Spectrum-Aware Beaconless Geographical Routing Protocol for Mobile Cognitive Radio Networks

Junseok Kim and Marwan Krunz
Dept. of Electrical and Computer Engineering, University of Arizona
Email: {junseok, krunz}@ece.arizona.edu

Abstract—A key challenge in routing in cognitive radio networks (CRNs) is how to adaptively and efficiently select a route and assign resources along that route according to the surrounding environment. In this work, we propose a distributed routing protocol for CRNs, in which path selection and resource allocation (e.g., spectrum, transmission power, and transmission rate) are determined by receivers. Because this process is done on a per-packet and per-hop basis, the proposed protocol can efficiently adapt to spectrum dynamics and node mobility. In addition, spectrum efficiency is increased through dynamic spectrum allocation and transmission power control. Simulation results show that delivery ratio and throughput are dramatically improved with our routing protocol.

I. INTRODUCTION

Cognitive radio (CR) is a promising concept to resolve the spectrum scarcity problem and meet a growing demand for wireless services. In contrast to classic spectrum assignment, CRs operate on unused licensed portions of spectrum without interfering with licensed primary radios (PRs). Routing in multi-hop CRNs faces unique challenges, compared with conventional mobile ad hoc networks (MANETs). In MANETs, the routing protocol is expected to adapt to node mobility and channel dynamics (e.g., fading and shadowing). In CRNs, in addition to that, PR activities and spectrum sharing among CRs necessitate modifying routes more quickly, according to spectrum availability.

Several routing protocols have been proposed for CRNs. Most of these protocols establish the route during the route discovery phase and try to change it when packets are dropped and/or new PR activity is detected. Such an approach suffers a significant performance degradation when spectrum availability and/or node locations change faster than the rate of route update (e.g., as in the case of vehicular ad hoc networks). In this paper, we propose a novel routing protocol for mobile CRNs, called Spectrum-Aware BEaconless geographical routing (SABE). We bring the concept of beaconless geographical routing to CRNs for the first time. The main idea in SABE is that routing decision as well as the resource allocation strategy (including the assignment of spectrum, transmission power, and transmission rate) are made by receivers on a per-packet and per-hop basis. A source or an intermediate CR broadcasts a forward request packet, and includes in it its available resources. Receivers calculate a link weight, considering the available spectrum at the sender and receiver, as well as the distance to the final destination. Then, a timer to reply to the request is set depending on the link weight. The receiver with the highest link weight replies first, establishing itself as the relay node.

Because the route selection is done on a per-packet and per-hop basis, SABE can efficiently adapt to spectrum dynamics and node mobility. Another feature of SABE is that the spectrum, transmission power, and transmission rate are jointly selected so that more CRs in the same vicinity can share the limited spectrum resource. In SABE, noncontiguous spectrum holes are used opportunistically with one transceiver based on the non-contiguous orthogonal frequency division multiplexing (NC-OFDM) technique, which is feasible with commercial off-the-shelf radios.

The rest of this paper is organized as follows. Section II gives an overview of related works. In Section III, we introduce the proposed routing protocol. Simulation results are presented in Section IV. Finally, Section V gives concluding remarks.

II. RELATED WORKS

Many routing protocols have been proposed for CRNs [5]. In here, we focus on distributed protocols that do not rely on full knowledge of spectrum availability and network topology. Many of these protocols are extensions of the ad-hoc on-demand distance vector (AODV) protocol. They use different metrics for path selection. AODV establishes an end-to-end route by broadcasting a route request (RREQ) packet over the network, and it tries to modify the route when packets are dropped [1]. In [6], the routing metric depends on delay factors, such as the channel switching delay, medium access delay, and queuing delay. The RREQ packet conveys the list of idle channels at each intermediate CR. The destination selects the path and the channel for each link on the path such that the total delay is minimized. In [7], the routing metric depends on link quality (e.g., packet error rate) as well as the packet delay. In [8], the PR activity degree, which represents how many channels are occupied by PRs, is piggybacked on the RREQ packet to minimize interference to PRs.

