2), when $N = \tilde{N}$, the inter wake-up times $\{U_k : 1 \leq k \leq \tilde{N}\}$ are identically distributed [26, Chapter 2], but not independent. Their common cumulative distribution function (cdf) is $F_{U_k|N}(u|\tilde{N}) = 1 - (1 - \frac{u}{T})^{\tilde{N}}$ for $u \in (0, T)$. From Fig. 5 we observe that the cdf of

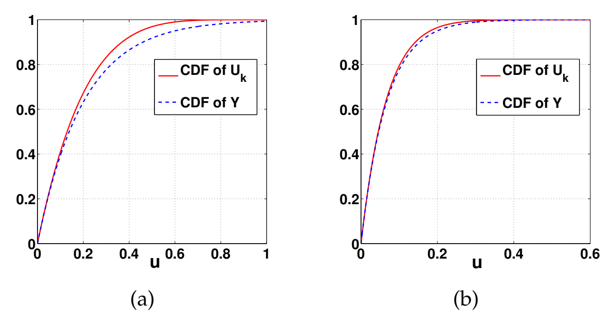


Fig. 5. The CDFs $F_{U_k|N}(.|\tilde{N})$ and $F_Y(.)$ where $Y \sim exp(\frac{\tilde{N}}{T})$ are plotted for (a) $\tilde{N} = 5$ and (b) $\tilde{N} = 15$.

$\{U_k : 1 \leq k \leq \tilde{N}\}$ is close to that of an exponential random variable of parameter $\frac{\tilde{N}}{T}$ and the approximation becomes better for large values of $\tilde{N}$ (for a fixed T). This motivates us to approximate the actual inter wake-up times by exponential random variable of rate $\frac{\tilde{N}}{T}$. Further in the simplified model we allow the inter wake-up times to be independent. Finally, observe that in the simplified model the average number of relays that wake-up within the duty cycle T is $\tilde{N}$ which is same as that in the exact model when $N = \tilde{N}$.

We will use the notations such as $\tilde{W}_k, \tilde{R}_k, \tilde{U}_k$, etc., to represent the analogous quantities that were defined for the exact model. For instance, $\tilde{W}_k$ represents the wake-up time of the k th relay. However, unlike in the exact model, here $\tilde{W}_k$ can be beyond T. As mentioned before, $\{\tilde{U}_k : k = 1, \cdots, \tilde{N}\}$ are simply iid exponential random variables with parameter $\frac{\tilde{N}}{T}$. $\{\tilde{R}_k : k = 1, \cdots, \tilde{N}\}$ are iid random rewards with common pdf f_R which is same as that in the exact model.

6.2 Optimal Policy for the Simplified Model

Again, here the decision instants are the times at which the relays wake-up. At some stage k, $1 \le k < \tilde{N}$, suppose $(\tilde{W}_k, \tilde{R}_k) = (w, b)$ then the one step cost of stopping is $-\eta b$ and that of continuing is $\tilde{U}_{k+1}$. Note that since $\tilde{U}_{k+1} \sim Exp(\frac{\tilde{N}}{T})$, the one step costs do not depend on w, which means that the optimal policy for the simplified model does not depend on the value of w. Also since the number of relays $\tilde{N}$ is a contant, we do not wish to retain it as a part of the state. Therefore we simplify the state space to be $\tilde{\mathcal{S}}_0 = \{0\}$ and for $k = 1, 2, \cdots, \tilde{N}$, $\tilde{\mathcal{S}}_k = [0, \overline{R}] \cup \{\psi\}$. As before ψ is the terminating state. In this section we will prove that the *one-step-look-ahead rule* is optimal for the simplified model. The idea is to show that the *one-step-stopping set is absorbing* [21, Section 4.4]. All these will now be defined.

