Spectrum Leasing via Cooperative Opportunistic Routing

Davide Chiarotto*, Osvaldo Simeone† and Michele Zorzi‡

*Department of Information Engineering, University of Padova, Italy
†CWCSPR, New Jersey Institute of Technology, New Jersey, USA
‡California Institute of Telecommunications and Information Technology, UC San Diego, USA

Email:{dchiarot,zorzi}@dei.unipd.it,osvaldo.simeone@njit.edu

Abstract—A spectrum leasing mechanism is proposed for the coexistence between a primary and a secondary network that is based on cooperation and opportunistic routing. The primary network consists of a source and a destination communicating via a number of primary relay nodes. In each transmission slot, the next hop is selected in an on-line fashion based on the decoding outcomes in the previous transmissions according to the idea of opportunistic routing. The secondary nodes may serve as potential next hops for the primary network, but only in exchange for leasing of spectral resources so as to satisfy secondary quality-of-service constraints. Four policies based on spectrum leasing via opportunistic routing are proposed that provide different trade-offs between gains in throughput and overall energy expenditure for the primary network. Analysis is carried out for networks with a linear geometry and quasi-static Rayleigh fading statistics by using Markov chain tools.

Index Terms—Cognitive radio networks, property-rights, spectrum leasing, cooperative transmission, opportunistic routing

I. INTRODUCTION

Regulation of the coexistence of primary and secondary networks is considered to be a key issue in the design of future wireless systems. Among the proposed paradigms, the approach proposed in [1], [2] circumvents the problems associated with the so called commons model (i.e., sensing and receiver-centric interference [3], [4]) and standard spectrum leasing (i.e., pricing [5]) by leveraging the idea of spectrum leasing via cooperation. In exchange for cooperation, primary users lease part of their spectral resource to secondary users, that in turn accept to cooperate if their desired Quality-of-Service (QoS) requirements are satisfied.

The cooperation between primary and secondary users in a multihop scenario is done by means of opportunistic routing, which aims at increasing the throughput of multihop networks over fading channels by exploiting the channel diversity offered by the availability of multiple possible next hops. In particular, selection of the next hop is made in an opportunistic fashion based on the channel conditions of previous transmissions of a given packet, thanks to appropriate feedback from the decoders [6], [7].

This paper aims to improve the performance of the primary network in terms of throughput and energy expenditure by using secondary nodes as potential next hops in an opportunistic fashion. In order to exploit the additional diversity provided by the secondary users via spectrum leasing, four routing schemes are proposed to offer different gains in terms of primary throughput and energy consumption.

II. SYSTEM MODEL

We consider the system sketched in Fig. 1, in which a primary and secondary network coexist via spectrum leasing. The primary source $P_0$ wishes to communicate with the primary destination $P_k$, at a normalized distance of one, possibly via multihop routing. There are two sets of additional nodes, which are placed along two parallel linear geometries with vertical distance $\Delta_V$. The first set is composed of $k-1$ primary nodes $P_1, \ldots, P_{k-1}$ whose only role is that of forwarding information from $P_0$ to $P_k$ and they are equally spaced with inter-node distance $\Delta_H = 1/k$. The second set of nodes consists of secondary (unlicensed) nodes $S_1, \ldots, S_{k-1}$, aligned with the primary, and thus with the same inter-node distance. We will also consider a partial secondary deployment in which only one every $\alpha$ secondary nodes in Fig. 1 is active so that the number of secondary nodes is $k/\alpha - 1$ (assumed to be an integer) with inter-node distance $\alpha \Delta_H$. For simplicity, where not stated otherwise, we will assume $\alpha = 1$ in the following. Secondary nodes access the channel only if spectrum is leased by the primary network, as will be discussed below.

All devices considered work in half-duplex mode (i.e., they cannot receive and transmit at the same time) and transmission is organized in slots, each corresponding to the transmission of a packet. The process starts when the source transmits a packet with transmission rate $R$ bits/s/Hz in the first slot. In the following ones, retransmissions take place, if necessary, according to a Type-I HARQ process (i.e., retransmissions are not combined at the destination). Retransmissions in each slot may be performed by the source $P_0$, or by the primary or the secondary relays, as long as the latter have correctly decoded in the previous slot. After the packet is correctly delivered to the destination, the primary source transmits a new packet and the process repeats.