Geographical routing protocols for CRNs are presented in [9], [10]. In geographical routing, each node knows its location, e.g., using a GPS device or by implementing a localization algorithm [1]. The greedy perimeter stateless routing (GPSR) is the best known geographical routing protocol for MANETs. In GPSR, a forwarder (a source or an intermediate node) forwards a message to the neighbor who is closest to the destination and is within the forwarding area. The forwarding...
area is generally defined as the intersection of two circles: the circle of the forwarder, defined by its maximum transmission range, and the circle centered at the destination whose radius equals the distance between the forwarder and the destination. The gray area in Figure 1 depicts the forwarding area of node $u$ when the destination is node $t$. We refer to any neighbor in the forwarding area as a candidate. GPSR can lead to a dead end, where no candidate can be found in the forwarding area of a node. In this case, the message is detoured around the dead end until reaching a node that has one or more candidates.

In [9], CRs are assumed to know the locations of PRs. If there is no PR activity, the candidate that is closest to the destination is selected as the next relay node. When PR activity is detected, the candidate that is farthest from the destination is selected and the transmission power is adjusted so as not to disrupt PR transmissions. CRs operate over a single channel, so opportunistic spectrum allocation is not addressed in that paper.

The spectrum-aware routing for cognitive ad-hoc networks (SEARCH) was proposed in [10]. SEARCH discovers routes similar to GPSR, but it repeats the route discovery process over each channel. The destination combines the routes and assigns data channels such that the end-to-end delay is minimized. When a dead end is encountered, SEARCH forwards a message to the closest node to the destination outside the forwarding area, until a node with candidates is encountered. As network size increases, SEARCH incurs large latency and message overhead for route discovery. Moreover, it is well-known that forwarding a message greedily to nodes outside the forwarding area often leads to a routing loop [4]. Other works related to routing in CRNs are discussed in [5].

In GPSR, every node periodically broadcasts a beacon packet to update its location. Intuitively, the rate at which beacons are generated should be high enough to maintain the local topology up-to-date. If not, the packet drop rate can increase drastically. Beaconless geographical routing (also called contention-based geographical routing) does not require nodes to transmit beacons. The routing decision is made by the receiver [2], [3]. In this type of protocols, a forwarder “broadcasts” a request to send (RTS), and candidates set their delay timer for the reply depending on their distance to the destination. The closer a node is to the destination, the shorter is its delay, allowing that node to be the first to reply. A drawback of this scheme is that a planar graph, which is used to avoid a dead end, cannot be constructed immediately, because a forwarder does not know the locations of its neighbors. To resolve this problem, the authors in [4] presented the beaconless forwarder planarization (BFP) technique. In BFP, when there is no candidate, the neighbor closest to the forwarder responds to the request first and other neighbors that overhear the response check whether that neighbor satisfies the planarity condition (i.e., no edge crosses any other edge). If this condition is not satisfied, one or more neighbors send a protest packet to cancel the previous response.

Because beaconless geographical routing does not involve the exchange of beacon packets, and receivers make a routing decision using current information, the message overhead is relatively low. The protocol can adapt to topological changes better than classic geographical routing, especially when nodes are mobile. In this paper, we extend beaconless geographical routing to accommodate dynamic-spectrum CRNs.

