Competitive Selection of Ephemeral Relays in Wireless Networks

K. P. Naveen, Eitan Altman, *Fellow, IEEE,* and Anurag Kumar, *Fellow, IEEE*

***Abstract*—We consider an opportunistic wireless communication setting, in which two nodes (referred to as forwarders) compete to choose a relay node from a set of relays, as they ephemerally become available (e.g., wake up from a sleep state). Each relay, when it becomes available (or *arrives*), offers a (possibly different) "reward" to each forwarder. Each forwarder's objective is to minimize a combination of the delay incurred in choosing a relay and the reward offered by the chosen relay. As an example, we develop the reward structure for the specific problem of *geographical forwarding* over a common set of sleep-wake cycling relays. In general, our model can be considered as a *game theoretic variant of the asset selling problem* studied in the operations research literature.**

We study two variants of the generic relay selection problem, namely, the completely observable (CO) and the partially observable (PO) cases. These cases are based on whether a forwarder (in addition to observing its reward) can also observe the reward offered to the other forwarder. Formulating both problems as a two person stochastic game, we characterize the solutions in terms of Nash Equilibrium Policy Pairs (NEPPs). For the CO case we provide a general structure of the NEPPs. For the PO case we prove that there exists an NEPP within the class of threshold policy pairs.

Through numerical work, for a one-hop forwarding example we compare the cost performance of various NEPPs with a simple forwarding (SF) policy which causes each forwarder to act as if the other is not present. We find that if the forwarders are not very close then the SF policy suffices. Insights gained from this numerical work are then used in an end-to-end simulation of geographical forwarding in a large network, in which we are concerned with delivery of packets from a tagged source to a sink, in the presence of competition from other packet flows destined to the same sink.

***Index Terms*—Competitive relay selection, geographical forwarding, stochastic games, Bayesian games.**

I. INTRODUCTION

We are concerned in this paper with a class of resource allocation problems in wireless networks, in which competing nodes need to acquire a resource, such as a physical radio relay (see the geographical forwarding example later in this paper) or a channel (as in a cognitive radio network [1]), when a sequence of such resources "arrive" (i.e., become available) over time, and stay available only fleetingly for acquisition. Each resource upon arrival offers a "reward" that reflects the goodness of the resource. In this paper, formulating such a problem for two nodes as a stochastic game, we characterize the solution in terms of Nash Equilibrium Policy Pairs (NEPP). We provide numerical results, and insights therefrom, for a specific reward structure derived from the problem of *geographical forwarding in sleep-wake cycling wireless networks*.

A. Geographical Forwarding Context

With the increasing importance of "smart" utilization of our limited resources (e.g., energy and clean water) there is a need for instrumenting our buildings and campuses with wireless sensor networks, towards achieving the vision of Internet of Things (IoT) [2], [3]. As awareness grows and sensing technologies emerge, new IoT applications will be implemented. While each application will require different sensors and backend analytics, the availability of a common wireless network infrastructure will promote the quick deployment of new applications. One approach for building such an infrastructure, say, in a large building setting, would be to deploy a large number of relay nodes, and employ the idea of *geographical forwarding*. If the phenomena to be monitored are slowly varying over time, the traffic on the network can be assumed to be light. In addition, such applications are *delay tolerant*, thus accommodating the approach of *opportunistic* geographical forwarding over *sleep-wake cycling* wireless networks [4], [5].

Sleep-wake cycling is an approach whereby, to conserve the relay battery power, the relay radios are kept turned OFF, while coming ON periodically to provide opportunities for packet forwarding. The problem of forwarding in such a setting was explored in [4], [5], where the formulation was limited to a *single packet* flowing through the network. However, when a common infrastructure of wireless network is being shared by several (competing) applications, it is possible that somewhere in the network more than one forwarding node, holding packets corresponding to different applications, are simultaneously seeking a next-hop relay from an overlapping set of potential relays. In this paper, we are interested in studying the problem of competitive relay selection that arises in the above context.

Formally, the geographical forwarding example we consider in this paper is the following. There are two forwarders[1]

Author's addresses: K. P. Naveen, Department of Electrical Engineering, Indian Institute of Technology Madras, Chennai 600036, India (Email: naveenkp@ee.iitm.ac.in); Eitan Altman, INRIA, Sophia Antipolis 06902, France (Email: Eitan.Altman@inria.fr); Anurag Kumar, Department of Electrical Communication Engineering, Indian Institute of Science, Bangalore 560012, India (Email: anurag@ece.iisc.ernet.in).

This work is supported by the Department of Science and Technology (DST) through an INSPIRE faculty award, a J.C. Bose National Fellowship, and project WINSON (Wireless Networks and Techniques with Applications to Social Needs); by the Indo-French Center for the Promotion of Advanced Research under Project 4000IT1 and IFCAM (Indo-French Centre for Applied Mathematics); and the INRIA Associates Project GANESH (Games, Optimization and Analysis of Networks: Theory and Applications).

[1] From an end-to-end simulation experiment (see Section VI-B for details) we have observed that, whenever competition occurs, more than 96% of the time only two forwarders are involved (when the traffic rate in the network is less than 20 packets/second); see Fig. 4(b). Thus, there is a practical significance in studying the two forwarder scenario.

and a large (approximated to be infinite) collection of relay nodes that are waking up from a sleep mode sequentially over time. Each forwarder has to choose a relay node to forward its packet to. When a relay wakes up, each forwarder first evaluates the relay based on a reward metric (which could be a function of the progress, towards the sink, made by the relay, and the power required to get the packet to the relay [5]), and then decides whether to compete (with the other forwarder) for this relay or continue to wait for further relays to wake-up. Such a one-hop geographical forwarding setting (described in detail in Section II-D) will serve as an example application for the stochastic game formulation developed in this paper. Although we are motivated by the geographical forwarding problem, we would like to emphasize that our model, in general, is applicable wherever the problem of competitive resource allocation occurs with the resources arriving sequentially over time.

B. Related Work

Geographical Forwarding: The problem of choosing a next-hop relay arises in the context of geographical forwarding [6], where the prerequisite is that the nodes know their respective locations as well as the sink's location. For instance, Zorzi and Rao in [7] propose an algorithm called GeRaF (Geographical Random Forwarding) which, at each forwarding stage, chooses the relay making the largest progress. For a sleep-wake cycling network, in our prior work [5], we have studied a basic version of the relay selection model comprising only one forwarder. For this basic model, the solution is completely in terms of a single threshold α: forward to the first relay whose reward is more than α. Here, we will again formally show that this is in fact the solution for one forwarder when the other forwarder has already terminated (Lemma 1). However when both the forwarders are present, the solution is more involved (studied in Section III-B). Thus, the competitive model studied here is a generalization of our basic relay selection model in [5].

Channel Selection: Akin to the relay selection problem is the problem of channel selection [8], [9] where a transmitter (i.e., a forwarder), given several channels, has to choose one for its transmissions. The transmitter's decision is based on the throughput it can achieve on a channel. Although there are game theoretic versions of the channel selection problem (e.g., [10] and references there in), these are interested in determining the order in which the different channels should be sensed, so as to minimize the interference among the transmitters. In contrast, we propose a framework for selecting a channel (i.e., relay), as and when the channels become available ephemerally over time; our objective is to minimize a combination of selection delay and the quality of the chosen channel.

Cooperative Communication: Game theory has been extensively used in the context of relay selection for cooperative communication (see e.g., [11]–[14] and references there in). For instance, the authors in [11] consider the problem of multiple forwarders (or sources) simultaneously choosing a relay node (from a set of relays) so as to maximize their respective achievable rates (which is a function of the transmit power available at the chosen relay). The above problem is formulated as a *Chinese restaurant game* [15] where the objective of the diners (i.e., forwarders) is to choose tables (i.e., relays) such that the satisfaction of their dining experience is maximized. Similarly, authors in [12] use *double auction theory* to efficiently match forwarders (i.e., buyers) with relays (i.e., sellers). However, in these work it is assumed that all the relays are a priori available to the forwarders for acquisition, and hence there is no notion of delay. This is in contrast to our framework where the relays become available sequentially over time, at which instants the forwarders choose whether or not to compete for this relay; forfeiting a relay will result in an increase in delay, however with the prospects of finding a better relay in the future. Thus, our work can be considered as a competitive version of the *exploration vs. exploitation* framework.

Asset Selling: Finally, our relay selection problem can be considered to be equivalent to the asset selling problem studied in the operations research literature. The basic asset selling problem [16, Section 4.4] [17] comprises a single seller (analogous to a forwarder in our model) and a sequence of independent and identically distributed (i.i.d.) offers (rewards in our case). The seller's objective is to choose an offer so as to maximize a combination of the offer value and the number of previous offers rejected. Our competitive model here can be considered as a game variant of this basic asset selling problem. Although a game variant has been studied in [18], where there are two sellers (like two forwarders in our case) and a sequence of buyers offering possibly different value to each seller, the specific cost structure in our problem enables us to prove results such as the existence of Nash equilibrium policy pair within the class of threshold rules (Theorem 4). We also study a completely observable case which is not considered in [18].

C. Outline and Our Contributions

We will formally describe the system model in Section II. In Sections III and IV we will study two variants of the problem (of progressive complexity), namely, the completely observable (CO) and the partially observable (PO) cases. In the CO case, each forwarder (or competitor), in addition to observing its reward, can also observe the reward offered to the other forwarder (or competitor), while in the PO case the latter is not observable. We use stochastic games (SGs) and partially observable SGs to characterize solution in terms of (stationary) Nash Equilibrium Policy Pairs (NEPPs). The following are our main technical contributions:

- For the CO case we obtain results illustrating the structure of NEPPs (see Fig. 2). In particular, for each forwarder we obtain two thresholds that partitions the *reward plane* (set of all possible reward pairs) into different regions. Within each region we identify strategies that are Nash equilibrium for a bimatrix game that is effectively played at each stage (Theorem 2). This result will enable us to construct NEPPs.
- For the PO case, we design a Bayesian game that is effectively played at each stage. For this Bayesian game, we prove the existence of a Nash equilibrium strategy

within the class of threshold strategies (Theorem 4). NEPPs are then constructed using this result.
- In Section V we will briefly discuss the Pareto optimal performance that is possible if the two forwarders cooperate. This result will provide a benchmark against which the performance of the various NEPPs can be compared.
- In Section VI, through numerical work (conducted using realistic parameter values) we make the following interesting observation: "Even for moderate separation between the two forwarders, the performance of all the NEPPs is close to the performance of a *simple forwarding (SF) strategy* where each forwarder behaves as if it is alone." End-to-end simulations are conducted to understand the performance achieved by the SF policy.

We will finally draw our conclusions in Section VII. Proofs of our results are available in our detailed technical report [19].

II. System Model

We will describe our formulation from a general context that can be applicable to model, in general, scenarios of resource acquisition by competing entities. For motivation from networks context, in Section II-D, we will briefly discuss the example of relay selection for geographical forwarding in sleep-wake cycling wireless networks.

Let $\mathscr{F}_1$ and $\mathscr{F}_2$ denote the two competing nodes (i.e., players in game theoretic terms), referred to as the *forwarders*. We will assume that there are an infinite number of *relay nodes* (or *resources* in general) that arrive sequentially at times $\{W_k : k \geq 0\}$, which are the points of a Poisson process of rate $\frac{1}{\tau}$. Thus, the *inter-arrival* times between successive relays, $U_k := W_k - W_{k-1}$, are i.i.d. exponential random variables of mean τ. We refer to the relay that arrives at the instant W_k as the *k-th relay*. Further, the k-th relay is only ephemerally available at the instant W_k.

When a relay arrives, either of the forwarders can compete for it, thereby obtaining a *reward*. Let $R_{\rho,k}$, $\rho = 1, 2$, denote the *reward* offered by the k-th relay to $\mathscr{F}_\rho$ (an example reward structure will be discussed later in this section). The rewards $R_{\rho,k}$ ($\rho = 1, 2$; $k \geq 1$) can take values from a finite set $\mathcal{R} = \{r_1, r_2, \cdots, r_n\}$, where $r_1 = -\infty$ and $r_i < r_j$ for $i < j$. The reward pairs $(R_{1,k}, R_{2,k})$ are i.i.d. across k, with their common joint p.m.f. (probability mass function) being $p_{R_1,R_2}(\cdot, \cdot)$, i.e., $\mathbb{P}(R_{1,k} = r_i, R_{2,k} = r_j) = p_{R_1,R_2}(r_i, r_j)$. For notational simplicity we will denote $p_{R_1,R_2}(r_i, r_j)$ as simply $p_{i,j}$. Further, let $p_i^{(1)}$ and $p_j^{(2)}$ denote the marginal p.m.fs of $R_{1,k}$ and $R_{2,k}$, respectively, i.e., $p_i^{(1)} = \sum_{j=1}^{n} p_{i,j}$ and $p_j^{(2)} = \sum_{i=1}^{n} p_{i,j}$.

A. Actions and Consequences

First we will study (in Section III) a *completely observable case* where the reward pair, $(R_{1,k}, R_{2,k})$, is revealed to both the forwarders. Later, in Section IV, we will consider a more involved (albeit more practical) *partially observable case* where only $R_{1,k}$ is revealed to $\mathscr{F}_1$, and $R_{2,k}$ is revealed to $\mathscr{F}_2$. However in either case, each time a relay becomes available, the two forwarders have to independently choose between one of the following actions:

- s: *stop* and forward the packet to the current relay, or
- c: *continue* to wait for further relays to arrive.

In case both forwarders choose to stop, then with probability (w.p.) ν_1, $\mathscr{F}_1$ gets the relay in which case $\mathscr{F}_2$ has to continue alone, while with the remaining probability ($\nu_2 = 1 - \nu_1$) $\mathscr{F}_2$ gets the relay and $\mathscr{F}_1$ continues alone. ν_ρ ($\rho = 1, 2$) could be thought of as the probability that $\mathscr{F}_\rho$ will win the contention when both forwarders attempt simultaneously. For mathematical tractability we will assume that the forwarders make their decision instantaneously at the relay arrival instants. Further, if a relay is not chosen by either forwarder (i.e., both forwarders choose to continue) we will assume that the relay disappears and is not available for further use.