This work was partially supported by the U. S. National Science Foundation under Grant # CCF-0914899, and by the European Commission under the MEDIEVAL project (grant agreement no FP7-258053).
As discussed above, secondary nodes may serve as relays for the current primary packet only if they are granted sufficient bandwidth for their own traffic as well. Specifically, if a secondary node $S_i$ is assigned a certain slot, it uses only a portion $\beta$ of the time-slot resources (time and/or frequency) to forward the primary packet, while in the remaining fraction $1-\beta$ of the time-slot it transmits its own data. Fraction $\beta$ is selected by the secondary nodes so as to satisfy their own QoS requirements in terms of rate and reliability. An example of how this selection may be done is discussed below.

Routing decisions are made by the primary nodes that can schedule transmissions in an opportunistic fashion based on the feedback received at the end of the previous slot from all nodes that have successfully received the packet. Such a feedback informs the primary network about which nodes were able to decode the packet transmitted in the previous slot. The exact mechanism as to where and how the decision is made is not of concern here, and has been studied in (e.g., [6], [7]).

The need to allocate channel resources for feedback will not be explicitly accounted for in the analysis, as is customary in the literature on opportunistic routing.

We now detail the signal model. Let $y_{N,j}(b,t)$ be the discrete-time (complex) baseband sample sent by node $N_i \in \{P_0, \ldots, P_k, S_1, \ldots, S_k-1\}$ and received by node $N_j \in \{P_1, \ldots, P_k, S_1, \ldots, S_k-1\}$ during the $b$-th slot, at channel use $t, t = 1, \ldots, n$:

$$y_{N,j}(b,t) = d_{N,j}^{\eta/2} h_{N,j}(b) x_{N,j}(b,t) + z_{N,j}(b,t),$$

where $x_{N,j}(b,t)$ is the sample transmitted by the scheduled node $N_i$ with the per-symbol power constraint $E[x_{N,j}(b,t)^2] \leq E_N$ for each $t = 1, \ldots, n$, where $E_N$ is equal to $E_P$ or $E_S$ when the transmitter is a primary or a secondary node, respectively. The path loss between the $N_i$-th transmitter and the $N_j$-th receiver with path-loss exponent $\eta$ is represented by $d_{N,i}^{\eta/2}$. The distance between two nodes $d_{N,N_i}$ can assume three forms: (a) $d_{N_i} = |j-i| \Delta d_i$ if both nodes $N_i$ and $N_j$ are of the same type (primary or secondary); (b) $d_{N_i} = \Delta_{i}^{a_\text{primary}}$ if $N_i$ is the source $P_0$ and the destination is a primary or secondary relay or $N_j$ is a relay and $N_i$ is the destination $P_k$, where $\Delta_{i}^{a_\text{primary}} = \sqrt{(\Delta H_i)^2 + (\Delta V_i/2)^2}$; (c) $d_{N_i} = \Delta_{i}^{a_\text{secondary}}$ if the transmission is between two relays, one in the primary network and one in the secondary, with $\Delta_{i}^{a_\text{secondary}} = \sqrt{(\Delta H_i)^2 + \Delta V_i}$. The term $h_{N,j}(b)$ represents the channel coefficient between transmitter $N_i$ and receiver $N_j$, modeled with a block-fading Rayleigh model with zero mean and unit power. Channels are known to the receivers but not to the transmitters. Finally, $z_{N,j}(b,t)$ is the complex white Gaussian noise term with zero mean and power given by $E[z_{N,j}(b,t)^2] = N_0$.

We define the signal-to-noise ratio (SNR) for primary users ($\gamma_P$) as the ratio between the maximum average energy received by $P_k$ directly from the source $P_0$ and the noise power $N_0$, $\gamma_P = E_P/N_0$. Hence, the SNR for a transmission from a primary node that covers distance $d$ is given by $\gamma_P d^{-\eta}$. For consistency, we define $\gamma_S = E_S/N_0$ so that the SNR for transmission from a secondary node that covers a distance $d$ is given by $\gamma_S d^{-\eta}$.