At stage k, $1 \le k < \tilde{N}$, when the state is b, the cost of stopping is simply $c_s(b) = -\eta b$. The cost of continuing for one more step (which is $\tilde{U}_{k+1}$) and then stopping at the next stage (where the state is $\max\{b, \tilde{R}_{k+1}\}$) is,

$$\begin{aligned} c_c(b) &= \mathbb{E}\Big[\tilde{U}_{k+1} - \eta \max\{b, \tilde{R}_{k+1}\}\Big] \\ &= -\eta\Big(\mathbb{E}[\max\{b, R\}] - \frac{T}{\eta\tilde{N}}\Big) \end{aligned}$$

By defining the function $\beta(\cdot)$ for $b \in [0, \overline{R}]$ as

$$\beta(b) = \mathbb{E}[\max\{b, R\}] - \frac{T}{\eta\tilde{N}}, \qquad (15)$$

we can write $c_c(b) = -\eta\beta(b)$. Note that both the costs, c_s and c_c, do not depend on the stage index k.

Definition 7: We define the *One-step-stopping set* as,

$$\mathcal{C}_{1step} = \Big\{b \in [0, \overline{R}] : -\eta b \le -\eta\beta(b)\Big\}. \qquad (16)$$

i.e., it is the set of all states $b \in [0, \overline{R}]$ where the cost of stopping, $c_s(b)$, is less than the cost of continuing for one more step and then stopping at the next stage $c_c(b)$. □

We will show that $\mathcal{C}_{1step}$ is characterized by a threshold α and can be written as $\mathcal{C}_{1step} = [\alpha, \overline{R}]$. This will require the following properties about $\beta(\cdot)$.

Lemma 7:

1) β is continuous, increasing and convex in b.
2) If $\beta(0) < 0$, then $\beta(b) < b$ for all $b \in [0, \overline{R}]$.
3) If $\beta(0) \ge 0$, then $\exists$ a unique α such that $\alpha = \beta(\alpha)$.
4) If $\beta(0) \ge 0$, then $\beta(b) < b$ for $b \in (\alpha, \overline{R}]$ and $\beta(b) > b$ for $b \in [0, \alpha)$.

Proof: See our technical report [19, Appendix III.A]. □

Discussion of Lemma 7: When $\beta(0) \ge 0$ then using Lemma 7.3 and 7.4, we can write $\mathcal{C}_{1step}$ in (16) as $\mathcal{C}_{1step} = [\alpha, \overline{R}]$. For the other case where $\beta(0) < 0$, from Lemma 7.2 it follows that $\mathcal{C}_{1step} = [0, \overline{R}]$. Thus by defining $\alpha = 0$ whenever $\beta(0) < 0$ we can write $\mathcal{C}_{1step} = [\alpha, \overline{R}]$ for either case. □

Definition 8: Depending on the value of $\beta(0)$ define α as follows: $\alpha = \beta(\alpha)$ if $\beta(0) \ge 0$. Otherwise fix $\alpha = 0$. □

Definition 9: A policy is said to be *one-step-look-ahead* if at stage k, $1 \le k < \tilde{N}$, it stops iff the $b \in \mathcal{C}_{1step}$, i.e., iff the cost of stopping, $c_s(b)$, is less than the cost of continuing for one more step and then stopping, $c_c(b)$. □

Definition 10: Let $\mathcal{C}$ be some subset of the state space $[0, \overline{R}]$, i.e., $\mathcal{C} \subseteq [0, \overline{R}]$. We say that $\mathcal{C}$ is *absorbing* if for every $b \in \mathcal{C}$, if the action at stage k, $1 \le k < \tilde{N}$, is to continue, then the next state, s_{k+1} at stage $k+1$, also falls into $\mathcal{C}$. □

Since we have expressed $\mathcal{C}_{1step}$ as $[\alpha, \overline{R}]$ and since $s_{k+1} = \max\{b, \tilde{R}_{k+1}\}$ it is clear that $\mathcal{C}_{1step}$ is absorbing. Finally, referring to [21, Section 4.4], it follows that, for optimal stopping problems, *whenever the one-step-stopping set is absorbing then the one-step-look-ahead rule is optimal*. Thus the optimal policy for the simplified model is to choose the first relay whose reward is more than α. If none of the relays' reward values are more than α then at the last stage choose the one with the maximum reward.

7 Numerical and Simulation Results

For simulations we have considered the context of geographical forwarding in a dense sensor network with sleep-wake cycling nodes, which is the primary motivation for this work. We will first perform simulations to compare the *one-hop performance* of the various policies obtained from the analysis in the previous sections. Next, after observing the good performance of A-SIMPL, we apply this policy to route a packet in a large network and study its *end-to-end performance*.