B. System State and Forwarding Policy

For the CO case, $(R_{1,k}, R_{2,k})$ can be regarded as the *state* of the system at stage k, provided both forwarders have not *terminated* (i.e., chosen a relay) yet. When one of the forwarder, say $\mathscr{F}_2$, terminates, we will represent the system state as $(R_{1,k}, \boldsymbol{t})$. Similarly, let $(\boldsymbol{t}, R_{2,k})$ and $(\boldsymbol{t}, \boldsymbol{t})$ represents the state of the system when only $\mathscr{F}_1$ has terminated and when both forwarders have terminated, respectively. Formally, we can define the *state space* to be

$$\mathcal{X} = \Big\{(r_i, r_j), (r_i, \boldsymbol{t}), (\boldsymbol{t}, r_j), (\boldsymbol{t}, \boldsymbol{t}) : r_i, r_j \in \mathcal{R}\Big\}. \tag{1}$$

Given a discrete set $\mathcal{S}$, let $\Delta(\mathcal{S})$ denote the set of all p.m.f.s on $\mathcal{S}$. Then, we have the following definition.

Definition 1: A *forwarding policy* π is a mapping, $\pi : \mathcal{X} \to \Delta(\{\mathsf{s}, \mathsf{c}\})$, such that $\mathscr{F}_1$ (or $\mathscr{F}_2$) using π will choose action s or c according to the p.m.f. $\pi(x_k)$ when the state of the system at stage $k \geq 1$ is $x_k \in \mathcal{X}$. A *policy pair* (π_1, π_2) is a tuple of policies such that $\mathscr{F}_1$ uses π_1 and $\mathscr{F}_2$ uses π_2.

Note that we have restricted to the class of stationary policies only. We will denote this class of policies as Π_S.

C. Problem Formulation

Suppose the forwarders use a policy pair (π_1, π_2), and $x \in \mathcal{X}$ is the system state at stage 1. Then, let K_ρ, $\rho = 1, 2$, denote the (random) stage at which $\mathscr{F}_\rho$ forwards its packet. Then, the delay incurred by $\mathscr{F}_\rho$ ($\rho = 1, 2$), starting from the instant $W_1 = U_1$ (first relay's arrival instant), is $D_{K_\rho} = U_2 + \cdots + U_{K_\rho}$, and the reward accrued is R_{ρ,K_ρ}. Let $\mathbb{E}^x_{\pi_1,\pi_2}[\cdot]$ denote the expectation operator corresponding to the probability law, $\mathbb{P}^x_{\pi_1,\pi_2}$, governing the system dynamics when the policy pair used is (π_1, π_2) and the initial state is x. Then, the expected total cost incurred by $\mathscr{F}_\rho$ is

$$J^{(\rho)}_{\pi_1,\pi_2}(x) = \mathbb{E}^x_{\pi_1,\pi_2}\Big[D_{K_\rho} - \eta_\rho R_{\rho,K_\rho}\Big], \tag{2}$$

where $\eta_\rho > 0$. Thus, the total cost is a linear combination of delay and reward, with η_ρ being the multiplier used to trade-off between the two quantities; for instance, a smaller value of η_ρ will give more emphasis on minimizing delay to achieve a smaller cost, while as η_ρ increases the emphasis shifts towards maximizing reward.

Definition 2: We say that a policy pair (π_1^*, π_2^*) is a Nash equilibrium policy pair (NEPP) if, for all $x \in \mathcal{X}$, $J^{(1)}_{\pi_1^*,\pi_2^*}(x) \leq J^{(1)}_{\pi_1,\pi_2^*}(x)$ for any policy $\pi_1 \in \Pi_S$, and $J^{(2)}_{\pi_1^*,\pi_2^*}(x) \leq J^{(2)}_{\pi_1^*,\pi_2}(x)$ for any policy $\pi_2 \in \Pi_S$. Thus, a unilateral deviation from an NEPP is beneficial neither for $\mathscr{F}_1$ nor $\mathscr{F}_2$.

Our objective in the subsequent sections will be to characterize the solution in terms of NEPPs. However, before proceeding further, as an example, we will construct a reward structure corresponding to the problem of geographical forwarding in sleep-wake cycling wireless networks. This development will be useful in our numerical results section as well (Section VI).

D. Geographical Forwarding Example

Due to space limitation, we will keep our discussion brief; a thorough justification for all the assumptions made here is available in our detailed technical report [19, Section III].

Let v_1 and v_2 denote the locations of the two forwarders, $\mathscr{F}_1$ and $\mathscr{F}_2$, respectively. The location of the sink node is v_0. Let d denote the range of both the forwarders. See Fig. 1. Given any location $\ell \in \Re^2$, we will define the *progress*, $Z_\rho(\ell)$, made by location ℓ with respect to (w.r.t.) $\mathscr{F}_\rho$ ($\rho = 1, 2$) as $Z_\rho(\ell) = \parallel v_\rho - v_0 \parallel - \parallel \ell - v_0 \parallel$ ($\parallel \cdot \parallel$ denotes the Euclidean norm), i.e., $Z_\rho(\ell)$ is the difference between $\mathscr{F}_\rho$-to-sink and ℓ-to-sink distances. Denoting $D_\rho(\ell) = \parallel \ell - v_\rho \parallel$ to be the distance between ℓ and $\mathscr{F}_\rho$, we define the forwarding region, $\mathcal{L}_\rho$, of $\mathscr{F}_\rho$ as, $\mathcal{L}_\rho = \Big\{\ell : D_\rho(\ell) \leq d, Z_\rho(\ell) \geq 0\Big\}$. Let $\mathcal{L} = \mathcal{L}_1 \cup \mathcal{L}_2$ denote the combined forwarding region of the two forwarders.

For simplicity, we will discretize $\mathcal{L}$ into a grid of finite set of m locations $\{\ell_1, \ell_2, \cdots, \ell_m\}$. Let $L_k \in \mathcal{L}$ denote the location of the k-th relay. The locations $\{L_k : k \geq 1\}$ are i.i.d. random variables with their common p.m.f. (probability mass function) being q, i.e., $\mathbb{P}(L_k = \ell) = q_\ell$. The aspect of the general model that the same relay does not appear again can be viewed as an approximation of the case where there is a large number of relays, each waking-up randomly within a large duty-cycling period.

Next, we will use the following standard model to obtain the transmission power required by $\mathscr{F}_\rho$ to achieve an SNR (signal to noise ratio) constraint of Γ at the k-th relay (i.e., the relay waking up at the instant W_k):

$$P_{\rho,k} = \frac{\Gamma N_0}{G_{\rho,k}} \left(\frac{D_\rho(L_k)}{d_{ref}}\right)^\xi \tag{3}$$

where, N_0 is the receiver noise variance, $D_\rho(L_k)$ is the distance between $\mathscr{F}_\rho$ and the k-th relay (recall whose location is L_k), $G_{\rho,k}$ is the gain of the channel between $\mathscr{F}_\rho$ and the k-th relay, ξ is the path-loss attenuation factor, and d_{ref} is the far-field reference distance beyond which the above expression is valid [20] (our discretization of $\mathcal{L}$ is such that the distance between $\mathscr{F}_\rho$ and any $\ell \in \mathcal{L}$ is more than d_{ref}). We will assume that the set of channel gains $\{G_{\rho,k} : k \geq 1, \rho = 1, 2\}$ are i.i.d. taking values from a finite set $\mathcal{G}$. Also, let P_{max} denote the maximum transmit power with which the two forwarders can transmit. Finally, combining progress and power, we define the reward offered by the k-th relay to $\mathscr{F}_\rho$ as,

$$R_{\rho,k} = \begin{cases} \frac{Z_\rho(L_k)^a}{P_{\rho,k}^{(1-a)}} & \text{if } P_{\rho,k} \leq P_{max} \\ -\infty & \text{otherwise,} \end{cases} \tag{4}$$

where $a \in [0, 1]$ is the parameter used to trade-off between progress and power. The reward being inversely proportional to power is clear because it is advantageous to use low power to get the packet across; $R_{\rho,k}$ is made proportional to $Z_\rho(L_k)$ to promote progress towards the sink while choosing a relay for the next hop. The motivation for the above reward function comes from our earlier work [5] where, using the reward function in (4), we first solve the local decision problem of choosing a next-hop relay by a (single) forwarder; end-to-end forwarding (as in Section VI-B) is then achieved by applying the locally-optimal policy at each hop en-route to the sink. We observed that (see [5, Fig. 8]), over some range of operation, the end-to-end performance (in terms of end-to-end delay and total transmission power) achieved by the locally-optimal forwarding policy is comparable with that achieved by a (computationally intensive, stochastic shortest path based) globally-optimal solution proposed by Kim et al. [4].

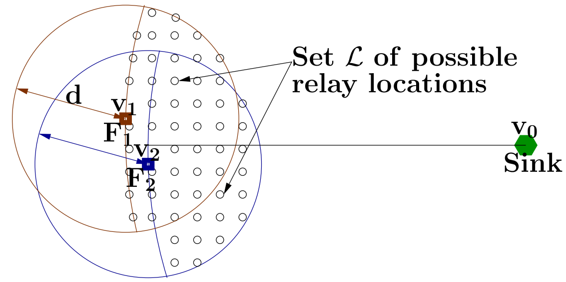

Fig. 1. Geographical forwarding example: v_0, v_1, and v_2 are the locations of the sink, $\mathscr{F}_1$, and $\mathscr{F}_2$, respectively; d is the range of each forwarder. Possible relay locations are shown as $\circ$.

III. Completely Observable (CO) Case

For the CO model we assume that the reward pair, $(R_{1,k}, R_{2,k})$, of the k-th relay is entirely revealed to both the forwarders. Recalling the geographical forwarding example, this case would model the scenario where the reward is simply the progress, $Z_\rho(L_k)$, the relay makes towards the sink, i.e., if $a = 1$ in (4). Thus, observing the location L_k of the k-th relay, both forwarders (assuming that both a-priori know the locations v_1, v_2 and v_0) can entirely compute $(R_{1,k}, R_{2,k})$.

We will first formulate the completely observable case as a stochastic game. Then, using a key theorem from Filar and Vrieze [21], We will characterize the structure of NEPPs.

A. Stochastic Game Formulation

Formulating our problem as a stochastic game [21], [22] will require us to first identify the players, state and action spaces, transition probabilities, and the one-step cost functions. In our case, the two forwarders, $\mathscr{F}_1$ and $\mathscr{F}_2$, constitute the players. $\mathcal{X}$ in (1) is the state space, and the action sets are $\mathcal{A}_1 = \mathcal{A}_2 = \{\mathsf{s}, \mathsf{c}\}$.

Transition Probabilities: Recall that $p_{i,j}$ is the joint p.m.f. of $(R_{1,k}, R_{2,k})$, $p_i^{(1)}$ and $p_j^{(2)}$ are the marginal p.m.f.s of $R_{1,k}$ and $R_{2,k}$, respectively, and ν_ρ ($\rho = 1, 2$) is the probability that $\mathscr{F}_\rho$ will win the contention if both forwarders cooperate. The transition probability when the current state is of the form $x = (r_i, r_j)$ can be written as,

$$T(x'|x, a) = \begin{cases} p_{i',j'} & \text{if } a = (\mathsf{c}, \mathsf{c}), x' = (r_{i'}, r_{j'}) \\ p_{i'}^{(1)} & \text{if } a = (\mathsf{c}, \mathsf{s}), x' = (r_{i'}, \boldsymbol{t}) \\ p_{j'}^{(2)} & \text{if } a = (\mathsf{s}, \mathsf{c}), x' = (\boldsymbol{t}, r_{j'}) \\ \nu_2 p_{i'}^{(1)} & \text{if } a = (\mathsf{s}, \mathsf{s}), x' = (r_{i'}, \boldsymbol{t}) \\ \nu_1 p_{j'}^{(2)} & \text{if } a = (\mathsf{s}, \mathsf{s}), x' = (\boldsymbol{t}, r_{j'}) \\ 0 & \text{otherwise.} \end{cases} \tag{5}$$

Note that when the joint-action is (s,s), $\nu_2 p^{(1)}_{i'}$ is the probability that $\mathscr{F}_2$ gets the current relay and the reward offered by the next relay to $\mathscr{F}_1$ is $r_{i'}$. Similarly, $\nu_1 p^{(2)}_{j'}$ is the probability (again when the joint-action is (s,s)) that $\mathscr{F}_1$ gets the relay and the reward value of the next relay to $\mathscr{F}_2$ is $r_{j'}$.

Next, when the state is of the form $x=(r_i,\boldsymbol{t})$ (i.e., $\mathscr{F}_2$ has already terminated) the transition probabilities depend only on the action a_1 of $\mathscr{F}_1$ and is given by,

$$T(x'|x,a)=\begin{cases} p^{(1)}_{i'} & \text{if } a_1=\mathsf{c}, x'=(r_{i'},\boldsymbol{t}) \\ 1 & \text{if } a_1=\mathsf{s}, x'=(\boldsymbol{t},\boldsymbol{t}) \\ 0 & \text{otherwise.}\end{cases} \tag{6}$$

Similarly one can write the expression for $T(x'|x,a)$ when the state is $x=(\boldsymbol{t},r_j)$. Finally, the state $(\boldsymbol{t},\boldsymbol{t})$ is absorbing so that $T((\boldsymbol{t},\boldsymbol{t})|(\boldsymbol{t},\boldsymbol{t}),a)=1$.

One-Step Costs: The one-step costs should be such that, for any policy pair (π_1,π_2), the sum of all one-step costs incurred by $\mathscr{F}_\rho$ $(\rho=1,2)$ should equal the total cost in (2). With this in mind, in Table 1 we write the pair of one-step costs, $(g_1(x,a),g_2(x,a))$, incurred by $\mathscr{F}_1$ and $\mathscr{F}_2$ for different joint-actions, $a=(a_1,a_2)$, when the current state is $x=(r_i,r_j)$.