III. SPECTRUM-AWARE BEACONLESS GEOGRAPHICAL ROUTING

A. Problem Setup

We consider an ad hoc CRN that coexists geographically with PRs (e.g., TV receivers or wireless microphones). As explained in Figure 2, we place OFDM subcarriers over TV channels from 21 to 50. Hereafter, we refer to an OFDM subcarrier as a subcarrier, for brevity. Subcarrier spacing is 0.2 MHz, so that the symbol duration, with the 1/4 cyclic prefix, is $(1/(0.2 \times 10^9)) \times 1/4 = 6.25$ us, which is similar to that of WiMAX [14]. A subcarrier constitutes a data or control subchannel. Three data subchannels1 are placed in a TV channel, and one control subchannel is placed between two consecutive TV channels. In [13], the authors showed that a few subcarriers can be placed between two consecutive TV channels so that the TV reception is not disturbed by CR nodes. The CR sender and receiver exchange control packets over the control channel, which consists of 31 non-contiguous control subchannels. The sender subsequently transmits a data packet over the data channel, which consists of multiple non-contiguous data subchannels, using one transceiver by suppressing subcarriers currently used by PRs and other CRs. A pilot subcarrier is placed in the middle of each control or data channel to aid the receiver with synchronization and channel estimation. Guard subcarriers are used to suppress the multi-access interference (MAI) from PRs and other CRs. As shown in Figure 2, there is no guard subcarrier between data subchannels 1 and 2, but to protect an ongoing CR transmission, three guard subcarriers should be placed between data subchannels 2 and 3.

1 More data subchannels can be placed, but there is a tradeoff between fine-grain spectrum allocation and packet header overhead.
Consider a CR node \( u \) that is about to transmit a data packet to CR node \( v \) using data subchannels of equal bandwidth \( W \). The maximum channel capacity between CRs \( u \) and \( v \) over the \( i \)th data subchannel, denoted by \( C_{uv}^{(i)} \), is given by Shannon’s equation:

\[
C_{uv}^{(i)} = W \log_2 \left( 1 + \frac{P_R^{(i)}}{N_0 + P_T^{(i)}} \right)
\]  

(1)

where \( N_0 \), \( P_R^{(i)} \), and \( P_T^{(i)} \) are the powers of the Gaussian noise, the desired signal, and the measured interference at receiver \( v \), respectively. We assume that CRs periodically sense the spectrum and learn what portions of it are used by PRs and other CRs. It is also assumed that CRs can distinguish a PR signal from noise using a cyclostationary or a waveform-based sensing technique [15]. Let \( \Psi_{uv} \) be a set of unused subchannels in the vicinity of link \((u,v)\). For a given rate demand \( D \) (in bits/second), we define the weight of link \((u,v)\) as follows:

\[
l_{uv} = (d_{ut} - d_{vt}) \cdot \min \left\{ \sum_{i \in \Psi_{uv}} C_{uv}^{(i)}, D \right\}
\]  

(2)

where \( d_{ct} \) denotes the Euclidean distance between CR node \( v \) and the CR destination \( t \). Each CR node knows its location via GPS or using some localization approach. The link weight represents how much further a message can be forwarded towards the destination per unit time over the given link. In this subsection, we rely on a simplified model (Shannon’s formula) and ignore frequency overhead such as pilot subcarriers and guard subcarriers. However, this simplified model is sufficient to differentiate between links and select the best relay node. In the next subsection, we show how to compute the link weight in a more realistic way.

**B. Protocol Description**

In SABE, CRs periodically scan the available spectrum and estimate the allowable transmission power for each data subchannel. A source or an intermediate node (say CR \( u \)) broadcasts a request-to-forward (RTF) packet using the maximum transmission power \( P_{\text{max}} \) over the control channel. It includes in this RTF packet the set of available data subchannels, the allowable transmission powers, the rate demand \( D \), the location of the broadcasting node, and the location of destination.

Upon receiving an RTF packet, a neighbor (say CR \( v \)) first checks whether it is located in the forwarding area of node \( u \). All CRs within the forwarding area of node \( u \) are candidates. CR \( v \) measures the received signal power \( P_R^{(j)} \) at each control subchannel \( j \) and derives the channel gain \( h \) for that subchannel, as follows:

\[
h_{uv}^{(j)} = \frac{P_R^{(j)}}{P_{\text{max}}}
\]  

(3)

For all \( j \in \Psi_{uv} \), the signal power received over the \( j \)th data subchannel is estimated from the \( i \)th control subchannel, as follows [12]:

\[
P_{R_{\text{min}}}^{(j)} = \frac{P_R^{(j)}}{\left( N_0 + P_T^{(j)} \right)} \text{SNR at node } v \text{ over data subchannel } j.
\]  