7.1 One-Hop Performance

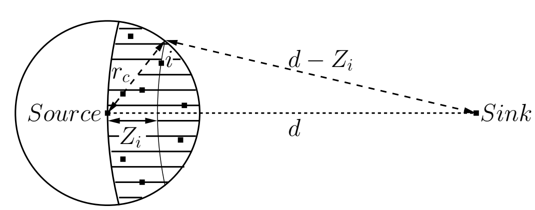
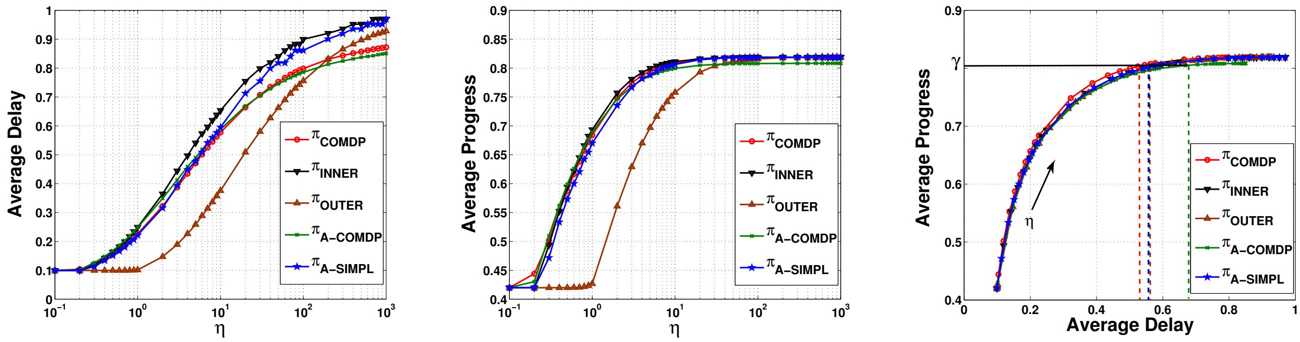


Fig. 6. Model used for the one-hop simulations.



(a) (b) (c)

Fig. 7. (a) Average delay obtained by various policies as a function of η (b) Average progress as a function of η (c) Average progress vs. average delay.

The source and sink are separated by a distance of $d = 10$ units (see Fig. 6). The source has a packet to forward at time 0. The communication radius of the source is $r_c = 1$. The potential relay nodes are the neighbors of the source that are closer to the sink than itself. The period of sleep-wake cycling is $T = 1$. Let Z_i represent the progress of relay i. Z_i is the difference between the source-sink and relay-sink distances. The reward associated with a relay i is simply the progress made by it, i.e., $R_i = Z_i$. We interchangeably use progress and reward in this section.

Each of the nodes is located uniformly in the forwarding set, independently of the other nodes. Therefore the progress made by them are iid with pdf

$$f_Z(r) = \frac{2(d-r)cos^{-1}\left(\frac{d^2+(d-r)^2-{r_c}^2}{2d(d-r)}\right)}{\text{Area of the forwarding region}}, \quad (17)$$

and the support of f_Z is $[0, r_c]$. Hence r_c is analogous to $\overline{R}$ in our model in Section 2. We take the bound on the number of relays as $K = 50$, and the initial pmf is taken as truncated Poisson with parameter 10, i.e., for $n = 1, 2, \cdots, K$, $p_0(n) = c\frac{10^n}{n!}e^{-10}$ where c is the normalization constant.

Implemented Policies (one-hop): Following is the description of the various policies that we have implemented:

1) π_{COMDP}: For this policy, the source knows the actual value of N. Suppose $N = n$, then the source begins with an initial belief with mass only on n. At any stage, $k = 1, 2, \cdots, n$, if the delay and best reward pair is (w, b) then transmit if $b \geq \phi_{n-k}(w, b)$, continue otherwise. See the remark following Lemma 3. This policy *serves as a lower bound for the cost* achieved by other policies.

2) π_{INNER}: We use the inner bound $\underline{\mathcal{C}}_k(w, b)$ to obtain a suboptimal policy. At stage k if the belief state is (p, w, b) $(\in \mathcal{S}_k)$, then transmit iff $p \in \underline{\mathcal{C}}_k(w, b)$.