$a=(a_1,a_2)$	$(g_1(x,a),g_2(x,a))$
(c,c)	(τ,τ)
(c,s)	$(\tau,-\eta_2 r_j)$
(s,c)	$(-\eta_1 r_i,\tau)$
(s,s)	$(-\eta_1 r_i,\tau)$ w.p. ν_1 $(\tau,-\eta_2 r_j)$ w.p. ν_2

TABLE 1
ONE-STEP COSTS WHEN $x=(r_i,r_j)$.

a_1	$(g_1(x,a),g_2(x,a))$
c	$(\tau,0)$
s	$(-\eta_1 r_i,0)$

TABLE 2
$x=(r_i,\boldsymbol{t})$

a_2	$(g_1(x,a),g_2(x,a))$
c	$(0,\tau)$
s	$(0,-\eta_2 r_j)$

TABLE 3
$x=(\boldsymbol{t},r_j)$

From Table 1 we see that if the joint action is (c,c) then both forwarders continue incurring a cost of τ which is the average time until the next relay wakes up. When one of the forwarder, say $\mathscr{F}_2$, chooses to stop (i.e., the joint action is (c,s)) then $\mathscr{F}_2$, forwarding its packet to the chosen relay, incurs a terminating cost of $-\eta_2 r_j$, while $\mathscr{F}_1$ simply continues incurring an average waiting delay of τ. Analogous is the case whenever the joint action is (s,c). Finally, if both forwarders compete (i.e., the case (s,s)), then with probability ν_ρ, $\mathscr{F}_\rho$ gets the relay incurring the terminating cost while the other forwarder has to continue.

When the state is of the form $(r_i,\boldsymbol{t})$ the cost incurred by $\mathscr{F}_2$ is 0 for any joint-action a, and further the one-step cost incurred by $\mathscr{F}_1$ depends only on the action a_1 of $\mathscr{F}_1$. Analogous situation holds for $\mathscr{F}_2$ when the state is $(\boldsymbol{t},r_j)$. These costs are given in Table 2 and 3, respectively. Finally, the cost incurred by both the forwarders once the termination state $(\boldsymbol{t},\boldsymbol{t})$ is reached is 0.

Total Cost: Finally, given a policy pair (π_1,π_2) (recall Definition 1) and an initial state $x\in\mathcal{X}$, let $\{X_k : k\geq 1\}$ denote the sequence of (random) states traversed by the system, and let $\{(A_{1,k},A_{2,k}) : k\geq 1\}$ denote the sequence of joint-actions. The total cost in (2) can now be expressed as the sum of all the one-step costs as follows:

$$J^{(\rho)}_{\pi_1,\pi_2}(x) \;=\; \sum_{k=1}^{\infty}\mathbb{E}^x_{\pi_1,\pi_2}\Big[g_\rho(X_k,(A_{1,k},A_{2,k}))\Big]. \tag{7}$$

B. Characterization of NEPPs

States of the form $(r_i,\boldsymbol{t})$, $(\boldsymbol{t},r_j)$: Once the system enters a state of the form $(r_i,\boldsymbol{t})$, since only $\mathscr{F}_1$ is present in the system, we essentially have an MDP problem where $\mathscr{F}_1$ is attempting to optimize its cost. Formally, if (π_1^*,π_2^*) is an NEPP then it follows that $J^{(1)}_{\pi_1^*,\pi_2^*}(r_i,\boldsymbol{t})$ is the optimal cost to $\mathscr{F}_1$ with π_1^* being an optimal policy; the cost incurred by $\mathscr{F}_2$ is 0 and π_2^* can be arbitrary, but for simplicity we fix $\pi_2^*(r_i,\boldsymbol{t})=\mathsf{s}$ for all $i\in[n]:=\{1,2,\cdots,n\}$. Hence $J^{(1)}_{\pi_1^*,\pi_2^*}(\cdot,\boldsymbol{t})$ satisfies the following Bellman optimality equation [16]:

$$J^{(1)}_{\pi_1^*,\pi_2^*}(r_i,\boldsymbol{t}) \;=\; \min\Big\{-\eta_1 r_i, D^{(1)}\Big\} \tag{8}$$

where

$$D^{(1)} \;=\; \tau+\sum_{i'}p^{(1)}_{i'}J^{(1)}_{\pi_1^*,\pi_2^*}(r_{i'},\boldsymbol{t}) \tag{9}$$

is the expected *cost of continuing alone in the system*; τ is the expected one-step cost (of waiting for the next relay) and the remaining term is the future cost-to-go. In the min-expression above, $-\eta_1 r_i$ is the cost of stopping. Thus, denoting $\frac{D^{(1)}}{-\eta_1}$ by $\alpha^{(1)}$, whenever the state is of the form $(r_i,\boldsymbol{t})$ an optimal policy is as follows:

$$\pi_1^*(r_i,\boldsymbol{t}) \;=\; \begin{cases}\mathsf{s} & \text{if } r_i\geq\alpha^{(1)}\\ \mathsf{c} & \text{otherwise.}\end{cases} \tag{10}$$

The following lemma provides an alternate method for obtaining $\alpha^{(1)}$.

Lemma 1: $\alpha^{(1)}$ is the unique fixed point of the function

$$\beta^{(1)}(x):=\mathbb{E}\Big[\max\{x,R_1\}\Big]-\frac{\tau}{\eta_1},\quad x\in(-\infty,r_n]. \tag{11}$$

Proof: We will first show that $\beta^{(1)}$ is a contraction mapping. Then, from the Banach fixed point theorem [23] it follows that there exists a unique fixed point α^* of $\beta^{(1)}$. Next, through an induction argument we will prove that $J^{(1)}_{\pi_1^*,\pi_2^*}(r_i,\boldsymbol{t}) = \min\Big\{-\eta_1 r_i,-\eta_1\alpha^*\Big\}$. Finally, substituting for $J^{(1)}_{\pi_1^*,\pi_2^*}(r_i,\boldsymbol{t})$ in $\alpha^{(1)}=\frac{D^{(1)}}{-\eta_1}$ (recall $D^{(1)}$ from (9)) and simplifying, we obtain the desired result. Details of the proof are available in [19, Appendix A]. ■

Similarly, when the state is of the form $(\boldsymbol{t},r_j)$ (i.e., $\mathscr{F}_1$ has already terminated), if (π_1^*,π_2^*) is an NEPP then, $J^{(1)}_{\pi_1^*,\pi_2^*}(\boldsymbol{t},r_j)=0$ and $\pi_1^*(\boldsymbol{t},r_j)=\mathsf{s}$, while $J^{(2)}_{\pi_1^*,\pi_2^*}(\boldsymbol{t},r_j)$ satisfies

$$J^{(2)}_{\pi_1^*,\pi_2^*}(\boldsymbol{t},r_j) \;=\; \min\Big\{-\eta_2 r_j, D^{(2)}\Big\} \tag{12}$$

$$\text{where,}\quad D^{(2)}=\tau+\sum_{j'}p^{(2)}_{j'}J^{(2)}_{\pi_1^*,\pi_2^*}(\boldsymbol{t},r_{j'}). \tag{13}$$

Further, $\alpha^{(2)}=\frac{D^{(2)}}{-\eta_2}$, is the unique fixed point of

$$\beta^{(2)}(x)=\mathbb{E}\Big[\max\{x,R_2\}\Big]-\frac{\tau}{\eta_2},\quad x\in(-\infty,r_n] \tag{14}$$

where now the expectation is w.r.t. the p.m.f. $p^{(2)}$ of R_2. Finally, an optimal policy π_2^* is given by

$$\pi_2^*(\boldsymbol{t}, r_j) \quad = \quad \begin{cases} \mathsf{s} & \text{if } r_j \geq \alpha^{(2)} \\ \mathsf{c} & \text{otherwise.} \end{cases} \tag{15}$$

States of the form (r_i, r_j): This is the more interesting case, where both forwarders are present in the system and are competing to choose a relay. When the state is of the form (r_i, r_j), if $\mathscr{F}_1$ decides to continue while $\mathscr{F}_2$ chooses to stop (i.e., the joint-action is (c, s)), then $\mathscr{F}_2$ terminates by incurring a cost of $-\eta_2 r_j$ so that the next state is of the form $(r_{i'}, \boldsymbol{t})$. Hence the expected total cost incurred by $\mathscr{F}_1$, if it uses the policy in (10) from the next stage onwards, is $D^{(1)}$ (recall (9)). Similarly, if the joint-action is (s, c) then $\mathscr{F}_1$ terminates incurring a cost of $-\eta_1 r_i$, and $\mathscr{F}_2$ incurs a cost of $D^{(2)}$.

If both forwarders decide to stop (joint-action is (s, s)) then with probability ν_1, $\mathscr{F}_1$ gets the relay in which case $\mathscr{F}_2$ continues alone, and with probability ν_2 it is vice versa. Thus, the expected cost incurred by $\mathscr{F}_1$ and $\mathscr{F}_2$, respectively, are

$$E^{(1)}(r_i) \quad = \quad \nu_1(-\eta_1 r_i) + \nu_2 D^{(1)} \tag{16}$$
$$E^{(2)}(r_j) \quad = \quad \nu_1 D^{(2)} + \nu_2(-\eta_2 r_j). \tag{17}$$

Finally, if both forwarders choose to continue (i.e., if the joint-action is (c, c)) then the next state is again of the form $(r_{i'}, r_{j'})$. Thus if (π_1, π_2) is the policy pair used from the next stage onwards then the expected cost incurred by $\mathscr{F}_1$ and $\mathscr{F}_2$ are, respectively,

$$C^{(1)}_{\pi_1,\pi_2} \quad = \quad \tau + \sum_{i',j'} p_{i',j'} J^{(1)}_{\pi_1,\pi_2}(r_{i'}, r_{j'}) \tag{18}$$
$$C^{(2)}_{\pi_1,\pi_2} \quad = \quad \tau + \sum_{i',j'} p_{i',j'} J^{(2)}_{\pi_1,\pi_2}(r_{i'}, r_{j'}). \tag{19}$$

We are now ready to state the following main theorem (adaptation of [21, Theorem 4.6.5]), which relates the "NEPPs of the stochastic game" with the "Nash equilibrium strategies of a certain static bimatrix game" played at a stage.

Theorem 1: Given a policy pair, (π_1^*, π_2^*), for each state $x = (r_i, r_j)$ construct the static bimatrix game given in Table 4. Then the following statements are equivalent:

	c	s
c	$C^{(1)}_{\pi_1^*,\pi_2^*}, C^{(2)}_{\pi_1^*,\pi_2^*}$	$D^{(1)}, -\eta_2 r_j$
s	$-\eta_1 r_i, D^{(2)}$	$E^{(1)}(r_i), E^{(2)}(r_j)$

TABLE 4
THE BIMATRIX STAGE GAME.

a) (π_1^*, π_2^*) is an NEPP.
b) For each $x = (r_i, r_j)$, $(\pi_1^*(x), \pi_2^*(x))$ is a *Nash equilibrium (NE) strategy* for the bimatrix game in Table 4. Further, the expected payoff pair at this NE strategy is, $(J^{(1)}_{\pi_1^*,\pi_2^*}(x), J^{(2)}_{\pi_1^*,\pi_2^*}(x))$.

Proof Outline: Although the proof of this theorem is along the lines of the proof of Theorem 4.6.5 in [21], however some additional efforts are required since the proof in [21] is for the case where the costs are discounted, while ours is a total cost undiscounted stochastic game. But the presence of cost-free absorption states for each player makes our problem *transient* by which we mean, when the policy of one player is fixed the problem of obtaining the optimal policy for the other player is a stopping problem [24]. Using this property we have modified the proof of [21, Theorem 4.6.5] appropriately so that the result holds for our case. Complete proof is available in [19, Appendix B]. ■

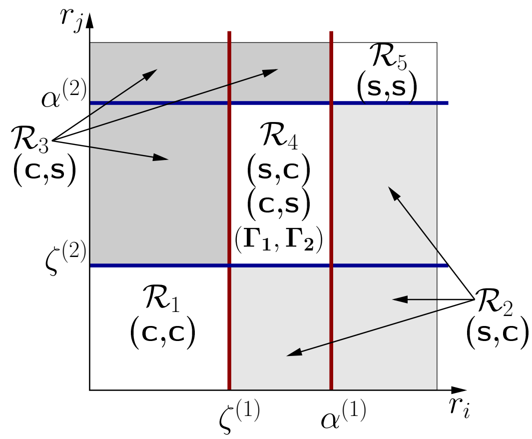

Fig. 2. Illustration of the various regions along with the NE strategy corresponding to these regions.

Assuming that the cost pair $(C^{(1)}_{\pi_1^*,\pi_2^*}, C^{(2)}_{\pi_1^*,\pi_2^*})$ is given, we now proceed to obtain all the NE strategies of the game in Table 4. We will first require the following key lemma.

Lemma 2: For an NEPP, (π_1^*, π_2^*), the various costs are ordered as follows:

$$D^{(1)} \leq C^{(1)}_{\pi_1^*,\pi_2^*} \quad \text{and} \quad D^{(2)} \leq C^{(2)}_{\pi_1^*,\pi_2^*}. \tag{20}$$

Proof: See [19, Appendix C]. ■

Discussion: The above lemma becomes intuitive once we recall that $D^{(1)}$ is the optimal cost incurred by $\mathscr{F}_1$ if it is alone in the system, while $C^{(1)}_{\pi_1^*,\pi_2^*}$ is the cost incurred if $\mathscr{F}_2$ is also present, and competing with $\mathscr{F}_1$ in choosing a relay. One would expect $\mathscr{F}_1$ to incur a lower cost without the competing forwarder.

For notational simplicity, from here on, we will denote the costs $C^{(1)}_{\pi_1^*,\pi_2^*}$ and $C^{(2)}_{\pi_1^*,\pi_2^*}$ as simply $C^{(1)}$ and $C^{(2)}$. We will write $\mathbf{C}$ for the pair $(C^{(1)}, C^{(2)})$. An important consequence of Lemma 2 is that, while solving the bimatrix game in Table 4, it is sufficient to only consider cost pairs, $(C^{(1)}, C^{(2)})$, which are ordered as in the lemma; the other cases (e.g., $D^{(1)} > C^{(1)}$ or $D^{(2)} > C^{(2)}$) cannot occur, and hence need not be considered. Further, for convenience let us denote the thresholds $\frac{C^{(1)}}{-\eta_1}$ and $\frac{C^{(2)}}{-\eta_2}$ by $\zeta^{(1)}$ and $\zeta^{(2)}$, respectively (recall that we already have, $\alpha^{(1)} = \frac{D^{(1)}}{-\eta_1}$ and $\alpha^{(2)} = \frac{D^{(2)}}{-\eta_2}$). Then, the solution (i.e., the NE strategies) to the bimatrix game in Table 4, for each (r_i, r_j) pair, is as depicted in Fig. 2.