$P_{\text{out},p}(d)$ represents the probability that a packet transmitted by a primary node is not decoded correctly by a (primary or secondary) node placed at distance $d$, and is given by:

$$P_{\text{out},p}(d) = \Pr \left\{ \log_2 \left[1 + \frac{|h|^2 \gamma_P}{d^{\eta}} \right] \leq R \right\} = 1 - e^{-\frac{R}{\gamma_P d^{-\eta}}}. $$

Similarly, secondary transmissions of a primary packet have an outage probability $P_{\text{out},s}(d) = 1 - \exp(-\frac{2R/\beta}{\gamma_S d^{-\eta}})$ for reception at a distance $d$. Notice that the rate for secondary transmissions of the primary packet is increased to $R/\beta$, to compensate for the fact that only a fraction of time (or bandwidth) $\beta$ is used for primary data.

As discussed above, the fraction $\beta$ depends on the QoS requirements of the secondary nodes. For instance, assume that each secondary node wants to transmit at rate $R_S$ to a node at distance $d_S$ with outage probability $\epsilon_S$. Recalling that a fraction $1-\beta$ of the time is used for the secondary node’s own traffic and imposing the condition on the outage probability as $1- \exp(-\frac{2R_S/(1-\beta)}{\gamma_S d_S^{-\eta}}) = \epsilon_S$ we obtain:

$$\beta = 1 - \frac{R_S}{\log_2 \left[1 - \log_2 (1 - \epsilon_S) \right]}.$$ (2)

The next section defines the performance criteria of interest in this work, namely primary throughput and average primary energy consumption. Throughout, we fix the spectrum leasing parameter $\beta$ and focus on the tradeoff between primary throughput and energy afforded by different spectrum leasing strategies based on opportunistic routing.

III. THROUGHPUT AND PRIMARY ENERGY ANALYSIS

The goal of this section is to first introduce the performance metrics of interest and then introduce four routing policies that exploit spectrum leasing via opportunistic routing to different degrees. As said, we fix parameter $\beta$, which is assumed to be calculated by the secondary network (e.g., based on (2)) and made known to the primary users, that can choose the most suitable policy for their QoS requirements.

Let $T(k, R, \beta)$ be the primary end-to-end throughput, defined as the average number of successfully transmitted bits per second per Hz, given the total number of hops $k$ and the transmission rate $R$. It is well known that this metric can be calculated, using renewal theory, as (see, e.g., [8], [9]):

$$T(k, R, \beta) = \frac{R}{E[N]},$$

where $N$ is the total number of slots, including both primary and secondary transmissions, necessary to transmit a given packet correctly from the source $P_0$ to the destination $P_k$. We also define the primary energy $E(k, R, \beta)$ as the average overall energy used by the primary network to deliver a packet successfully. We measure this quantity via $N_P$, which represents the number of primary transmissions required to

\footnote{In the rest of this work, for simplicity, we do not write explicitly the expressions of distance, $d_{N_i,N_j}$, and channel coefficient, $h_{N_i,N_j}$, between transmitter $N_i$ and receiver $N_j$, but only $d$ and $h$, with the understanding that the subscript $N_i,N_j$ is implied.}
correctly deliver a packet from the source $P_0$ to the destination $P_k$, 

\[ E(k, R, \beta) = E[N_p]. \quad (4) \]

We now detail the four proposed transmission policies for the primary packets. We remark that all four policies are based on a Type-I HARQ so that decoding is performed in each slot by discarding previous retransmissions. Extension to more complex forms of HARQ is possible but would lead to a different analysis and is left as future work.

A. Policy 1: only Primary (only-P)

The only-P policy, introduced here for reference, does not exploit spectrum leasing and uses only the primary relays $P_1, \ldots, P_{k-1}$ to opportunistically forward the packet from $P_0$ to $P_k$. Accordingly, in each slot, the transmitter is selected opportunistically as the primary node that has decoded the previous transmission and is the closest to the destination. Since we assume Type-I HARQ, the current transmitter retransmits the packet until at least one of the downstream nodes has successfully decoded.