3) π_{OUTER}: We use the outer bound $\overline{\mathcal{C}}_k(w, b)$ to obtain a suboptimal policy. At stage k if the belief state is (p, w, b) $(\in \mathcal{S}_k)$, then transmit iff $p \in \overline{\mathcal{C}}_k(w, b)$.

4) $\pi_{A-COMDP}$: (Average-COMDP) The source assumes that N is equal to its average value $\overline{N} = \lceil \mathbb{E}N \rceil$ ($\lceil x \rceil$ represents the smallest integer greater than x), and begins with an initial pmf with mass only on $\overline{N}$. Suppose $N = n$ (which the source does not know), then at some stage $k = 1, 2, \cdots, \min\{n, \overline{N}\}$ if the delay and best reward pair is (w, b) then transmit iff $b \geq \phi_{\overline{N}-k}(w, b)$. In the case when $\overline{N} > n$, if the source has not transmitted until stage n and further at stage n if the action is to continue, then since there are no more relays to go, the source ends up waiting until time T and then forwards to the node with the best reward.

5) $\pi_{A-SIMPL}$: (Average-Simple) This policy is derived from the simplified model described in Section 6. The source considers the simplified model assuming that there are $\overline{N} = \lceil \mathbb{E}N \rceil$ number of relays. It computes the threshold α accordingly, using Definition 8. The policy is to transmit to the first relay that wakes up and offers a reward (progress in this case) of more than α. If there is no such relay then the source ends up waiting until time T, and then transmits to the node with the best reward.

In Fig. 7(a), we plot the average delays of the policies described above as a function of η. The average reward is plotted in Fig. 7(b).

Discussion: As a function of η both, the average delay and the average reward are increasing. This is because for larger η we value the progress more so that we tend to wait for longer time to do better in progress. For very small values of η, all the thresholds ($\{\phi_\ell\}$ and α) are very small and most of the time, the packet is forwarded to the first node. For very high values of η the policies end up waiting for all the relays and then choose the one with the best reward. Therefore, as η increases, the average progress of all the policies (excluding $\pi_{A-COMDP}$) converge to $\mathbb{E}[\max\{Z_1, \cdots, Z_N\}]$ which is about 0.82 (see Fig. 7(b)). However the average progress for $\pi_{A-COMDP}$ converges to a value less than 0.82. This is because whenever $\overline{N} < N$ and for large η (where all the thresholds $\{\phi_\ell\}$ are large) $\pi_{A-COMDP}$ ends up waiting for the first $\overline{N}$ relays and obtains a progress of $\max\{Z_1, \cdots, Z_{\overline{N}}\}$ which is less than (or equal to) the progress made by the other policies (which is $\max\{Z_1, \cdots, Z_N\}$).

Recall that the main problem we are interested in is the one in (1). We should be comparing the average delay obtained by the above policies when the average reward provided by each of them is equal to γ. To illustrate the solution, in Fig. 7(c) we first plot the average progress (y-axis) obtained by the various policies as a function of the average delay (x-axis). Then, for a given γ, the projection of the point of intersection of the horizontal line of height γ with each of the curves onto the x-axis, gives the average delay obtained by the respective policies. In Fig. 7(c) for a particular choice of γ we have depicted such projections. As expected π_{COMDP}, being optimal, obtains minimum delay for any value of γ. The naive policy, $\pi_{A-COMDP}$, does not perform well, in the sense that it obtains the maximum delay among all the policies. The delay values of the remaining policies (namely π_{INNER}, π_{OUTER} and $\pi_{A-SIMPL}$) are only slightly higher than that of π_{COMDP}. A closer inspection of the curves reveal that the performance of π_{INNER} and $\pi_{A-SIMPL}$ are very close to each other and is slightly better than that of π_{OUTER} (this not visible from Fig. 7(c), we had to zoom into the curves to conclude this result).

These observations are for the particular case where the reward is simply the progress and the initial belief is truncated Poisson. In our technical report [19, Appendix IV] we have shown simulation results for other reward structures and initial beliefs. We observe the good performance of $\pi_{A-SIMPL}$ there as well.