We see that the thresholds $(\alpha^{(1)}, \zeta^{(1)})$ and $(\alpha^{(2)}, \zeta^{(2)})$ partition the reward pair set, $\{(r_i, r_j) : i, j \in [n]\}$, into 5 regions $(\mathcal{R}_1, \cdots, \mathcal{R}_5)$[2] such that the NE strategy (strategies) corresponding to each region are different. For instance, for any $(r_i, r_j) \in \mathcal{R}_1$, (c, c) (i.e., both forwarders continue) is the only NE strategy, while within $\mathcal{R}_2$, (s, c) is the NE strategy, and so on. All regions contain a unique pure NE strategy

[2]These regions depend on the cost pair $\mathbf{C}$; for simplicity we neglect $\mathbf{C}$ in their notation. We will invoke this dependency when required.

except for $\mathcal{R}_4$ where $(\mathtt{s},\mathtt{c})$, $(\mathtt{c},\mathtt{s})$, and a mixed strategy (Γ_1,Γ_2) (where Γ_ρ is the probability with which $\mathscr{F}_\rho$ chooses $\mathtt{s}$) are all NE strategies. For details on how to solve the game in Table 4 to obtain the various regions, see [19, Appendix D]. Finally, we summarize the observations made thus far in the form of the following theorem.

Theorem 2: The NE strategies of the game in Table 4 are completely characterized by the threshold pairs $(\alpha^{(\rho)},\zeta^{(\rho)})$, $\rho=1,2$ as follows (recall Fig. 2 for illustration):

- If r_i is less than $\zeta^{(1)}$, then the NE strategy recommends $\mathtt{c}$ for $\mathscr{F}_1$ irrespective of the reward value r_j of $\mathscr{F}_2$.
- On the other hand, if r_i is more than $\alpha^{(1)}$, then the NE strategy recommends action $\mathtt{s}$ for $\mathscr{F}_1$ irrespective of the value of r_j. Note that this is exactly the action $\mathscr{F}_1$ would choose if it was alone in the system; recall the discussion following (10).
- Finally, the presence of the competing forwarder $\mathscr{F}_2$ is felt by $\mathscr{F}_1$ only when its reward value r_i is between $\zeta^{(1)}$ and $\alpha^{(1)}$, in which case the NE strategies are: $(\mathtt{s},\mathtt{c})$ if $r_j<\zeta^{(2)}$; $(\mathtt{s},\mathtt{c})$, $(\mathtt{c},\mathtt{s})$ and (Γ_1,Γ_2) if $\zeta^{(2)}\le r_j\le\alpha^{(2)}$; and $(\mathtt{c},\mathtt{s})$ if $r_j>\alpha^{(2)}$.

Analogous results hold for $\mathscr{F}_2$.

Remark: We believe that the above result can be extended to the case of more than two forwarders. For instance, when three forwarders are competing, we believe that the structure of a *Nash equilibrium policy triple (NEPT)* would be analogous to the NEPP's structure depicted in Fig. 3, but with a 3-dimensional reward region partitioned into different subregions.

C. Constructing NEPPs from NE strategies

First note that $D^{(1)}$ (similarly $D^{(2)}$) can be easily computed by solving the optimality equation in (8). Now, suppose (π_1^*,π_2^*) is a NEPP such that for all $x=(r_i,r_j)\in\mathcal{R}_4(\mathbf{C})$ the NE strategy, $(\pi_1^*(x),\pi_2^*(x))$, is $(\mathtt{s},\mathtt{c})$ (recall Fig. 2); for convenience, we denote $(\pi_1^*(x),\pi_2^*(x))$ as $(\pi_1^{SC}(x),\pi_2^{SC}(x))$. Then, using Part 2(b) of Theorem 1 we can write,

$$J^{(1)}_{\pi_1^{SC},\pi_2^{SC}}(r_i,r_j)=\begin{cases} C^{(1)} & \text{if } (r_i,r_j)\in\mathcal{R}_1(\mathbf{c})\\ -\eta_1 r_i & \text{if } (r_i,r_j)\in\mathcal{R}_2(\mathbf{c})\\ D^{(1)} & \text{if } (r_i,r_j)\in\mathcal{R}_3(\mathbf{c})\\ -\eta_1 r_i & \text{if } (r_i,r_j)\in\mathcal{R}_5(\mathbf{c})\\ E^{(1)}(r_i) & \text{if } (r_i,r_j)\in\mathcal{R}_4(\mathbf{c}). \end{cases} \quad (22)$$

Using the above in (18), $C^{(1)}$ can be expressed as $C^{(1)}=\mathcal{T}_1(\mathbf{C})$, where the function $\mathcal{T}_1(\mathbf{C})$ is as given in (21) (where, for simplicity, we have used (i,j) for (r_i,r_j)). Similarly, $C^{(2)}$ can be written as $C^{(2)}=\mathcal{T}_2(\mathbf{C})$. Thus, $\mathbf{C}$ can be expressed as a fixed point of the mapping $\mathcal{T}(\mathbf{C}):=(\mathcal{T}_1(\mathbf{C}),\mathcal{T}_2(\mathbf{C}))$.

We do not have results showing that $\mathcal{T}$ indeed has a fixed point, although such a result holds for the discounted stochastic game [21, Theorem 4.6.4] (recall that ours is a transient stochastic game). However, in our numerical results section (Section VI) we are able to obtain $\mathbf{C}$ by iteration: we begin with an initial $\mathbf{C}(0)$ such that $C^{(1)}(0)<D^{(1)}$ and $C^{(2)}(0)<D^{(2)}$ (recall Lemma 2), and inductively iterate to obtain $\mathbf{C}(k)=\mathcal{T}(\mathbf{C}(k-1))$ until convergence is achieved. Finally, given a fixed point $\mathbf{C}$, we obtain the corresponding NEPP (π_1^{SC},π_2^{SC}) by constructing the various regions as in Fig. 2.

Other NEPPs: To obtain $(C^{(1)},C^{(2)})$ we had restricted (π_1^{SC},π_2^{SC}) to use NE strategy $(\mathtt{s},\mathtt{c})$ whenever $(r_i,r_j)\in\mathcal{R}_4(\mathbf{C})$. We can similarly obtain NEPPs (π_1^{CS},π_2^{CS}) and (π_1^{MX},π_2^{MX}) (whose corresponding cost pairs are denoted $\mathbf{C}_{CS}$ and $\mathbf{C}_{MX}$) by restricting, respectively, to the NE strategies $(\mathtt{c},\mathtt{s})$ and the MiXed strategy, (Γ_1,Γ_2), whenever $(r_i,r_j)\in\mathcal{R}_4(\mathbf{C}_{CS})$ and $(r_i,r_j)\in\mathcal{R}_4(\mathbf{C}_{MX})$, respectively. In Section VI we will numerically compare the performances of all these various NEPPs.

IV. Partially Observable (PO) Case

Let us first formally introduce a finite *location set* $\mathcal{L}$. Let L_k denote the *location* of the k-th relay. The locations $\{L_k:k\ge 1\}$ are i.i.d with their common p.m.f. being $(q_\ell:\ell\in\mathcal{L})$. Recall that for the PO case we assume that only $R_{\rho,k}$ is revealed to $\mathscr{F}_\rho$ $(\rho=1,2)$. In addition, we will assume that L_k is revealed to both the forwarders.

Recalling the geographical forwarding example from Section II-D, the PO case corresponds to the scenario where, in addition to L_k, the gains $G_{\rho,k}$ are required to compute $R_{\rho,k}$, i.e., if $a<1$ in (4). Hence, $\mathscr{F}_1$ not knowing $G_{2,k}$ cannot compute $R_{2,k}$. However, knowing the channel gain distribution (recall that the gains are identically distributed) it is possible for $\mathscr{F}_1$ to compute the probability distribution of $R_{2,k}$ given L_k. Similarly, $\mathscr{F}_2$ can compute the distribution of $R_{1,k}$ given L_k. Further, since the gains, $(G_{1,k},G_{2,k})$, are independent, it follows that $R_{1,k}$ and $R_{2,k}$ are independent given L_k (but unconditionally they may be dependent).

Formally, given that $L_k=\ell$, we will assume the following *independence condition*:

$$p_{R_1,R_2|L_k}(r_i,r_j|\ell)=p_{R_1|L_k}(r_i|\ell)p_{R_2|L_k}(r_j|\ell). \quad (23)$$

For simplicity, we will denote the conditional p.m.f.s $p_{R_1|L_k}(r_i|\ell)$ and $p_{R_2|L_k}(r_j|\ell)$, $i,j\in[n]$, by $p^{(1)}_{i|\ell}$ and $p^{(2)}_{j|\ell}$, respectively. The above independence condition is essential to prove a key result later (see the remark following Lemma 4).

We will formulate our PO model as a partially observable stochastic game (POSG). We will first formally describe the problem setting, before proceeding to our main results.

A. Problem Formulation

If both forwarders are still competing when the k-th relay arrives, then the observations of $\mathscr{F}_1$ and $\mathscr{F}_2$ are of the form (r_i,ℓ) and (ℓ,r_j), respectively, where (r_i,r_j) is the actual state, $L_k=\ell$ is the location of the k-th relay. Suppose $\mathscr{F}_2$ has already terminated before stage k then the location information is no more required by $\mathscr{F}_1$, and hence we will denote its

$$\mathcal{T}_1(\mathbf{c}) \;=\; \tau+\sum_{(i,j)\in\mathcal{R}_1(\mathbf{c})}p_{i,j}c^{(1)}+\sum_{(i,j)\in\mathcal{R}_2(\mathbf{c})\cup\mathcal{R}_4(\mathbf{c})}p_{i,j}(-\eta_1 r_i)+\sum_{(i,j)\in\mathcal{R}_3(\mathbf{c})}p_{i,j}d^{(1)}+\sum_{(i,j)\in\mathcal{R}_5(\mathbf{c})}p_{i,j}e^{(1)}(r_i) \quad (21)$$

observation as $(r_i, \boldsymbol{t})$ which is simply the system state. Finally, when $\mathscr{F}_1$ terminates we will use $\boldsymbol{t}$ to denote its subsequent observations. Thus, we can write the *observation space* of $\mathscr{F}_1$ as

$$\mathcal{O}_1 \quad = \quad \Big\{(r_i, \ell), (r_i, \boldsymbol{t}), \boldsymbol{t} : i \in [n], \ell \in [m]\Big\}. \tag{24}$$

Similarly, the observation space of $\mathscr{F}_2$ is given by $\mathcal{O}_2 = \Big\{(\ell, r_j), (\boldsymbol{t}, r_j), \boldsymbol{t} : j \in [n], \ell \in [m]\Big\}$.

Definition 3: We will modify[3] the definition of a policy pair, $(\overline{\pi}_1, \overline{\pi}_2)$ (see Definition 1), such that $\overline{\pi}_1 : \mathcal{O}_1 \to \{\mathtt{s}, \mathtt{c}\}$ and $\overline{\pi}_2 : \mathcal{O}_2 \to \{\mathtt{s}, \mathtt{c}\}$. Thus, the decision to stop or continue by $\mathscr{F}_1$ and $\mathscr{F}_2$, when the k-th relay arrives is based on their respective observations $o_{1,k} \in \mathcal{O}_1$ and $o_{2,k} \in \mathcal{O}_2$.

Note that we have restricted the PO policies to be deterministic (and as before stationary), i.e., $\overline{\pi}_1(o_1)$ is either $\mathtt{s}$ or $\mathtt{c}$ without mixing between the two. Let Π_D denote the set of all such deterministic policies. Restricting to Π_D is primarily to simplify the analysis. However, it is not immediately clear if a partially observable NEPP (to be formally defined very soon) should even exist within the class Π_D. Our main result is to construct a Bayesian stage game and prove that this game contains pure strategy (or deterministic) NE vectors (Theorem 4), using which PO-NEPPs in Π_D can be constructed.

Let $\{(O_{1,k}, O_{2,k})\colon k \geq 1\}$, denote the sequence of joint-observation at stage k, and let $\{X_k : k \geq 1\}$ as before denote the sequence of actual states. Then the expected cost incurred by $\mathscr{F}_\rho$, $\rho = 1, 2$, when the PO policy pair used is $(\overline{\pi}_1, \overline{\pi}_2)$, and when its initial observation is o_ρ, can be written as

$$G^{(\rho)}_{\overline{\pi}_1, \overline{\pi}_2}(o_\rho) = \sum_{k=1}^{\infty} \mathbb{E}^{o_\rho}_{\overline{\pi}_1, \overline{\pi}_2}\Big[g_\rho(X_k, (A_{1,k}, A_{2,k}))\Big], \tag{25}$$

where $A_{1,k} = \overline{\pi}_1(O_{1,k})$ and $A_{2,k} = \overline{\pi}_2(O_{2,k})$.

Similar to the completely observable case, the objective for the partially observable (PO) case will be to characterize PO-NEPPs which are defined as follows:

Definition 4: We say that a PO policy pair $(\overline{\pi}_1^*, \overline{\pi}_2^*)$ is a PO-NEPP if $G^{(1)}_{\overline{\pi}_1^*, \overline{\pi}_2^*}(o_1) \leq G^{(1)}_{\overline{\pi}_1, \overline{\pi}_2^*}(o_1)$ for all $o_1 \in \mathcal{O}_1$ and PO policy $\overline{\pi}_1 \in \Pi_D$, and $G^{(2)}_{\overline{\pi}_1^*, \overline{\pi}_2^*}(o_2) \leq G^{(2)}_{\overline{\pi}_1^*, \overline{\pi}_2}(o_2)$ where $o_2 \in \mathcal{O}_2$ and $\overline{\pi}_2 \in \Pi_D$.

We will end this section with the expressions for the various cost terms corresponding to a PO-NEPP, which are analogues of the cost terms in Section III.