(4)

Let \( \mu^{(j)} = \frac{P_R^{(j)}}{\left( N_0 + P_T^{(j)} \right)} \) be the SINR at node \( v \) over data subchannel \( j \). Let \( \eta^{(j)} \) be the SINR threshold that maps to the highest possible transmission rate in the rate-SINR relationship. The minimum transmission power at CR \( u \) over data subchannel \( j \) that meets the SINR threshold at CR \( v \) is:

\[
P_{R_{\text{min}}}^{(j)} = \frac{\eta^{(j)}}{h_{uv}^{(j)}}
\]  

(5)

If \( P_{R_{\text{min}}}^{(j)} > P_{R_{\text{max}}}^{(j)} \) due to interference from other CRs, data subchannel \( j \) is excluded from \( \Psi_{uv} \).

Finally, the link weight \( l \) between CR \( u \) and \( v \) is taken as:

\[
l_{uv} = (d_{ut} - d_{vt}) \cdot \min \left\{ \sum_{j \in \Psi_{uv}} \rho^{(j)}, D \right\}
\]  

(6)

where \( \rho^{(j)} \) is the transmission rate over data subchannel \( j \), obtained from the rate-SINR table. Note that \( \rho^{(j)} \leq W \log_2 \left( 1 + \mu^{(j)} \right) \). If a link has enough idle data subchannels to meet the rate demand, the best subchannels should be selected. We later explain how to select such subchannels.

After obtaining \( l_{uv} \), CR \( v \) sets a delay timer \( \delta \) to reply to the RTF. The value of \( \delta \) is set as follows:

\[
\delta = \delta_{\text{max}} \cdot \left( 1 - \frac{l_{uv}}{\tau \cdot D} \right)
\]  

(7)

where \( \delta_{\text{max}} \) and \( \tau \) denote, respectively, the maximum allowable delay and the maximum transmission range. The value of \( \delta_{\text{max}} \) is set to the minimum contention window size times the slot time. \( \tau \) can be calculated as prior to the path loss formula. The higher the link weight, the smaller is the delay. If
the timer expires, node \( v \) transmits an accept-to-forward (ATF) packet to node \( u \) and includes in this ATF the data subchannel indexes that are to be used for the upcoming transmission, their transmission powers, their transmission rates, and the tolerable interference power of each data subchannel. The last parameter is the minimum additional interference power that the receiver can tolerate (i.e., SINR will not go below the threshold by adding this amount of interference), and is used to calculate the allowable transmission power \( P_{I}^{(j)} \). We will explain how to determine this tolerable interference power in the next subsection.

While waiting for an ATF packet, there are three scenarios that can happen: (1) Candidates exist and all of them satisfy \( D \), (2) candidates exist but some of them do not satisfy \( D \), and (3) no candidates exist. In the first case, the candidate that is closest to the destination sends its ATF first, and is thus selected as the relay node. In the second case, the candidate with the highest link weight is selected as the relay node. For example, in Figure 1, if CR \( v \) has enough available data subchannels to meet the rate demand, it sends ATF first (because \( d_{at} - d_{et} > d_{at} - d_{wt} \)). When the PR activity is high and/or many CR links are active around CR \( v \), CR \( w \) in Figure 1 may be the relay node although it is not the closest neighbor to the destination. The third case occurs when a dead end is encountered (i.e., no neighbor is closer to the destination than the sending node). In this case, the BFP scheme, which was explained earlier, is used. Because BFP does not take into account interference to PRs, we slightly modify it by excluding, from the planar graph, CRs that have no available channels.

If a candidate detects a carrier before its ATF timer expires, it cancels its timer. Because the carrier sensing range is typically larger than the transmission range, it is assumed that all candidates of a given forwarding node can sense the carrier of each other. Upon receiving the ATF packet, the forwarder (CR \( v \)) transmits a decide-to-send (DTS) packet. It is possible that more than one candidate have the same link weight, resulting in multiple ATF transmissions that could potentially collide at node \( u \). So, if a DTS packet does not arrive within a certain amount of time, the ATF packet is retransmitted after a random backoff.