7.2 End-to-End Performance

Example-1: Hop Count as the Total Cost

We have already observed that, the one-hop performance of $\pi_{A-SIMPL}$ is close to that of the optimal policy π_{COMDP}. Further, the implementation of $\pi_{A-SIMPL}$ requires only a single threshold α in contrast to the sequence of threshold functions $\phi_\ell(w,b)$ (for each $(w,b) \in [0,T]\times[0,r_c]$ and $\ell = 1,2,\cdots,K$), required to implement the other policies. These features motivates us to study the end-to-end performance (i.e., average total delay and average hop count) obtained by heuristically applying the policy A-SIMPL at each hop enroute to the sink in a large network.

We will also make a comparison with the work of Kim et al. [1], where the problem of routing a packet in a network with duty cycling nodes is considered as a stochastic shortest path problem. The authors in [1] have developed a distributed Bellman-Ford algorithm (referred to as the LOCAL-OPT algorithm) to minimize the average total delay[3]. The LOCAL-OPT algorithm yields, for each neighbor j of node i, an integer threshold $h_j^{(i)}$ such that if j wakes up and listens to the h-th *beacon signal*[4] from node i and if $h \le h_j^{(i)}$, then j will send an ACK to receive the packet from i. Otherwise (if $h > h_j^{(i)}$) j will go back to sleep. In fact, here we have adopted the work of Kim et al. to minimize

$$\text{Avg. Total Delay} + \lambda \text{ Avg. Hop Count} \tag{18}$$

where $\lambda \ge 0$ is the trade-off parameter. λ could be thought of as the power required for each transmission in which case the hop count is proportional to the energy expended by the network.

Implemented Policies (end-to-end): We have fixed a network comprising 500 nodes located uniformly in the region $[0,L]^2$ where $L = 10$ (thus the node density is 5). An additional sink node is placed at the location $(0,L)$ as shown in Fig. 1. As before we fix the radius of communication of each node to be $r_c = 1$. Events occur at random locations within $[0,L]^2$. Each time an event occurs, a node nearest to its location generates an alarm packet which needs to be forwarded to the sink, possibly through multiple hops. We will refer to the time at which the packet is generated as time 0. Now the wake-up times of the nodes are sampled independently and randomly from $[0,T]$, where $T = 1$ is the period of the sleep-wake cycle (for each event we generate fresh samples for the wake-up times). Thus a node i wakes-up at the periodic instances $T_i, T+T_i, 2T+T_i, \cdots$, where T_i is uniform on $[0,T]$. At each wake-up instant, node i listens for a beacon signal, if any, before going back to sleep. The duration of the beacon signal is $t_I = 5$ msec. Thus a forwarding node has to send at most $\frac{T}{t_I} = 200$ beacon signals before all its neighbors wake-up. Description of the forwarding policies that we have implemented is given below.

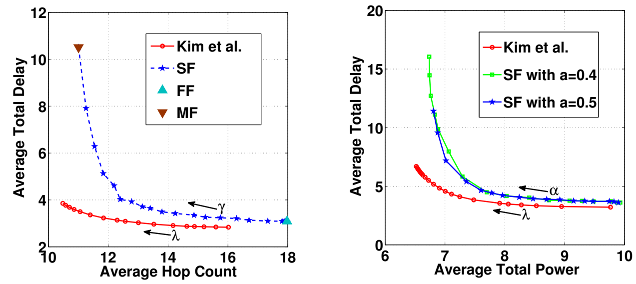
1) FF (First Forward): Forward to the first node that wakes up within the forwarding region irrespective of the progress it makes towards the sink.

2) MF (Max Forward): Wait of the entire duration T and then choose a neighbor with the maximum progress.