Various Cost Terms: Recall the expression for $D^{(1)}$ from (9). Given a NEPP (π_1^*, π_2^*), $D^{(1)}$ is the cost incurred by $\mathscr{F}_1$ if it continues alone. Similar expression can be written for a PO-NEPP $(\overline{\pi}_1^*, \overline{\pi}_2^*)$: For $\mathscr{F}_1$ and $\mathscr{F}_2$, the cost of continuing along, respectively, are given by

$$\overline{D}^{(1)} \quad = \quad \tau + \sum_{i'} p^{(1)}_{i'} G^{(1)}_{\overline{\pi}_1^*, \overline{\pi}_2^*}(r_{i'}, \boldsymbol{t}) \tag{26}$$

$$\overline{D}^{(2)} \quad = \quad \tau + \sum_{j'} p^{(2)}_{j'} G^{(2)}_{\overline{\pi}_1^*, \overline{\pi}_2^*}(\boldsymbol{t}, r_{j'}). \tag{27}$$

The following lemma will be useful.

[3] In this section we will apply overline to most of the symbols in order to distinguish them from the corresponding symbols that have already appeared in Section III.

Lemma 3: Let (π_1^*, π_2^*) be an NEPP and $(\overline{\pi}_1^*, \overline{\pi}_2^*)$ be a PO-NEPP then $J^{(1)}_{\pi_1^*, \pi_2^*}(r_i, \boldsymbol{t}) = G^{(1)}_{\overline{\pi}_1^*, \overline{\pi}_2^*}(r_i, \boldsymbol{t})$ and $J^{(2)}_{\pi_1^*, \pi_2^*}(\boldsymbol{t}, r_j) = G^{(2)}_{\overline{\pi}_1^*, \overline{\pi}_2^*}(\boldsymbol{t}, r_j)$.

Proof: Whenever $\mathscr{F}_1$ is alone in the system, all its observations (which are of the form $(r_i, \boldsymbol{t})$ until $\mathscr{F}_1$ terminates) are exactly the actual states traversed by the system. Hence the problem of obtaining $G^{(1)}_{\overline{\pi}_1^*, \overline{\pi}_2^*}(r_i, \boldsymbol{t})$ is identical to the MDP problem of obtaining $J^{(1)}_{\pi_1^*, \pi_2^*}(r_i, \boldsymbol{t})$ in Section III-B, so that $G^{(1)}_{\overline{\pi}_1^*, \overline{\pi}_2^*}(r_i, \boldsymbol{t})$ satisfies the Bellman equation in (8). Since the solution to (8) is unique [24] we obtain $J^{(1)}_{\pi_1^*, \pi_2^*}(r_i, \boldsymbol{t}) = G^{(1)}_{\overline{\pi}_1^*, \overline{\pi}_2^*}(r_i, \boldsymbol{t})$. Similarly it follows that $J^{(2)}_{\pi_1^*, \pi_2^*}(\boldsymbol{t}, r_j) = G^{(2)}_{\overline{\pi}_1^*, \overline{\pi}_2^*}(\boldsymbol{t}, r_j)$. ■

Discussion: An immediate consequence of the above lemma is that $\overline{D}^{(1)} = D^{(1)}$ and $\overline{D}^{(2)} = D^{(2)}$. Further, if $(\overline{\pi}_1^*, \overline{\pi}_2^*)$ is a PO-NEPP then for states of the form $(r_i, \boldsymbol{t})$, $\overline{\pi}_1^*(r_i, \boldsymbol{t})$ is same as $\pi_1^*(r_i, \boldsymbol{t})$ in (10). Similarly, for states of the form $(\boldsymbol{t}, r_j)$, $\overline{\pi}_2^*(\boldsymbol{t}, r_j)$ is same as $\pi_2^*(r_i, \boldsymbol{t})$.

However, the analogues of the cost terms $C^{(1)}_{\pi_1, \pi_2}$ and $C^{(2)}_{\pi_1, \pi_2}$ (recall (18)) and (19)) are different for the PO case. The corresponding expressions are

$$\overline{C}^{(1)}_{\overline{\pi}_1, \overline{\pi}_2} \quad = \quad \tau + \sum_{\ell', i'} q_{\ell'} \cdot p^{(1)}_{i'|\ell'} \cdot G^{(1)}_{\overline{\pi}_1, \overline{\pi}_2}(r_{i'}, \ell') \tag{28}$$

$$\overline{C}^{(2)}_{\overline{\pi}_1, \overline{\pi}_2} \quad = \quad \tau + \sum_{\ell', j'} q_{\ell'} \cdot p^{(2)}_{j'|\ell'} \cdot G^{(2)}_{\overline{\pi}_1, \overline{\pi}_2}(\ell', r_{j'}). \tag{29}$$

Finally, similar to the result in Lemma 2, we can show that for a PO-NEPP $(\overline{\pi}_1^*, \overline{\pi}_2^*)$,

$$\overline{D}^{(1)} \leq \overline{C}^{(1)}_{\overline{\pi}_1^*, \overline{\pi}_2^*} \text{ and } \overline{D}^{(2)} \leq \overline{C}^{(2)}_{\overline{\pi}_1^*, \overline{\pi}_2^*}. \tag{30}$$

The proof of these is exactly along the lines of the proof of Lemma 2.

B. Bayesian Stage Game

We are now ready to provide a solution to the PO case in terms of a certain Bayesian game [25, Chapter 9] which is effectively played at the stages where both forwarders are still contending. For the CO case, given a policy pair (π_1, π_2), corresponding to each (r_i, r_j) pair we constructed the normal-form game in Table 4. However here, given a PO policy pair (π_1, π_2) and given the observation (r_i, ℓ), $\mathscr{F}_1$'s belief that the game in Table 4 (with $(C^{(1)}_{\pi_1, \pi_2}, C^{(2)}_{\pi_1, \pi_2})$ replaced by $(\overline{C}^{(1)}_{\overline{\pi}_1, \overline{\pi}_2} \overline{C}^{(2)}_{\overline{\pi}_1, \overline{\pi}_2})$) will be played is $p^{(2)}_{j|\ell}$, $j \in [n]$. Hence, $\mathscr{F}_1$ needs to first compute the costs incurred for playing $\mathtt{s}$ and $\mathtt{c}$, averaged over all observations (ℓ, r_j), $j \in [n]$, of $\mathscr{F}_2$. We will formally develop these in the following.

Strategy vectors and corresponding costs: Fixing the PO-policy pair to be $(\overline{\pi}_1, \overline{\pi}_2)$ (unless otherwise stated), we will refer to the subsequent development (which includes, the strategy vectors, various costs, best responses and NE vectors, to be discussed next) as the *Bayesian game corresponding to* $(\overline{\pi}_1, \overline{\pi}_2)$, denoted $\mathcal{G}(\overline{\pi}_1, \overline{\pi}_2)$.

Definition 5: For $\ell \in \mathcal{L}$ we define a *strategy vector*, f_ℓ, of $\mathscr{F}_1$ as $f_\ell : \{r_i : i \in [n]\} \to \{\mathtt{s}, \mathtt{c}\}$. Similarly, a strategy vector g_ℓ of $\mathscr{F}_2$ is $g_\ell : \{r_j : j \in [n]\} \to \{\mathtt{s}, \mathtt{c}\}$. Thus, given the observation (r_i, ℓ) of $\mathscr{F}_1$, f_ℓ decides for $\mathscr{F}_1$ whether to stop or continue.

Now, given the strategy vector g_ℓ of $\mathscr{F}_2$, and the location information ℓ, $\mathscr{F}_1$'s belief that $\mathscr{F}_2$ will choose action c is

$$\widetilde{g}_\ell = \sum_{j:g_\ell(r_j)=\mathsf{c}} p^{(2)}_{j|\ell}; \tag{31}$$

$(1-\widetilde{g}_\ell)$ is the probability that $\mathscr{F}_2$ will stop. Thus, the expected cost incurred by $\mathscr{F}_1$ for playing s when its observation is (r_i,ℓ) and when $\mathscr{F}_2$ uses g_ℓ is

$$C^{(1)}_{\mathsf{s},g_\ell}(r_i) = \widetilde{g}_\ell(-\eta_1 r_i) + (1-\widetilde{g}_\ell)E^{(1)}(r_i), \tag{32}$$

where, recall from (16) that $E^{(1)}(r_i) = \nu_1(-\eta_1 r_i) + \nu_2 D^{(1)}$. The various terms in (32) can be understood as follows: $\widetilde{g}_\ell$ is the probability that $\mathscr{F}_2$ will continue in which case $\mathscr{F}_1$ (having chosen the action s) stops, incurring a terminating cost of $-\eta_1 r_i$, while $(1-\widetilde{g}_\ell)$ is the probability that $\mathscr{F}_2$ will stop in which case the expected cost is, $\nu_1(-\eta_1 r_i) + \nu_2 D^{(1)}$; ν_1 is the probability that $\mathscr{F}_1$ gets the relay and terminates incurring a cost of $(-\eta_1 r_i)$, otherwise w.p. ν_2, $\mathscr{F}_2$ gets the relay in which case $\mathscr{F}_1$ continues alone, the expected cost of which is $\overline{D}^{(1)} = D^{(1)}$ (from Lemma 3).

Similarly, the expected cost of continuing when $\mathscr{F}_1$'s observation is (r_i,ℓ) can be written as,

$$C^{(1)}_{\mathsf{c},g_\ell}(r_i) = \widetilde{g}_\ell \overline{C}^{(1)}_{\overline{\pi}_1,\overline{\pi}_2} + (1-\widetilde{g}_\ell)D^{(1)}. \tag{33}$$

From the above expression we see that the cost of continuing is a constant in the sense that it does not depend on the value of r_i. Hence we will denote it as simply $C^{(1)}_{\mathsf{c},g_\ell}$. Further, note that $C^{(1)}_{\mathsf{c},g_\ell}$ depends on the PO policy pair $(\overline{\pi}_1,\overline{\pi}_2)$, but for simplicity we do not shown this dependence in the notation.

Similarly for $\mathscr{F}_2$, when its observation is (ℓ, r_j) and when $\mathscr{F}_1$ uses f_ℓ, expressions for the costs $C^{(2)}_{\mathsf{s},f_\ell}(r_j)$ and $C^{(2)}_{\mathsf{c},f_\ell}$ can be written as,

$$\begin{aligned} C^{(2)}_{\mathsf{s},f_\ell}(r_j) &= \widetilde{f}_\ell(-\eta_2 r_j) + (1-\widetilde{f}_\ell)E^{(2)}(r_j) \\ C^{(2)}_{\mathsf{c},f_\ell} &= \widetilde{f}_\ell \overline{c}^{(2)}_{\overline{\pi}^*_1,\overline{\pi}^*_2} + (1-\widetilde{f}_\ell)D^{(2)}, \end{aligned}$$

where $\widetilde{f}_\ell = \sum_{i:f_\ell(r_i)=\mathsf{c}} p^{(1)}_{i|\ell}$.

Definition 6: We say that f_ℓ is the *best response vector of* $\mathscr{F}_1$ against the strategy vector g_ℓ played by $\mathscr{F}_2$, denoted $f_\ell = BR_1(g_\ell)$, if $f_\ell(r_i) = \mathsf{s}$ iff $C^{(1)}_{\mathsf{s},g_\ell}(r_i) \le C^{(1)}_{\mathsf{c},g_\ell}$. Note that such an f_ℓ is unique. Similarly, g_ℓ is the (unique) best response against f_ℓ if, $g_\ell(r_j) = \mathsf{s}$ iff $C^{(2)}_{\mathsf{s},f_\ell}(r_j) \le C^{(2)}_{\mathsf{c},f_\ell}$. We denote this as $g_\ell = BR_2(f_\ell)$.

Definition 7: For $\ell \in \mathcal{L}$, a pair of strategy vectors (f^*_ℓ, g^*_ℓ) is said to be a *Nash equilibrium (NE) vector* for the game $\mathcal{G}(\overline{\pi}_1,\overline{\pi}_2)$ iff $f^*_\ell = BR_1(g^*_\ell)$, and $g^*_\ell = BR_2(f^*_\ell)$.

We will end this section with the following theorem which is similar to Theorem 1-(b), that was used to obtain NEPPs. This theorem will enable us to construct PO-NEPPs.

Theorem 3: Given a PO policy pair $(\overline{\pi}^*_1,\overline{\pi}^*_2)$, construct the strategy vector pair $\{(f^*_\ell,g^*_\ell) : \ell \in \mathcal{L}\}$ as follows: $f^*_\ell(r_i) = \overline{\pi}^*_1(r_i,\ell)$ and $g^*_\ell(r_j) = \overline{\pi}^*_2(\ell,r_j)$ for all $i,j \in [n]$. Now, suppose for each ℓ, (f^*_ℓ,g^*_ℓ) is a NE vector for the game $\mathcal{G}(\overline{\pi}^*_1,\overline{\pi}^*_2)$ such that,

$$\min\left\{C^{(1)}_{\mathsf{s},g^*_\ell}(r_i), C^{(1)}_{\mathsf{c},g^*_\ell}\right\} = G^{(1)}_{\overline{\pi}^*_1,\overline{\pi}^*_2}(r_i,\ell) \tag{34}$$

$$\min\left\{C^{(2)}_{\mathsf{s},f^*_\ell}(r_j), C^{(2)}_{\mathsf{c},f^*_\ell}\right\} = G^{(2)}_{\overline{\pi}^*_1,\overline{\pi}^*_2}(\ell,r_j). \tag{35}$$

Then $(\overline{\pi}^*_1,\overline{\pi}^*_2)$ is a PO-NEPP.

Proof: Available in [19, Appendix E]. ■

Discussion: If $\{(f^*_\ell,g^*_\ell)\}$ happens to be a NE vector, then from Definition 7 it simply follows that the LHS of (34) (resp. (35)) is simply the cost incurred by $\mathscr{F}_1$ (resp. $\mathscr{F}_2$) for playing the action, $f^*_\ell(r_i)$ (resp. $g^*_\ell(r_j)$), suggested by its NE vector. Thus, (34) and (35) collective say that the cost-pair obtained by playing the NE vector (f^*_ℓ,g^*_ℓ) in the Bayesian game $\mathcal{G}(\overline{\pi}^*_1,\overline{\pi}^*_2)$, is equal to the cost-pair incurred by the PO policy pair $(\overline{\pi}^*_1,\overline{\pi}^*_2)$ in the original POSG. Hence, this result could be thought as the analogue of Theorem 1-(b) proved for the completely observable case.