B. Policy 2: only Secondary (only-S)

The only-S policy aims minimizing the primary transmissions, and thus the primary energy consumption, thanks to spectrum leasing, while possibly suffering some throughput loss. This is accomplished by forcing the source to send the information only through secondary nodes, i.e., without exploiting any primary relay \{P_1, \ldots, P_{k-1}\} and thus the multiuser diversity arising from their presence as well. The only primary (re)transmissions allowed are thus from the source $P_0$. So, only-S has the same topology of only-P, but a different exploitation of the relays. In fact, when a secondary relay is considered as transmitter, only a portion $\beta$ of the timeslot is used to transmit the primary packet and also the transmission power may be different (i.e., $E_p \neq E_s$ in general). An opportunistic routing scheme is used on the secondary network, where transmission is granted to the secondary node that has decoded the previous transmission and that is the closest to the destination.

C. Policy 3: Primary to Secondary (P-to-S)

With the P-to-S policy, the idea is to use primary relays until a secondary node in a “sufficiently good” position, as indicated by a parameter $m$, is able to decode. From that point on, the packet is handled by the secondary network as in the only-S strategy. Specifically, at each slot in which a primary node is transmitting, the primary network first determines the type of relay closest to the destination that has successfully decoded. If the latter node is secondary, it is selected as the next hop. If it is primary, in order to save primary energy, the node is selected only if the next best secondary nodes is at least $m$ hops behind. In other words, $m$ measures the maximum number of backward hops that one is willing to accept in order to deliver the packet to the secondary network. This window will be referred to as backward window and we generally have $0 \leq m \leq k-2$. With this policy, as with only-S, once a packet enters the secondary network, it cannot return to the primary one, except for the final destination $P_k$. This is again done in an attempt to save the primary energy, but can come at some cost in terms of throughput.

D. Policy 4: Primary and Secondary (P-and-S)

The P-and-S policy aims at fully favoring throughput maximization by removing the constraint on secondary transmissions contained in only-S and P-to-S policies. Specifically, if the transmitter is a primary node, the strategy works as for the P-to-S policy. However, if the transmitter is secondary, in order to enhance throughput, the P-and-S policy enables the selection also of primary relays $P_1, \ldots, P_{k-1}$, depending on the parameter $m$. If $m = 0$, the node closest to the destination among the ones that have decoded is selected, irrespective of whether such node is primary or secondary. This strategy clearly privileges primary throughput, since it exploits all the transmission opportunities afforded by the network. In order to obtain a more controllable trade-off between throughput and energy, we then propose to generalize this policy by letting $m > 0$. In this case, the secondary transmitter enables the selection also of primary nodes, as long as the primary node to be selected is at least $m$ hops ahead of the most advanced secondary decoding node. Thus, a primary relay can receive the packet from a secondary node only if it is outside a window (called forward window) which is of size $m$ hops\(^2\) and starts from the most advanced decoding secondary node.

We remark that the use of backward window and forward window is quite different: the backward window includes the secondary relays that can be selected as next hops when a primary node is the current transmitter; the forward window includes the primary relays that cannot be selected as the next hop when a secondary node transmits.

E. Evaluating Primary Throughput and Energy

In order to evaluate the performance metrics throughput (3) and average primary energy (4) for the protocols discussed above, we use the theory of Markov chains. We model the network with a chain of $2k$ states, one for each primary and secondary node. State $P_0$ refers to a situation where the current packet is at the source $P_0$, primary states $P_i$ and secondary $S_i$, $i = 1, \ldots, k-1$, are similarly defined, and $P_k$ represents the state where the destination has successfully decoded. Recalling that we assume Type-I HARQ, the current transmitter retransmits the packet until at least one of the nodes admitted by the specific policy has successfully decoded. Based on this, the transition matrix can be organized in four slots as

\[ \Phi = \begin{bmatrix} \Phi_{PP} & \Phi_{PS} \\ \Phi_{SP} & \Phi_{SS} \end{bmatrix}, \quad (5) \]

where the states are ordered as $P_0, P_1, \ldots, P_k, S_1, \ldots, S_{k-1}$, and $\Phi_{AB}$, $A, B \in \{P, S\}$ are the submatrices that collect all the transition probabilities from nodes of type $A$ to $B$.

In general, in matrix $\Phi_{AB}$ the term $\Phi_{AB}(i, j)$ represents the probability that, given the current state $A$, the next state is $B_j$, with $j = i, i+1, \ldots, k$ when $B = P$ and $j = 1, \ldots, k-1$ if $B = S$ (this is because the policies allow backward transmissions only if the receiver is a secondary node).