3) SF (Simple Forward): Obtained by applying the one-hop policy, $\pi_{A-SIMPL}$, at each hop enroute to the sink. First, knowing the node density each node i computes the average number of neighbors, $\overline{N}_i$, within its forwarding region. Then, for a given γ, node i computes a threshold α_i such that the average progress obtained (by assuming the simplified model with $\overline{N}_i$ relays) using α_i is γ. Now, when a node j wakes up and if it hears a beacon signal from i, it waits for the ID signal and then sends an ACK signal containing its location information. If the progress made by j is more than the threshold α_i, then i forwards the packet to j. If the progress made by j is less than the threshold, then i asks j to stay awake if its progress is the maximum among all the nodes that have woken up thus far, otherwise i asks j to return to sleep. If more than one node wakes up during the same beacon signal, then contentions are resolved by selecting the one which makes the most progress among them. In the simulation, this happens instantly (as also for the Kim et al. algorithm that we compare with); in practice this will require a splitting algorithm; see, for example, [27, Chapter 4.3]. We assume that within $t_I = 5$

3. Total delay is the sum of the waiting times (i.e., one-hop delays) incurred at each hop due to the sleep-wake process.

4. The beacon signal contains the node ID of the forwarding node and other control data. A node i with a packet to forward continuously transmits these beacon signals until it gets an ACK from a neighbor



(a) (b)

Fig. 8. End-to-end performance trade-off curves obtained for, (a) Example-1, where total cost is the hop count, and (b) Example-2, where total cost is the total power.

msec all these transactions (beacon signal, ID, ACK and contention resolution if any) are over. FF and MF can be thought of as special cases of A-SIMPL with thresholds of 0 and 1 respectively.

4) Kim et al.: For a given λ, we run the LOCAL-OPT algorithm on the network and obtain the values $h_j^{(i)}$ for each neighbor pair (i,j). We use these thresholds to route from source to sink. Contentions, if any, are resolved (instantly, in the simulation) by selecting a node j with the highest $h_j^{(i)}$ index.

In Fig. 8(a) we plot average total delay vs. average hop count obtained by the policies described above. The average is over the random event location and sleep-wake cycling. For the policy SF, by varying γ we obtain its performance curve, the two extreme points of which is the performance of the policies FF and MF. Similarly by varying λ we obtain the performance curve for the Kim et al. policy.

Discussion: As expected, the Kim et al. policy, being optimal, performs better. However, interestingly we observe that for any target hop count (along x-axis) within $[12.5, 16]$ the difference in corresponding delays obtained by SF and by Kim et al. is less than one duty cycle $T = 1$. However the difference in delay grows rapidly for target hop count less than 11.5. This is because, lower values of target hop count corresponds to the SF policy being operated at larger values of γ, resulting in each node i using a larger threshold α_i. Thus, there will be very few nodes (or none) whose progress in more than α_i so that the packet ends up waiting for close to entire duty cycle at each hop.

A major drawback with the Kim et al. policy is that a *pre-configuration* phase, involving a lot of control packet exchanges, is required to run the LOCAL-OPT algorithm. In contrast the SF policy can be implemented immediately after deployment. All it requires is for each node to know the node density, its location and the location of the sink (which are the prerequisites for geographical forwarding). Then, for a given γ a forwarding node can compute the threshold α it needs to use. An interesting approach would be to allow the source node to set the value of γ depending on the "type" of the event detected. For *delay sensitive* events it is appropriate to use a smaller value of γ so that the delay is small, whereas, for energy constrained applications (where the network energy needs to conserved) it is better to use large γ so that the number of hops (and hence the number of transmissions) is reduced. For other applications, moderate values of γ can be used. γ can be a part of the beacon signal so that it is made available to the next hop relay.

Example-2: Power as the Total Cost

We have also performed end-to-end simulations by imposing a constraint on the average total power required to route an alarm packet. In this case, analogous to (18), we attempt to minimize

$$\text{Avg. Total Delay} + \lambda \text{ Avg. Total Power.} \qquad (19)$$

We have assumed a model where the one-hop power required by the forwarding node to forward the alarm packet to relay i at a distance D_i from it is $P_i = P_{min} + \Gamma D_i^{\beta}$, where β is the path loss attenuation factor usually in the range 2 to 5 and $\Gamma > 0$ is a constant containing the noise variance and the SNR (signal to noise ratio) threshold beyond which decoding is sucessful. In our simulations we have fixed $P_{min} = 0.1$ and $\Gamma = 1$. Now the total power is simply the sum of all one-hop powers. For the local problem we define the reward associated with a relay i as $R_i = \frac{Z_i^a}{P_i^{(1-a)}}$ where $a \in [0,1]$ is used to trade-off between the progress and the one-hop power. Z_i in the reward expression is essential to give a sense of direction, towards the sink, to the packet.