Existence of a NE Vector: We will fix a PO policy pair $(\overline{\pi}^*_1,\overline{\pi}^*_2)$ that satisfies the inequalities in (30). In this section we will prove that there exists a NE vector for $\mathcal{G}(\overline{\pi}^*_1,\overline{\pi}^*_2)$. Before proceeding to the main theorem we need the following results (Lemma 4 and 5).

Lemma 4: For any $\ell \in \mathcal{L}$, the best response vector, f_ℓ, against any vector g_ℓ of $\mathscr{F}_2$ is a *threshold vector*, i.e., there exists an $\Phi_\ell \in \{0,1,\cdots,n\}$ such that $f_\ell(r_i) = \mathsf{s}$ iff $i > \Phi_\ell$. We refer to Φ_ℓ as the *threshold of* f_ℓ. Similarly, if g_ℓ is the best response against any vector f_ℓ of $\mathscr{F}_1$, then g_ℓ is a threshold vector with threshold Ψ_ℓ.

Proof: Since $r_{i'} \le r_i$ whenever $i' \le i$, we can write $C^{(1)}_{\mathsf{s},g_\ell}(r_{i'}) \ge C^{(1)}_{\mathsf{c},g_\ell}(r_i)$ (see (32)). Then the proof follows by recalling Definition 6. ■

Remark: The above lemma is possible essentially because of the independence condition imposed in (23). Suppose we had worked with the model where, given only r_i, $\mathscr{F}_1$'s belief about $\mathscr{F}_2$'s observation is simply the conditional p.m.f. $p_{R_1,R_2}(r_j|r_i)$, $j \in [n]$, then, as in (31), we can write the expression for the continuing probability as

$$\widetilde{g}_{\ell,r_i} = \sum_{j:g_\ell(r_j)=\mathsf{c}} p_{R_1,R_2}(r_j|r_i), \tag{36}$$

which is now a function of r_i. If we replace $\widetilde{g}_\ell$ in (32) by $\widetilde{g}_{\ell,r_i}$ it is not possible to conclude, $C^{(1)}_{s,g_\ell}(r_{i'}) \ge C^{(1)}_{s,g_\ell}(r_i)$ whenever $i' \le i$, as required for the proof of the above lemma.

The following is an immediate consequence of Lemma 4: if (f^*_ℓ,g^*_ℓ) is a NE vector then f^*_ℓ and g^*_ℓ are both threshold vectors. Thus, we can restrict our search for NE vectors over the class of all pairs of threshold vectors. Since a threshold vector f_ℓ can be equivalently represented by its threshold Φ_ℓ we can alternatively work with the thresholds. Thus $\Phi_\ell \in \mathcal{A}_0 := \{0,1,\cdots,n\}$ represents the $n+1$ thresholds that $\mathscr{F}_1$ can use. Similarly, we will represent the $n+1$ thresholds that $\mathscr{F}_2$ can use by $\Psi_\ell \in \mathcal{A}_0$. We will write $\Phi_\ell = BR_1(\Psi_\ell)$ whenever their corresponding threshold vectors, f_ℓ and g_ℓ, respectively, are such that $f_\ell = BR_1(g_\ell)$. Similarly, we will write $\Psi_\ell = BR_2(\Phi_\ell)$ whenever $g_\ell = BR_2(f_\ell)$.

Lemma 5: (1) Let $\Psi_\ell, \Psi^o_\ell \in \mathcal{A}_0$ be two thresholds of $\mathscr{F}_2$ such that $\Psi_\ell < \Psi^o_\ell$, then the best response of $\mathscr{F}_1$ to these are ordered as, $BR_1(\Psi_\ell) \ge BR_1(\Psi^o_\ell)$. (2) Similarly, if $\Phi_\ell, \Phi^o_\ell \in \mathcal{A}_0$ are two thresholds of $\mathscr{F}_1$ such that $\Phi_\ell < \Phi^o_\ell$ then $BR_2(\Phi_\ell) \ge BR_2(\Phi^o_\ell)$.

Proof: See [19, Appendix F]. ■

We are now ready to prove the following main theorem. We will present the complete proof here because the proof technique will be required to understand the construction of PO-NEPPs discussed in the next section.

Theorem 4: For every $\ell \in \mathcal{L}$, there exists a NE vector (f_ℓ^*, g_ℓ^*) for the game $\mathcal{G}(\overline{\pi}_1^*, \overline{\pi}_2^*)$.

Proof: As mentioned earlier, a consequence of Lemma 4 is that it is sufficient to restrict our search for NE vectors within the class of all pairs of threshold vectors. Let $\mathcal{A}_0 := \{\Phi_\ell : 0 \le \Phi_\ell \le n\}$ denote the set of all $n+1$ thresholds of $\mathscr{F}_1$. Now, for $1 \le k \le n$, inductively define the sets $\mathcal{B}_k$ and $\mathcal{A}_k$ as follows: $\mathcal{B}_k = \Big\{BR_2(\Phi_\ell) : \Phi_\ell \in \mathcal{A}_{k-1}\Big\}$ and $\mathcal{A}_k = \Big\{BR_1(\Psi_\ell) : \Psi_\ell \in \mathcal{B}_k\Big\}$.

It is easy to check that through this *inductive process* we will finally end up with non-empty sets $\mathcal{A}_n$ and $\mathcal{B}_n$ such that

- for each $\Phi_\ell \in \mathcal{A}_n$ there exists a unique $\Psi_\ell \in \mathcal{B}_n$ such that $\Phi_\ell = BR_1(\Psi_\ell)$, and
- for each $\Psi_\ell \in \mathcal{B}_n$ there exists a unique $\Phi_\ell \in \mathcal{A}_n$ such that $\Psi_\ell = BR_2(\Phi_\ell)$.

Since best responses are unique, these would also mean that $|\mathcal{A}_n| = |\mathcal{B}_n|$.

Note that there is nothing special about this inductive process in the sense that for any normal form game with two player, each of whose action set is $\mathcal{A}_0$, this inductive process will still yield sets $\mathcal{A}_n$ and $\mathcal{B}_n$ satisfying the above properties whenever the best responses are unique. However, it is possible that there exists no pair $(\Phi_\ell, \Psi_\ell) \in \mathcal{A}_n \times \mathcal{B}_n$ such that $\Phi_\ell = BR_1(\Psi_\ell)$ and $\Psi_\ell = BR_2(\Phi_\ell)$. For instance, $\mathcal{A}_n = \{\Phi_\ell, \Phi'_\ell\}$, $\mathcal{B}_n = \{\Psi_\ell, \Psi'_\ell\}$ and $BR_2(\Phi_\ell) = \Psi_\ell$ and $BR_2(\Phi'_\ell) = \Psi'_\ell$ while $BR_1(\Psi_\ell) = \Phi'_\ell$ and $BR_1(\Psi'_\ell) = \Phi_\ell$. This is precisely where Lemma 5 will be useful, due to which such a situation cannot arise in our case.

Now, arrange the $N = |\mathcal{A}_n|$ $(= |\mathcal{B}_n|)$ remaining thresholds in $\mathcal{A}_n$ and $\mathcal{B}_n$ as, $\Phi_{\ell,1} < \Phi_{\ell,2} < \cdots < \Phi_{\ell,N}$ and $\Psi_{\ell,1} < \Psi_{\ell,2} < \cdots < \Psi_{\ell,N}$, respectively. Then $\Phi_{\ell,1} = BR_1(\Psi_{\ell,N})$, since if not then using Lemma 5 we can write $\Phi_{\ell,1} < BR_1(\Psi_{\ell,N}) \le BR_1(\Psi_{\ell,t})$ for every $t = 1, 2, \cdots, N$ contradicting the fact that $\Phi_{\ell,1}$ being in $\mathcal{A}_n$ has to be the best response for some $\Psi_{\ell,t} \in \mathcal{B}_n$. Similarly $\Psi_{\ell,N} = BR_2(\Phi_{\ell,1})$, otherwise again from Lemma 5 we obtain $\Psi_{\ell,N} > BR_2(\Phi_{\ell,1}) \ge BR_2(\Phi_{\ell,t})$ for every $t = 1, 2, \cdots, N$ leading to a contradiction that $\Psi_{\ell,N}$ is not the best response of any $\Phi_{\ell,t} \in \mathcal{A}_n$. Thus the threshold strategy pair (f_ℓ^*, g_ℓ^*) corresponding to the threshold pair $(\Phi_{\ell,1}, \Psi_{\ell,N})$ is a NE vector. By an inductive argument, it can be shown that all the threshold vector pairs corresponding to the threshold pairs $(\Psi_{\ell,t}, \Psi_{\ell,N-(t-1)})$, $t = 1, 2, \cdots, N$, are NE vectors. ■

C. PO-NEPP Construction from NE Vectors

Once we have obtained NE vectors (f_ℓ^*, g_ℓ^*), for each $\ell \in [m]$, The procedure for constructing PO-NEPP from NE vectors is along the same lines as the construction of NEPP from NE strategies (see Section III-C).

We begin with a pair of cost terms, $\overline{\mathbf{C}} = (\overline{C}^{(1)}, \overline{C}^{(2)})$, satisfying (30). Using the procedure in the proof of Theorem 4, we obtain, for each $\ell \in \mathcal{L}$, the NE vector $(f_\ell^{LH}, g_\ell^{LH})$ corresponding to the threshold pair $(\Phi_{\ell,1}, \Psi_{\ell,N})$ (i.e., $\mathscr{F}_1$ using lowest threshold while $\mathscr{F}_2$ uses the highest; hence the superscript LH in the notation). Then we define

$$\begin{aligned} G^{(1)}(r_i, \ell) &= \min\Big\{C^{(1)}_{\mathsf{s}, g_\ell^{LH}}(r_i), C^{(1)}_{\mathsf{c}, g_\ell^{LH}}\Big\} \\ G^{(2)}(\ell, r_j) &= \min\Big\{C^{(2)}_{\mathsf{s}, f_\ell^{LH}}(r_j), C^{(2)}_{\mathsf{c}, f_\ell^{LH}}\Big\}. \end{aligned}$$

Now recall the expressions for the costs $\overline{C}^{(1)}$ and $\overline{C}^{(2)}$ (see 28)). Compute the RHS of these expressions by replacing $G^{(1)}_{\overline{\pi}_1^*, \overline{\pi}_2^*}(\cdot)$ and $G^{(2)}_{\overline{\pi}_1^*, \overline{\pi}_2^*}(\cdot)$ by the functions $G^{(1)}(\cdot)$ and $G^{(2)}(\cdot)$, respectively. Denote the computed sums as $\overline{\mathcal{T}}_1(\overline{\mathbf{C}})$ and $\overline{\mathcal{T}}_2(\overline{\mathbf{C}})$, respectively. Suppose $\overline{\mathbf{C}}$ is such that $\overline{\mathbf{C}} = (\overline{\mathcal{T}}_1(\overline{\mathbf{C}}), \overline{\mathcal{T}}_2(\overline{\mathbf{C}}))$ (we inductively iterate to obtain such a $\overline{\mathbf{C}}$) then using Theorem 3 we can construct the PO-NEPP, $(\overline{\pi}_1^{LH}, \overline{\pi}_2^{LH})$ using $(f_\ell^{LH}, g_\ell^{LH})$ as follows: for each $i, j \in [n]$ and $\ell \in \mathcal{L}$, $\overline{\pi}_1^{LH}(r_i, \ell) = f_\ell^{LH}(r_i)$ and $\overline{\pi}_2^{LH}(\ell, r_j) = g_\ell^{LH}(r_j)$.

Finally, since the threshold vector $(f_\ell^{HL}, g_\ell^{HL})$ corresponding to the threshold pair $(\Phi_{\ell,N}, \Psi_{\ell,1})$ (i.e., $\mathscr{F}_1$ using the Highest threshold while $\mathscr{F}_2$ uses the Lowest; hence the superscript HL) is also a NE vector, one can similarly construct the PO-NEPP, $(\overline{\pi}_1^{HL}, \overline{\pi}_2^{HL})$, using $(f_\ell^{HL}, g_\ell^{HL})$.

V. Cooperative Case

It will be interesting to benchmark the best performance that can be achieved if both forwarders would *cooperate* with each other. In this section, we will briefly discuss this case leading to the construction of a *Pareto optimal* performance curve.

We will assume the completely observable case. The definition of a policy pair (π_1, π_2) and the costs $J^{(1)}_{\pi_1,\pi_2}(x)$ and $J^{(2)}_{\pi_1,\pi_2}(x)$ will remain as in Section III. However, here our objective is instead to optimize a linear combination of the two costs. Formally, let $\gamma \in (0,1)$, then the problem we are interested in is,

$$\text{Minimize}_{(\pi_1,\pi_2)} \Big(\gamma J^{(1)}_{\pi_1,\pi_2}(x) + (1-\gamma) J^{(2)}_{\pi_1,\pi_2}(x)\Big). \qquad (37)$$

Let $(\pi_1^\gamma, \pi_2^\gamma)$ denote the policy pair which is optimal for the above problem. Then, using (18) and (19), it is easy to show that $(\pi_1^\gamma, \pi_2^\gamma)$ is also optimal for

$$\text{Minimize}_{(\pi_1,\pi_2)} \Big(\gamma C^{(1)}_{\pi_1,\pi_2} + (1-\gamma) C^{(2)}_{\pi_1,\pi_2}(x)\Big). \qquad (38)$$

We have the following lemma.

Lemma 6: The policy pair $(\pi_1^\gamma, \pi_2^\gamma)$ is *Pareto optimal*, i.e., for any other policy (π_1, π_2),

(1) if $C^{(1)}_{\pi_1,\pi_2} < C^{(1)}_{\pi_1^\gamma,\pi_2^\gamma}$ then $C^{(2)}_{\pi_1^\gamma,\pi_2^\gamma} < C^{(2)}_{\pi_1,\pi_2}$, and
(2) if $C^{(2)}_{\pi_1,\pi_2} < C^{(2)}_{\pi_1^\gamma,\pi_2^\gamma}$ then $C^{(1)}_{\pi_1^\gamma,\pi_2^\gamma} < C^{(1)}_{\pi_1,\pi_2}$.

Proof: See [19, Appendix G]. ■

Thus, by varying $\gamma \in (0,1)$, we obtain a Pareto optimal boundary whose points are $(C^{(1)}_{\pi_1^\gamma,\pi_2^\gamma}, C^{(2)}_{\pi_1^\gamma,\pi_2^\gamma})$. Details on how to obtain $(\pi_1^\gamma, \pi_2^\gamma)$ is available in [19, Appendix G].