\(^2\)In principle, one could choose two different sizes for forward and backward windows, but this is not further investigated here.
and primary energy of a given policy for different policies, especially for high $m$. The only-S policy in the full secondary relay scenario is very close to the performance of P-to-S for high $m$, and is not visible in the graph, confirming the strict relationship between these two policies, especially for high $m$.

Referring to matrix (5), the first set of $k$ states and the last set of $k-1$ states are transient nodes, whereas the $k+1$-th state, corresponding to the packet being received at the destination, is absorbing. Depending on the routing policy adopted, the transition probabilities will assume different expressions and will be detailed in the Appendix. Based on the matrix (5) the average primary energy and throughput can be obtained as detailed in the Lemma below, which follows from standard Markov chain theory [8, Cap. 3] (see also [10]).

\textbf{Lemma 1.} The end-to-end throughput (3) and the primary energy (4) for fixed transmission rate $R$ are given by $T(k, R, \beta) = R/\nu_{P0}$ and $E(k, R, \beta) = w_{P0}$, where $\nu_{P0}$ and $w_{P0}$ are the first elements of vectors $\nu = [\nu_P, \nu_S]$ and $w = [w_P, w_S]$ which are evaluated as $\nu = (I - Q)^{-1} 1$ and $w = (I - Q)^{-1} r$, where $I$ is a $[(2k-1) \times 1]$ vector with all entries equal to 1 and $r$ is the reward vector $r = [r_P, r_S] = [r_{P0}, \ldots, r_{P(k-1)}, r_S]$ where $r_P$ is a $[k \times 1]$ vector with all ones and $r_S$ is a $[(k-1) \times 1]$ vector with all zero elements and $I$ is the $(2k-1) \times (2k-1)$ identity matrix. Finally, matrix $Q$ is obtained from $\Phi$ removing the $(k+1)$-th row and the $(k+1)$-th column.

\section*{IV. Numerical Results}

In this section we numerically evaluate the impact of secondary relays on the primary network using the analysis above, with the following fixed parameters: number of hops $k = 12$, path loss exponent $\eta = 3$, geometry of the network $\Delta_V = \Delta_H = 1/k$ and transmitting power of secondary users $E_S = E_P$. We consider two secondary deployments: (i) Full ($\alpha = 1$) and (ii) Partial ($\alpha > 1$).

Fig. 2 shows end-to-end throughput and primary energy by varying $m$ for full and partial secondary deployment with $\alpha = 4$ and for parameters $\gamma = -8$ dB ($E_S = E_P$, then $\gamma_S = \gamma_P = \gamma$), $R = 2.7 \, \text{bits/s/Hz}$ and $\beta = 0.8$. Each curve is obtained by evaluating the pair end-to-end throughput and primary energy of a given policy for different $m$ (ranging from 0 to 10) and keeping all the other parameters fixed. From Fig. 2, it is seen that, with the given parameters, spectrum leasing policies with full secondary deployment are more energy efficient with respect to the only-P (i.e., no spectrum leasing) policy, while their throughput performance is similar to only-P. Moreover, P-and-S outperforms only-S and P-to-S in terms of throughput but at the cost of a larger primary energy, especially in the partial secondary deployment scenario. Moreover, the roles of the parameter $m$ and of both backward and forward window are clear in allowing to trade-off energy and throughput. For the policies P-to-S and P-and-S, increasing $m$ ($m \geq 3$) trades throughput for a decreased primary energy consumption, due to the larger number of secondary transmissions admitted. When $m$ is low enough ($m \leq 3$), the throughput increases differently in P-to-S and P-and-S. For the P-to-S policy, which employs only the backward window and slots secondary transmissions to primary relays, the throughput and primary energy are larger due to the lower number of secondary nodes available to lease the spectrum. In P-and-S this limit is overcome by removing the slot from the secondary transmissions and by introducing the forward window. Thus, due to the capability of exploiting more path diversity, the P-and-S policy is able to obtain larger throughput (though at the cost of a larger energy consumption) than P-to-S.