In Fig. 8(b) we have plotted the end-to-end performance trade-off curves obtained by the simple policy SF (for two different values of the parameter a namely, $a = 0.4$ and $a = 0.5$) and that obtained by Kim et al. This time, to further ease the implementation of SF, we allow all the nodes to use the same threshold α. Again, as in Example-1 (see Fig. 8(a)), we observe that the SF policy (for both the values of a) performs well for target total power in the range $[7.5, 9.75]$ and worsens for target total power less than 7. We have performed simulations for few other values of a as well (plots are not shown). However, we found that the performance of SF for these values of a are not as good as for $a = 0.4$ and $a = 0.5$, thus suggesting that these values of a best capture the tradeoff between progress and power in the "reward" expression.

8 Conclusion

Our work in this paper was motivated by the problem of geographical forwarding of packets in a wireless sensor networks whose function is to detect certain infrequent events and forward these alarms to a base station, and whose nodes are sleep-wake cycling to conserve energy. This end-to-end problem motivated the local problem faced by a packet forwarding node, i.e., that of choosing one among a set of potential relays, so as to minimize the average delay in selecting a relay subject to a constraint on the average progress (or some reward, in general). Further the source does not know the number of available relays. We formulated the problem as a finite horizon POMDP and characterized the optimal policy in terms of optimum stopping sets. We proved inner and outer bounds for this set (Theorem 1 and Theorem 2, respectively). We also obtained a simple threshold rule by formulating an alternate simplified model (Section 6). We performed one-hop simulations and observed the good performance of the simple policy ($\pi_{A-SIMPL}$). Finally, we applied the policy $\pi_{A-SIMPL}$ to route an alarm packet in a large network and observed that its performance, over some range of target hop count (or total cost), is comparable to that of a distributed Bellman-Ford algorithm proposed by Kim et al.

Acknowledgements

This research was supported in part by the Indo-French Center for the Promotion of Advanced Research (IFCPAR Project 4000-IT-1), and in part by the following funding agencies of the Government of India: Defence Research and Development Organisation (DRDO), and Department of Science and Technology (DST).

References

[1] J. Kim, X. Lin, and N. Shroff, "Optimal Anycast Technique for Delay-Sensitive Energy-Constrained Asynchronous Sensor Networks," in *INFOCOM 2009. The 28th Conference on Computer Communications. IEEE*, April 2009, pp. 612–620.

[2] Q. Cao, T. Abdelzaher, T. He, and J. Stankovic, "Towards Optimal Sleep Scheduling in Sensor Networks for Rare-Event Detection," in *IPSN '05: Proceedings of the 4th international symposium on Information processing in sensor networks*, 2005.

[3] K. Premkumar, A. Kumar, and J. Kuri, "Distributed Detection and Localization of Events in Large Ad Hoc Wireless Sensor Networks," in *Communication, Control, and Computing, 2009. Allerton 2009. 47th Annual Allerton Conference on*, sept. 2009, pp. 178 –185.

[4] K. P. Naveen and A. Kumar, "Tunable Locally-Optimal Geographical Forwarding in Wireless Sensor Networks with Sleep-Wake Cycling Nodes," in *INFOCOM 2010. The 29th Conference on Computer Communications. IEEE*, march 2010, pp. 1–9.

[5] K. Akkaya and M. Younis, "A Survey on Routing Protocols for Wireless Sensor Networks," *Ad Hoc Networks*, vol. 3, no. 3, pp. 325 – 349, 2005.

[6] M. Mauve, J. Widmer, and H. Hartenstein, "A Survey on Position-Based Routing in Mobile Ad-Hoc Networks," *IEEE Network*, vol. 15, pp. 30–39, 2001.

[7] J. H. Chang and L. Tassiulas, "Maximum Lifetime Routing in Wireless Sensor Networks," *IEEE/ACM Trans. Netw.*, vol. 12, no. 4, pp. 609–619, 2004.

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

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

[10] S. Liu, K. W. Fan, and P. Sinha, "CMAC: An Energy Efficient MAC Layer Protocol using Convergent Packet Forwarding for Wireless Sensor Networks," in *Sensor, Mesh and Ad Hoc Communications and Networks, 2007. SECON '07. 4th Annual IEEE Communications Society Conference on*, June 2007, pp. 11–20.