VI. Numerical Results for the Geographical Forwarding Example

A. One-Hop Study

Setting: Recall the geographical forwarding example from Section II-D. We will fix the locations of $\mathscr{F}_1$ and $\mathscr{F}_2$ to be

$v_1 = [0, \frac{\theta}{2}]$ and $v_2 = [0, -\frac{\theta}{2}]$, respectively. Thus, the distance of separation between the two forwarders is θ meters (m); we will vary θ and study the performance of the various policies. The range of each forwarder is $d = 80$ m. The combined forwarding region is discretized into a uniform grid where the distance between the neighboring points is 5 m. Finally, the sink node is placed at $v_0 = [1000, 0]$.

Next, recall the power and reward expressions from (3) and (4), respectively. 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} milliWatts (mW) (equivalently -90 dBm) [26]. The maximum transmit power available at a node is $P_{max} = 1$ mW (equivalently 0 dBm). We allow for four different channel gain values: 0.4×10^{-3}, 0.6×10^{-3}, 0.8×10^{-3}, and 1×10^{-3}, each occurring with equal probability. We fix $\eta_1 = \eta_2 = 100$ (recall that η_ρ is the parameter used to trade-off between delay and reward (see (2)); we will denote the common value of η_ρ as simple η. The contention winning probability is $\nu_1 = 1 - \nu_2 = 0.5$. Finally, the mean inter-wake-up time $\tau = 10$ milliseconds (mSec).

We first set $\theta = 0$ m (recall that θ is the distance between the two forwarders) and, in Fig. 3(a), depict the performance of various NEPPs and PO-NEPPs as pair of costs $\mathbf{C} = (C^{(1)}, C^{(2)})$ where $C^{(\rho)}$ is the cost incurred by $\mathscr{F}_\rho$ starting from time 0 if the particular NEPP or PO-NEPP is used. Also shown in Fig. 3(a) is the performance of a *simple forwarding (SF) policy* (the point marked $\times$; to be describe next) along with the Pareto optimal boundary (the solid curve). Since, from Fig. 3(a) it is not easy to distinguish between the various points, we show a section of Fig. 3(a) as Fig. 3(b). Fig. 3(c) corresponds to $\theta = 10$ m.

Policy Pairs: The description of various policies shown in Fig. 3 are as follows (we use $\mathbf{C}_{policy}$ to denote the cost pair corresponding to a $policy$):

- SC, CS, and MX (★, ◯, and □ in Fig. 3, respectively): Performances of the NEPPs that uses the NE strategies (s, c), (c, s), and the MiXed strategy (Γ_1, Γ_2), respectively, whenever $(r_i, r_j) \in \mathcal{R}_4(\mathbf{C}_{SC})$, $\mathcal{R}_4(\mathbf{C}_{CS})$, and $\mathcal{R}_4(\mathbf{C}_{MX})$, respectively (recall Fig. 2).
- LH and HL (∇ and Δ in Fig. 3, respectively): Performances of the PO-NEPPs that are constructed by choosing, for each $\ell \in \mathcal{L}$, the thresholds $(\Phi_{\ell,1}, \Psi_{\ell,N})$ (Lowest-Highest) and $(\Phi_{\ell,N}, \Psi_{\ell,1})$ (Highest-Lowest), respectively (recall the proof of Theorem 4).
- SF ($\times$ in Fig. 3): Performance of a Simple Forwarding policy where each forwarder $\mathscr{F}_\rho$ ($\rho = 1, 2$) chooses s if and only if its reward value $r_\rho \geq \alpha^{(\rho)}$. Such a policy is optimal whenever $\mathscr{F}_\rho$ is alone in the system (recall (10)). Thus, using the simple policy each forwarder behaves as if it is alone in the system.
- Solid curve: Pareto optimal boundary achieved by the cooperative policy pair $(\pi_1^\gamma, \pi_2^\gamma)$, $\gamma \in (0, 1)$ (Section V).

Results: First, from Fig. 3(a) observe that there is a range of γ values over which the cooperative policy pair, $(\pi_1^\gamma, \pi_2^\gamma)$, outperforms other policies in terms of the cost incurred to both forwarders. However, cooperation between forwarders may not be always possible, for instance, due to lack of communication between the forwarders or simply because the forwarders are selfish. Hence, it will be useful to understand the performance of NEPPs and PO-NEPPs (depicted in Fig. 3(a), (b) and (c)) that constitutes solutions to the competitive scenario.

From Fig. 3(b) we see that operating at NEPP SC is most favorable for $\mathscr{F}_2$ since $C_{SC}^{(2)}$ is less than the cost to $\mathscr{F}_2$ at the other two NEPPs, $C_{CS}^{(2)}$ and $C_{MX}^{(2)}$. This is because, whenever $(r_i, r_j) \in \mathcal{R}_4(\mathbf{C}_{SC})$ the joint-action (s, c) played by SC fetches the least cost (of $D^{(2)}$) possibly by any strategy. In contrast, $\mathscr{F}_1$ incurs highest possible cost (of $-\eta_1 r_i$) because of which NEPP SC is least favorable for $\mathscr{F}_1$. For a similar reason, operating at NEPP CS is most favorable for $\mathscr{F}_1$ while being least favorable for $\mathscr{F}_2$. The NEPP MX which chooses the mixed strategy (Γ_1, Γ_2) whenever $(r_i, r_j) \in \mathcal{R}_4(\mathbf{C}_{MX})$ achieves a fairer cost to both forwarders, however at the expense of increased cost to both when compared with the other two NEPPs.

The performances of the PO-NEPPs, LH and HL, are worse than that of the NEPPs, thus exhibiting the loss in performance due to partial information. The PO-NEPP LH which uses the NE vector corresponding to the Lowest-Highest best response pair, $(\Phi_{\ell,1}, \Psi_{\ell,N})$ (for each $\ell \in \mathcal{L}$), provides lower cost to $\mathscr{F}_2$ than the PO-NEPP HL. This is because, $\mathscr{F}_1$ using a lower threshold will essentially choose an initial relay, thus leaving $\mathscr{F}_2$ alone in the system which can now accrue a better cost. For a similar reason, HL results in $\mathscr{F}_1$ achieving a lower cost. Finally, from Fig. 3(b) we observe that the simple forwarding policy SF has the worst performance in comparison with all other points, suggesting that it may not be wise to be operating using this policy pair when the forwarders are co-located.

However, when the separation between the forwarders is increased to $\theta = 10$ m from Fig. 3(c), we see that the performance of the SF policy improves; in fact the performances achieved by the various policy pairs are practically indistinguishable when $\theta = 10$ m (note that the magnitude of the scales in plots Fig. 3(a) and 3(c) is the same). This observation motivates us to study the effect of the separation distance θ on the resulting performance. In Fig. 3(d) we plot the expected cost incurred by $\mathscr{F}_1$ as a function of θ (similar observation applies for the expected cost of $\mathscr{F}_2$). From Fig. 3(d) we identify a threshold of $\theta_1 = 8$ m above which the SF policy and the NEPPs yield identical cost; NEPPs perform better when θ is below 8 m (this region is marked as "NEPP>SF" in Fig. 3(d)). Interestingly, we identify a second threshold of $\theta_2 = 26$ m such that whenever θ is greater than θ_2, we have $\zeta^{(\rho)} = \alpha^{(\rho)}$ (for $\rho = 1, 2$; recall Fig. 2). Thus, in the region $\theta > \theta_2$ (marked "NEPP=SF" in Fig. 3(d)) the NEPP policies themselves are identical to the SF policy; hence the performances are identical. On the other hand, within the region $[\theta_1, \theta_2]$ (marked "NEPP≈SF" in Fig. 3(d)) the SF policy achieves the same performance as NEPPs, inspite of the policies being different.

Finally, we study the effect of varying η on performance. The following plots correspond to $\theta = 10$ m (i.e., θ in the region NEPP≈SF). In Fig. 3(e) we plot the expected cost incurred by $\mathscr{F}_1$ as a function of η. We observe that, as η varies, SF yields the same cost as any NEPP would incur (similar

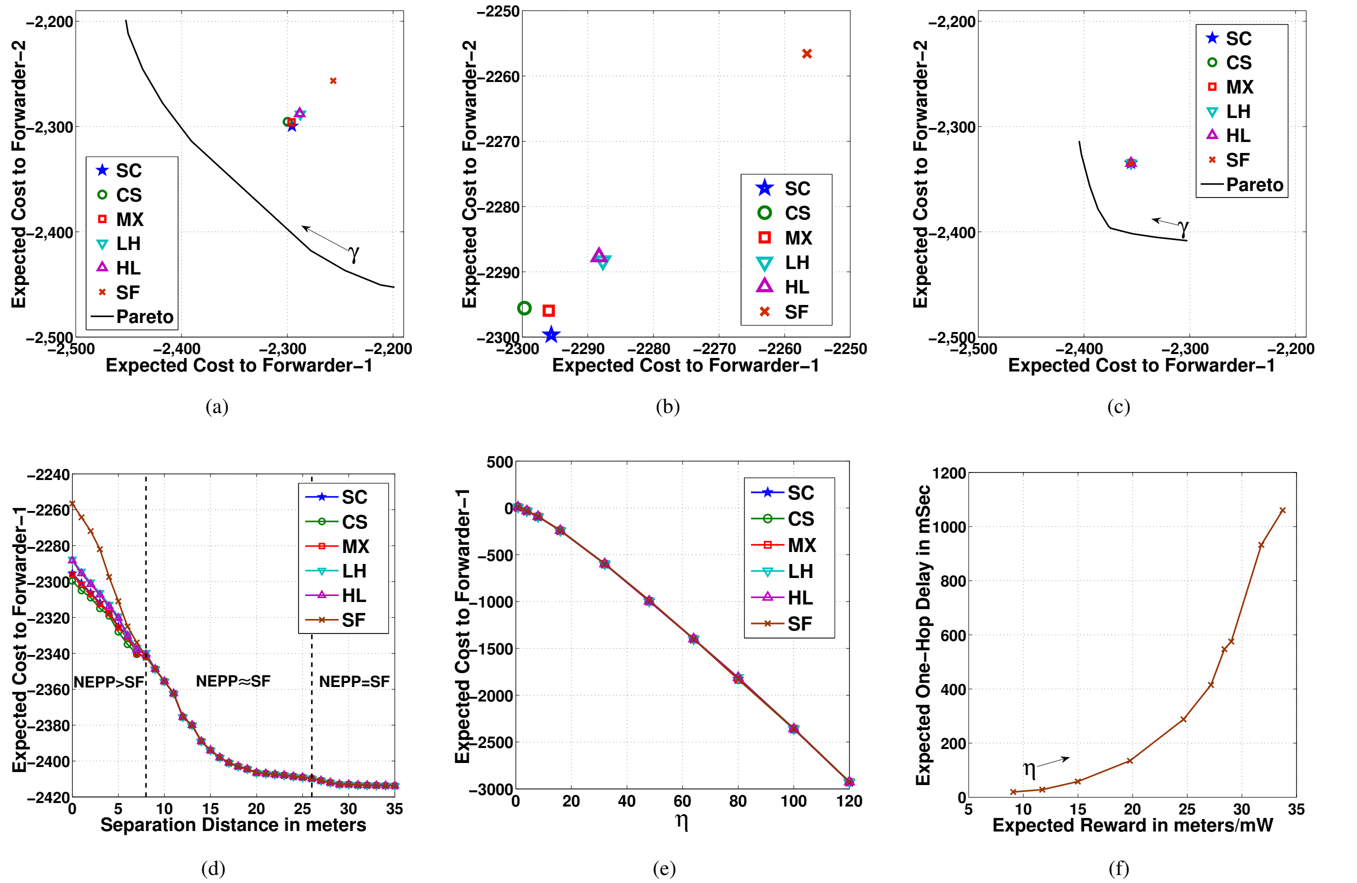

Fig. 3. Performance of various NEPPs and PO-NEPPs depicted as points in $\Re^2$ where the first (second) coordinate is the expected cost incurred by $\mathscr{F}_1$ ($\mathscr{F}_2$). Fig. (a) corresponds to the case when the forwarders are co-located (i.e., $\theta = 0$ m). A portion of Fig. (a) is enlarged and shown in Fig. (b). Fig. (c) corresponds to the case $\theta = 10$ m. In Fig. (d) and (e) we show the expected cost incurred by $\mathscr{F}_1$ as a function of θ (separation distance) and η, respectively. Fig. (f) depicts the trade-off between one-hop delay and reward that forwarder-1 can achieve using the SF policy.

observation holds for $\mathscr{F}_2$). In Fig. 3(f) we show the trade-off between one-hop delay and reward that $\mathscr{F}_1$ can achieve using the SF policy by tuning the multiplier η; as expected, a larger η yields an higher reward but at the expense of increased delay and vice verse.

Key Insight: Whenever the forwarders are even moderately separated (e.g., 8 m as in our study), the SF policy pair (where each forwarder behaves as if it is alone in the system) yields little (or, practically, no) loss in performance when compared with the performance of any NEPP or a PO-NEPP. This observation motivates us to study the end-to-end performance that the SF policy can achieve, when used for forwarding packets in a large sleep-wake cycling network.

B. *End-to-End Study*

Setting: We form a network by randomly placing $N = 5000$ nodes in a square region of area 1 km^2. A source node is placed at the coordinates $[0 \text{ km}, 0 \text{ km}]$, while a sink node is located at the diagonally opposite corner $[1 \text{ km}, 1 \text{ km}]$. Each node is allowed to asynchronously and periodically sleep-wake cycle with period $T = 100$ ms, i.e., each node i wakes up and stays ON for a small duration (which we neglect, given the other time scales) at the periodic instants $T_i + kT$, $k \geq 1$ where $\{T_i\}$ are i.i.d. uniform on $[0, T]$ (recall the discussion on the sleep-wake process from Section II). Each node i, assuming an inter-wake-up time of $\frac{T}{N_i}$ (where N_i = node density $\times$ area of the forwarding region, which is the expected number of nodes in the forwarding region of node i), computes $\alpha^{(i)}$ which is the threshold (on reward) required to implement the SF policy by node i; the values of all the other parameters required to compute the threshold, e.g., P_{max}, ξ, etc., remain the same as in our one-hop study.