The DTS packet contains the data subchannel indexes and the transmission power of each data subchannel. It is used to inform \( u \)’s neighbors of the upcoming transmission and spectrum allocation. Such information along with the sensing results is used to estimate the interference power \( P_{I}^{(j)} \). After transmitting the DTS packet, CR \( u \) transmits a DATA packet to the relay node (say CR \( v \)) over data subchannels \( j \in \Psi_{uv} \subseteq \Psi_{uv} \), using the transmission power \( P_{T}^{(j)} \) and the transmission rate \( \rho^{(j)} \). We later explain how to obtain \( \Psi_{uv} \). Upon receiving the DATA packet, CR \( v \) replies with an acknowledgment (ACK) packet, sent over the control channel. Note that only the DATA packet is transmitted over the data channel. All control packets (i.e., RTF, ATF, DTS, and ACK) are transmitted over the control channel using the maximum transmission power and the lowest transmission rate. ACK can be delayed until the control channel becomes idle.

C. Concurrent Transmissions via Power Control

As mentioned before, \( \Psi_{uv} \) consists of all data subchannels not used by PRs in the vicinity of both the sender \( u \) and the receiver \( v \), and the allowable transmission power of \( u \) over data subchannel \( j \in \Psi_{uv} \) (\( P_{T}^{(j)} \)) is not less than the lowest transmission power supported by the transceiver. \( P_{I}^{(j)} \) depends on the tolerable interference power, which is indicated in the ATF packet.

We now show how to derive the tolerable interference power. Suppose that CR \( u \) wishes to transmit a data packet to CR \( v \), while in the meantime CR \( w \) is transmitting to CR \( z \). From the SINR definition, the following equation must be met to secure the ongoing transmission from \( w \) to \( z \):

\[
P_{I}^{(j)} \leq \frac{P_{R}^{(j)}}{\mu_{R}} - N_{0}.
\]

(8)

When CR \( u \) overhears an ATF packet from \( z \), it adjusts its allowable transmission power over data subchannel \( j \) so as not to disrupt \( z \)’s reception:

\[
P_{T}^{(j)} = P_{I}^{(j)} \cdot \frac{1}{h(u)} \cdot \epsilon
\]

(9)

where \( \epsilon \) is a deflation factor that scales down the interference effect to CR \( z \), \( 0 < \epsilon < 1 \). If there are several ongoing transmissions in the vicinity, the lowest of the allowable transmission powers is used in (4). Note that SABE can improve spectrum efficiency not only through subchannel allocation, but also by using transmission power control.

D. Data Subchannel Selection

As mentioned before, if a link has enough available data subchannels to meet the rate demand, the best subchannels should be selected. Hereafter, we refer to a data subchannel as a subchannel, for brevity. In WiMAX networks, the access point assigns subchannels to its nodes, so that the sum of the transmission powers of all nodes is minimized [14]. In SABE, because data transmissions are carried out on a per-packer and per-hop basis, optimal subchannel allocation is nontrivial. After deciding the available subchannels and transmission rates (as explained in Figure 3), a candidate selects the fewest subchannels that meet the rate demand.

If only one subchannel is used in a TV channel (e.g., subchannel 1 in Figure 3), guard subcarriers should be placed at the ends of that subchannel to suppress MAI. If more than one contiguous subchannel is selected (e.g., subchannels 7 and 8), guard subcarriers between those subchannels can be used for data transmission. This problem can be formulated as a variant of the Knapsack problem, which is known to be NP-hard. Note that \( \beta \) is the ratio of the number of guard subcarriers and the number of data subcarriers (used to stream data symbols) in that noncontiguous subchannel.
In Figure 3, $\beta$ equals to 3/14. Formally, the data subchannel selection problem can be stated as follows:

$$\text{minimize} \sum_{j \in \Psi_{uv}} X(j)$$

s.t. $$\sum_{j \in \Psi_{uv}} \left( X(j) + \sum_{k \in \Psi_{uv}} Y(j, k) \cdot \frac{\beta}{2} \right) \rho^{(j)} \geq D$$

where $X(j)$ is the indicator function, taking a value of 1 if subchannel $j$ is selected, and $Y(j, k)$ is the indicator function, taking the value 1 if two contiguous subchannels $j$ and $k$ are selected.