[11] V. Paruchuri, S. Basavaraju, A. Durresi, R. Kannan, and S. S. Iyengar, "Random Asynchronous Wakeup Protocol for Sensor Networks," *International Conference on Broadband Networks*, vol. 0, pp. 710–717, 2004.

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

[13] N. B. Chang and M. Liu, "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*, 2007, pp. 27–38.

[14] M. Sakaguchi, "Dynamic Programming of Some Sequential Sampling Design," *Journal of Mathematical Analysis and Applications*, vol. 2, no. 3, pp. 446 – 466, 1961.

[15] S. Karlin, *Stochastic Models and Optimal Policy for Selling an Asset, Studies in Applied Probability and Management Science / edited by Kenneth J. Arrow, Samuel Karlin, Herbert Scarf*. Stanford University Press, Stanford, Calif, 1962.

[16] B. K. Kang, "Optimal Stopping Problem with Double Reservation Value Property," *European Journal of Operational Research*, vol. 165, no. 3, pp. 765 – 785, 2005.

[17] I. David and O. Levi, "A New Algorithm for the Multi-item Exponentially Discounted Optimal Selection Problem," *European Journal of Operational Research*, vol. 153, no. 3, pp. 782 – 789, 2004.

[18] M. S. Ee, "Asset-Selling Problem with an Uncertain Deadline, Quitting Offer, and Search Skipping Option," *European Journal of Operational Research*, vol. 198, no. 1, pp. 215 – 222, 2009.

[19] K. P. Naveen and A. Kumar, "Relay Selection with Partial Information in Wireless Sensor Networks," Tech. Rep., 2011. [Online]. Available: http://arxiv.org/abs/1101.3835

[20] G. E. Monahan, "A Survey of Partially Observable Markov Decision Processes: Theory, Models, and Algorithms," *Management Sciences*, vol. 28, no. 1, pp. 1–16, January 1982.

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

[22] W. S. Lovejoy, "A Survey of Algorithmic Methods for Partially Observed Markov Decision Processes," *Ann. Oper. Res.*, vol. 28, no. 1-4, pp. 47–66, 1991.

[23] R. D. Smallwood and E. J. Sondik, "The Optimal Control of Partially Observable Markov Processes Over a Finite Horizon," *Operations Research*, vol. 21, no. 5, pp. 1071–1088, 1973.

[24] J. M. Porta, N. Vlassis, M. T. J. Spaan, and P. Poupart, "Point-Based Value Iteration for Continuous POMDPs," *Journal of Machine Learning Research*, vol. 7, pp. 2329–2367, 2006.

[25] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge University Press, March 2004.

[26] H. A. David and H. N. Nagaraja, *Order Statistics (Wiley Series in Probability and Statistics)*. Wiley-Interscience, August 2003.

[27] D. Bertsekas and R. Gallager, *Data networks*. Upper Saddle River, NJ, USA: Prentice-Hall, Inc., 1992.

K. P. Naveen obtained his B.E. degree from the Visveswaraiah Technological University (VTU), Belgaum in 2005. Subsequently he has worked at the ISRO Satellite Centre, Bangalore. Currently he is a Ph.D student in the Department of ECE, Indian Institute of Science (IISc), Bangalore. His research interests include stochastic control and optimization of wireless networks.

Anurag Kumar (B.Tech., IIT Kanpur, PhD Cornell University, both in EE) was with Bell Labs, Holmdel, for over 6 years. He is now a Professor in the ECE Department at the Indian Institute of Science (IISc), Bangalore, and Chair of the Division of Electrical Sciences. His area of research is communication networking, and he has recently focused primarily on wireless networking. He is a Fellow of the IEEE, the Indian National Science Academy (INSA), the Indian National Academy of Engineering (INAE), and Indian Academy of Sciences (IASc). He was an associate editor of IEEE Transactions on Networking, and of IEEE Communications Surveys and Tutorials. He is a coauthor of the graduate text-books "Communication Networking: An Analytical Approach", and "Wireless Networking", by Kumar, Manjunath and Kuri, published by Morgan-Kaufman/Elsevier.