We consider the impact of secondary QoS requirements, quantified by parameter $\beta$, for full and partial secondary deployment with $\alpha = 4$ in Fig. 3 for $\gamma = -8$ dB, $R = 2.7 \, \text{bits/s/Hz}$ and $m = 1$. We note that increasing the secondary QoS requirements (i.e., decreasing $\beta$) leads to a decreased throughput without affecting the primary energy for all policies, except P-and-S. Indeed, in all policies except P-and-S modifying $\beta$ does not change the number of primary transmissions, but only the portion of the spectrum leased to the secondary node that is used to serve primary traffic. Instead for P-and-S a smaller $\beta$ leads to both a decreased throughput and an increased primary energy, due to the larger number of secondary transmissions towards the primary network. In fact, the number of relays that are potentially reachable at
each transmission from a secondary relay decreases with $\beta$. So, when $\beta$ is low the number of secondary transmissions increases, but this only affects throughput, not primary energy. Even so, the primary energy increases. This is due to the following fact: if secondary transmissions increase, also the possibility of returning to the primary network increases. If this happens, the next hop will be covered by a primary transmission that must be taken into account in the primary energy. Moreover, with partial secondary deployment for $\alpha = 4$, P-and-S still confirms the best adaptability to the lack of secondary nodes among all the other policies, increasing primary transmissions.

Finally, Fig. 4 focuses on the impact of the transmission rate in a full secondary deployment for $\gamma = -8$ dB, $m = 1$ and $\beta = 0.8$. It is seen that, for each policy, there exists a rate that maximizes the throughput. Such rate is different for distinct policies. For example, the optimal rates for only-P is $R = 3.4$ bits/s/Hz, for only-S and P-to-S is $R = 2.7$ bits/s/Hz and for P-and-S is $R = 2.9$ bits/s/Hz. In order to reduce the primary energy consumption, at the cost of a reduced throughput, one can decrease the transmission rate $R$ for all policies except P-to-S. In fact, in general, a lower transmission rate causes a wider coverage range of the primary transmission. So, when the transmission rate is low the progress towards the destination increases, and this makes it less likely that a packet enters the secondary network. Thus, when the transmission rate decreases in the P-to-S policy, it is more likely to remain in the primary network, which leads to a higher primary energy consumption. It is finally noted that the gains of spectrum leasing over the only-P policy are substantial, irrespective of the choice of the transmission rate.

V. CONCLUDING REMARKS

This paper has proposed a novel approach to regulate the coexistence of primary and secondary nodes in multihop networks based on spectrum leasing and opportunistic routing. In particular, it is proposed that primary nodes may, in a local and dynamic fashion, select secondary nodes as next hops for primary traffic by allowing the latter to exploit the spectral resource for secondary data with some QoS guarantees. This approach is an implementation of the previously proposed idea of spectrum leasing via cooperation. We have designed different routing strategies based on this principle, that provide different trade-offs between gains in terms of primary throughput and energy. Numerical results confirm that secondary nodes permit the primary network to achieve gains in both throughput and primary energy consumption, provided that secondary QoS is satisfied. We finally remark that this work relies on several simplifying assumptions, such as the linear topology, and does not account for other physical layer techniques. These aspects will be considered in future work.

APPENDIX

In this Appendix, we sketch the derivation of the transition probability matrix (5) for the only-S policy described in Section III-B. The probabilities of the remaining policies described in Section III are reported in [11].

In only-S the only primary transmissions allowed are from the source, which leads to the following probabilities for submatrices $\Phi_{PP}$ and $\Phi_{PS}$:

$$\Phi_{PP}(0,0) = P_{out,P}(k\Delta H)\prod_{q=1}^{k-1} P_{out,P}(\Delta^{TD}_q);$$

$$\Phi_{PP}(0,k) = 1 - P_{out,P}(k\Delta H);$$

$$\Phi_{PS}(0,j) = P_{out,P}(k\Delta H)(1 - P_{out,P}(\Delta^{TD}_j)).$$

The probabilities of submatrices $\Phi_{SP}$ and $\Phi_{SS}$ reflect the fact that secondary transmissions can reach only other secondary relays or the destination:

$$\Phi_{SP}(i,k) = 1 - P_{out,S}(\Delta^{TD}_{k-i});$$

$$\Phi_{SS}(i,j) = (1 - P_{out,S}(j-i\Delta H)) P_{out,S}(\Delta^{TD}_{j-i}).$$

Finally, $\Phi_{AB}(i,j) = 0$ with $A, B \in \{P, S\}$, in all other cases.

REFERENCES