We present a simple heuristic algorithm to solve this problem. Let $\Phi$ be a set of subchannel combinations. For each combination, each element has one or more contiguous available subchannels. For example, in Figure 3, $\Phi = \{(1, 2), (1, 3), (2, 4, 5, 6), \ldots\}$ where numbers in parentheses denote subchannel indexes. A candidate sorts $\Phi$ in a decreasing order of the transmission rate. If there are elements whose transmission rates are more than $D$, the smallest of these elements is selected. If not, the candidate selects subchannels from the first element until the cumulative transmission rate exceeds $D$.

IV. EXPERIMENTAL RESULTS

We used NS-3 to simulate SABE and compare its performance with GPSR. In our simulations, 46 CRs are randomly distributed over 1500 m x 1500 m square. They move according to the random way-point model with speed in the range [5, 20] m/s and 10 second pause time. Two source / destination pairs have fixed positions. Each source CR generates a 1000-byte UDP packet every 1 ms. The source CRs are placed at locations (0, 0) and (0, 1500), and the corresponding destinations are placed at (1500, 1500) and (1500, 0), so that two routes cross in the middle. We use the rate-SINR table shown in Table I for data subchannels. The SINR values are taken from the 802.16 standard [14]. The cumulative transmission rate of control subchannels is 4.96 Mbps. The maximum transmission power of CRs is 16 dBm. CRs opportunistically use the spectrum of TV channels 21 to 50 (512 ~ 692 MHz). Each TV channel is used by one PR. PRs behave as an independent ON/OFF source with an activity factor $\alpha$. PR activities are homogeneous. GPSR is modified to allow for adjusting the transmission power and selecting one data channel with the highest transmission rate. In GPSR, every node broadcasts HELLO packet every 1 or 2 second(s).

Figure 2 shows the end-to-end message delivery ratio as a function of node speed with $\alpha = 0.2$. It is observed that SABE losses significantly fewer messages compared with GPSR. The message delivery ratio of GPSR decreases as the node speed increases. The reason is that GPSR forwards a message using the locations of neighbors and these locations are updated through HELLO packets. If the rate of HELLO broadcasts is not high enough relative to node speeds, the sender can transmit a message to a node not in its transmission range anymore. In contrast, in SABE the sender does not need locations of neighbors and the route is decided by receivers with their up-to-date information.

Figure 3 illustrates the end-to-end latency as a function of $\alpha$ when the node speed is 15 m/s. SABE results in a lower latency than GPSR because it uses multiple non-contiguous spectrum holes and adjusts transmission powers to improve spectrum efficiency. The latency in SABE increases monotonically with $\alpha$ but is still much lower than that of GPSR. The latency in GPSR does not depend on the HELLO period or $\alpha$ because GPSR can exploit only one spectrum hole at a time.

The spectrum efficiency of SABE is also validated by the end-to-end goodput performance, shown in Figure 4. GPSR gives lower goodput than SABE because it can use only one contiguous spectrum hole. The goodput of GPSR decreases with the node speed, due to a higher rate of lost messages. From the simulation results, we conclude that SABE is robust to spectrum dynamics and node mobility.

V. CONCLUSION

We presented a novel routing protocol, called SABE, which uses beaconless geographical routing for CRNs. Inherited from the beaconless geographical routing, SABE can efficiently adapt to spectrum dynamics and node mobility. SABE jointly selects routes, transmission powers, and transmission rates to increase spectrum efficiency. Simulation results showed that
SABE provides high delivery ratio and throughput performance even when node locations and spectrum availabilities change frequently.

REFERENCES