The source node generates an alarm packet at time 0. We introduce competition by generating additional packets at randomly chosen nodes, randomly in time at the points of a Poisson process of rate λ. Thus, λ is the packet rate in the network, while $\frac{\lambda}{N}$ is the rate at which each node is generating packets. For instance, a network packet rate of $\lambda = 10$ packets/sec corresponds to a node packet rate of 0.002 packets/sec/node (i.e., each node is generating packets at a rate of 1 packet every 500 seconds).

All the packets are destined for the sink node (at $[1 \text{ km}, 1 \text{ km}]$). For any packet, end-to-end forwarding is achieved by simply applying the SF policy at each hop en-route to the sink. At some hop, if a relay node is simultaneously chosen by more than one forwarder (each carrying different packets), then randomly one of them will win the contention and gets the relay to forward. Thus, we allow for more than two forwarders to simultaneously compete for a relay.

Results: In Fig. 4(a) we have plotted, for different values of λ, the average end-to-end delay and the average end-to-

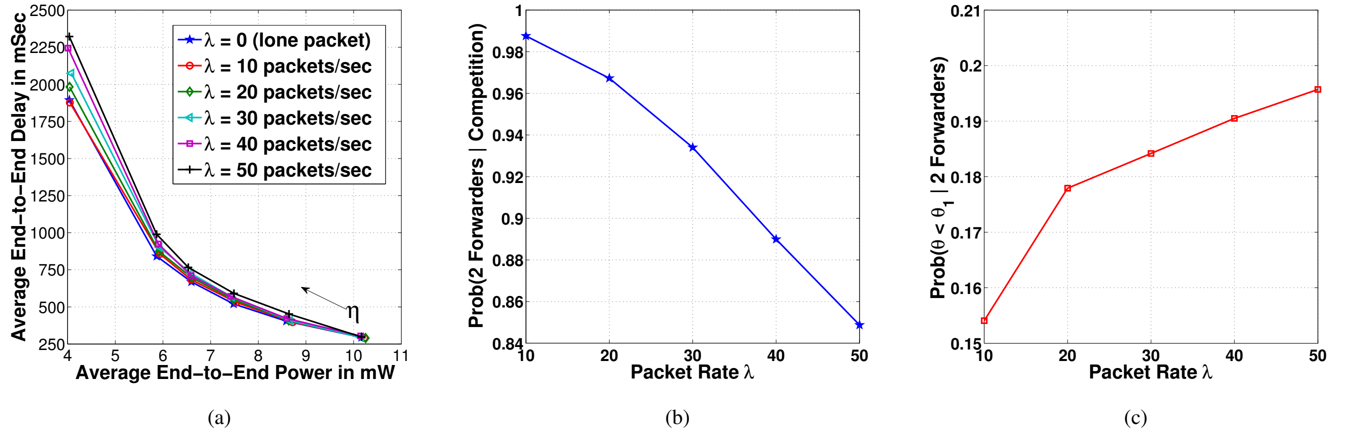

Fig. 4. (a) Average power vs. average delay curves for different values λ. (b) Conditional probability that only two forwarders are involved in the competion, given that competition occurs. (c) Conditional probability that the separation distance is less than $\theta_1 = 8$ m, given that two forwarders are competing.

end power incurred by the packets generated by the source node. Each data point in Fig. 4(a) is the average of the respective quantities over 100 alarm packets generated by the source node. The curve corresponding to each λ is obtained by varying η, which is the multiplier used to trade-off between delay and reward in the one-hop forwarding problem (recall (2) and see Fig. 3(f)). Thus, from these curves we see that by tuning the local parameter η it is possible to achieve a global trade-off between end-to-end delay and power.

Also shown in Fig. 4(a) is the performance curve corresponding to the "lone packet case" where no additional packets are generated (i.e., $\lambda = 0$). From our prior work (see [5, Fig. 8]) we already know that, for the lone packet case, the performance of the SF policy is comparable with the performance achieved by a (computationally intensive, stochastic shortest path based) globally-optimal solution proposed by Kim et al. [4]. Hence, the lone packet case serves as a benchmark for performance comparison. Thus, from Fig. 4(a) we can claim that the end-to-end performance of the SF policy is good for small values of λ (e.g., $\lambda \leq 20$); the performance, however, gradually degrades as λ increases.

In Fig. 4(b) we show the percentage of times a forwarder, currently holding the source packet, competes with just one other forwarder whenever competition occurs (i.e., probability that two forwarder are involved, given that competition occurs). We observe that for $\lambda = 10$, more than 98% of the time the two forwarder scenario is encountered. Although the above probability decreases as λ increase, it remains to be a significant percentage for all values of λ considered. This observation serves as a practical justification for studying the two forwarder scenario.

Finally, in Fig. 4(c) we show the percentage of times a competing forwarder is within $\theta_1 = 8$ m of the source-packet's forwarder, given that two forwarders compete; recall from Fig. 3(d) that when the separation distance is less than θ_1, the NEPPs are better than the SF policy in terms of the one-hop performance. We note that the fraction of times the separation distance is within $[\theta_1, \theta_2]$ (where $\theta_2 = 26m$; recall Fig. 3(d)) is approximately 70% for all values of λ considered. From Fig. 3(d) we see that the probability of θ being less than 8 m increases with λ. Thus, for large λ it may be possible to improve the end-to-end performance (recall Fig. 4(a)) by resorting to any of the NEPPs whenever the competing forwarders are located less than 8 m apart; when the separation distance is more than 8 m, the SF policy suffices as its performance is identical to the performance achieved by any NEPP. However, implementing NEPPs or PO-NEPPs for end-to-end routing has the following difficulties: (1) for a given pair of neighboring nodes, obtaining NEPPs will require fixed point iterations, (2) NEPPs are node pair dependent, so that all possible neighboring node pairs are required to compute the corresponding NEPPs; during actual forwarding a node may competing with any of its neighbors. Thus, there is a large complexity involved in implementing NEPPs. In contrast, the SF policy, being single-threshold based, is easy to implement.

VII. Conclusion

We studied the problem of competitive relay selection that arises when two forwarders are faced with the scenario of competing for a next-hop relay. In general, our formulation can be considered as a game theoretic variant of the asset selling problem (or equivalently the asset buying problem) studied in the operations research literature. We first considered the model where complete information is available to both the forwarders. We formulated the problem as a stochastic game and proceeded to obtain solution in terms of Nash equilibrium policy pairs (NEPPs). We provided valuable insights into the structure of NEPPs (Theorem 2). We next studied a partially observable case for which we constructed a Bayesian game which is effective played at each stage. For this Bayesian game, we proved the existence of a Nash equilibrium strategy within the class of (pure) threshold vectors (Theorem 4). The proof method of this result enabled us to construct NEPPs for the partial case. For the geographical forwarding example, through numerical experiments we observed that, even for moderate separation between the two forwarders, the performance of our simple forwarding (SF) policy is as good as the performance of any other NEPP/PO-NEPP. In the context of end-to-end forwarding, through simulations we established (for the considered setting) that for packet rates of less than 20 packets/second, the performance of the SF policy is good in comparison with the lone packet case.

References

[1] Y. Yao, S. Ngoga, D. Erman, and A. Popescu, "Competition-Based Channel Selection for Cognitive Radio Networks," in *IEEE Wireless Communications and Networking Conference*, April 2012.

[2] D. Miorandi, S. Sicari, F. D. Pellegrini, and I. Chlamtac, "Internet of Things: Vision, Applications and Research Challenges," *Ad Hoc Networks*, vol. 10, no. 7, pp. 1497–1516, 2012.

[3] L. Mainetti, L. Patrono, and A. Vilei, "Evolution of Wireless Sensor Networks Towards the Internet of Things: A Survey," in *19th International Conference on Software, Telecommunications and Computer Networks (SoftCOM)*, Sept 2011.

[4] J. Kim, X. Lin, and N. Shroff, "Optimal Anycast Technique for Delay-Sensitive Energy-Constrained Asynchronous Sensor Networks," *IEEE/ACM Transactions on Networking*, vol. 19, no. 2, April 2011.

[5] K. P. Naveen and A. Kumar, "Relay Selection for Geographical Forwarding in Sleep-Wake Cycling Wireless Sensor Networks," *IEEE Transactions on Mobile Computing*, vol. 12, no. 3, pp. 475–488, 2013.

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

[7] 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.

[8] P. Yang, B. Li, J. Wang, X. Y. Li, Z. Du, Y. Yan, and Y. Xiong, "Online Sequential Channel Accessing Control: A Double Exploration vs. Exploitation Problem," *IEEE Transactions on Wireless Communications*, vol. 14, no. 8, pp. 4654–4666, Aug 2015.

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

[10] Y. Xu, Q. Wu, J. Wang, L. Shen, and A. Anpalagan, "Robust Multiuser Sequential Channel Sensing and Access in Dynamic Cognitive Radio Networks: Potential Games and Stochastic Learning," *IEEE Transactions on Vehicular Technology*, vol. 64, no. 8, pp. 3594–3607, Aug 2015.

[11] M. W. Baidas and M. S. Bahbahani, "Game-Theoretic Modeling and Analysis of Relay Selection in Cooperative Wireless Networks," *Wireless Communication and Mobile Computing*, vol. 16, no. 5, pp. 500–518, April 2016.

[12] Y. Li, C. Liao, Y. Wang, and C. Wang, "Energy-Efficient Optimal Relay Selection in Cooperative Cellular Networks Based on Double Auction," *IEEE Transactions on Wireless Communications*, vol. 14, no. 8, pp. 4093–4104, Aug 2015.

[13] C. Ding, H. Yang, L. Shen, and D. Liu, "A Game Theoretic Solution for Relay Selection in Dynamic Cooperative Networks," in *WOCC 16, 25th IEEE Wireless and Optical Communication Conference*, May 2016.

[14] Y. Huo, L. Liu, L. Ma, W. Zhou, X. Cheng, T. Jing, and X. Jiang, "A Coalition Formation Game based Relay Selection Scheme for Cooperative Cognitive Radio Networks," *Wireless Networks*, pp. 1–12, 2016.

[15] C. Y. Wang, Y. Chen, and K. J. R. Liu, "Sequential Chinese Restaurant Game," *IEEE Transactions on Signal Processing*, vol. 61, no. 3, pp. 571–584, Feb 2013.

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

[17] 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.

[18] J. I. Nakagami, "A Two-Person Noncooperative Game for Assets Selling Problem: Independent Case," *Computers and Mathematics with Applications*, vol. 37, pp. 207 – 212, 1999.

[19] K. P. Naveen, E. Altman, and A. Kumar, "Competitive Selection of Ephemeral Relays in Wireless Networks," Tech. Rep., 2015. [Online]. Available: http://arxiv.org/abs/1506.06215

[20] D. Tse and P. Viswanath, *Fundamentals of wireless communication*. NY, USA: Cambridge University Press, 2005.

[21] J. Filar and K. Vrieze, *Competitive Markov Decision Processes*. NY, USA: Springer-Verlag New York, Inc., 1996.

[22] E. Altman, "Non Zero-Sum Stochastic Games in Admission, service and Routing Control in Queueing Systems," *Queueing Systems*, vol. 23, pp. 259–279, 1996.

[23] V. Pata, "Fixed Point Theorems and Applications," *Mimeo*, 2014.

[24] D. P. Bertsekas and J. N. Tsitsiklis, "An Analysis of Stochastic Shortest Path Problems," *Mathematics of Operations Research*, vol. 16, pp. 580–595, 1991.

[25] M. J. Osborne, *An Introduction to Game Theory*. Oxford University Press, USA, August 2003.

[26] A. Bhattacharya, S. Ladwa, R. Srivastava, A. Mallya, A. Rao D Sahib, S. Anand, and A. Kumar, "SmartConnect: A System for the Design and Deployment of Wireless Sensor Networks," in *COMSNETS 13': Fifth International Conference on Communication Systems and Networks*, Jan 2013.

K. P. Naveen received the B.E. degree in ECE from the Visveswaraiah Technological University, Belgaum (2005), and the Ph.D. degree from the Department of Electrical Communication Engineering, Indian Institute of Science, Bangalore (2013). Subsequently, he was a post-doctoral fellow with the INFINE team at INRIA Saclay, France. He has also worked at the ISRO Satellite Centre, Bangalore. He is currently an INSPIRE faculty at the Department of Electrical Engineering, Indian Institute of Technology (IIT) Madras. His research interests include modeling and performance analysis of wireless networks, content delivery networks, stochastic games and optimal control.

Eitan Altman Eitan Altman received the B.Sc. degree in Electrical Engineering (1984), the B.A. degree in Physics (1984) and the Ph.D. degree in Electrical Engineering (1990), all from the Technion-Israel Institute, Haifa. In 1990 he received his B.Mus. degree in music composition in Tel-Aviv University. Since 1990, Dr. Altman has been a researcher at INRIA (National research institute in computer science and control). He has been in the editorial boards of the journals Wireless Networks (WINET), Computer Networks (COMNET), Computer Communications (ComCom), J. Discrete Event Dynamic Systems (JDEDS), SIAM J. of Control and Optimisation (SICON), Stochastic Models, and Journal of Economy Dynamic and Control (JEDC). He received the best paper award in the Networking 2006, in Globecom 2007, in IFIP Wireless Days 2009 and in CNSM 2011 (Paris) conferences. His areas of interest include network engineering games, social networks and their control and the analysis through game theoretical models of network neutrality issues. He received in 2012 the Grand Prix de France Telecom from the French Academy of Sciences. On July 2014 he received the ISAACS Award from the Society of Dynamic Games for his contribution in game Theory. More information can be found at www-sop.inria.fr/members/Eitan.Altman/

Anurag Kumar (B.Tech., Indian Institute of Technology (IIT) Kanpur; Ph.D., Cornell University, both in Electrical Engineering) was with Bell Labs, Holmdel, N.J., for over 6 years. Since then he has been on the faculty of the ECE Department at the Indian Institute of Science (IISc), Bangalore; he is at present the Director of the Institute. 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 the Indian Academy of Sciences (IASc). He was an associate editor of IEEE Transactions on Networking, and of IEEE Communications Surveys and Tutorials.