QoS-Aware Parallel Sensing/Probing Architecture and Adaptive Cross-Layer Protocol Design for Opportunistic Networks

Mohammad J. Abdel-Rahman, *Member, IEEE*, Harish Kumar Shankar, and Marwan Krunz, *Fellow, IEEE*

***Abstract*—The opening of the ultrahigh frequency (UHF) TV band by the Federal Communications Commission for opportunistic operation promises to relieve the demand on the industrial, scientific, and medical (ISM) bands. However, supporting bandwidth-intensive applications over TV white spaces can be quite challenging, due to the unpredictable nature of spectrum availability and the fluctuations in channel quality. The realization of this Herculean feat through unlicensed usage, while providing protection to licensed primary users, requires intelligent and adaptive protocol design. In this paper, we propose a quality-of-service-aware parallel sensing/probing architecture (QASPA), which exploits inherent channel and user diversity exhibited by the wireless system. Aiming at maximizing the sensing efficiency while achieving high detection accuracy, QASPA incorporates an optimal adaptive double-threshold-based sensing mechanism. It also embodies a cross-layer protocol, which uses an adaptive framing structure to minimize the control overhead, as well as a novel spectrum assignment strategy targeted at improving the spatial reuse of the network. The proposed spectrum assignment strategy supports both channel bonding and aggregation. Our simulations validate the ability of QASPA in guaranteeing the demands of high-bandwidth opportunistic flows while supporting low-bandwidth flows. They also show the superior performance of QASPA compared with the scheme used in the ECMA-392 standard for opportunistic indoor streaming.**

***Index Terms*—Channel allocation, channel probing, multimedia communication, opportunistic access, optimal stopping theory, spectrum sensing.**

I. Introduction

TRADITIONALLY, much of the spectrum is statically licensed for a given use in a given geographic area. Exceptions to this norm include the industrial, scientific, and medical (ISM) bands, which facilitate many indoor and short-range communications (e.g., wireless local area networks, Bluetooth, etc.). However, these bands are reaching their capacity limit as more bandwidth-hungry multimedia traffic is being pushed through them. Dynamic spectrum access tries to address the rising demand by allowing spectrum-agile devices with cognitive radio (CR) capabilities to operate *opportunistically* as secondary users (SUs) over certain licensed bands, including TV white spaces (TVWSs). However, supporting the quality-of-service (QoS) requirements of high-bandwidth applications using the DSA paradigm is a Herculean feat, to say the least. These applications require sustained throughput, which is difficult to guarantee in a DSA environment characterized by spatiotemporal variations in spectrum availability. In fact, the mere presence of spectrum holes that, on average, exhibit low primary user (PU) activity is not enough to enable bandwidth-intensive communications. One also needs to take into account the *quality* of these holes. To minimize the disruption to the high-bandwidth flows, the spectrum sensing process needs to identify *stable* idle channels, i.e., ones that are expected to remain idle for an extended period of time.

The main objective of this paper is to provide a framework for opportunistically transporting heterogeneous traffic, which includes high-bandwidth flows and best-effort flows, over TVWS. To this end, we introduce a QoS-aware parallel sensing/probing architecture called quality-of-service-aware parallel sensing/probing architecture (QASPA), which exhibits five unique features. First, it uses estimated PU activity profiles to construct a schedule for parallel (concurrent) sensing/probing of different channels and for determining the best channel to use for control and management over the next frame. In contrast to [2], in which channel quality is inferred by periodically randomly probing spectral bands without any scheduling, in our scheme, links are scheduled for sensing/probing such that the rate demands of prospective traffic flows are probabilistically guaranteed. Second, QASPA adopts an adaptive time-division multiple-access (TDMA)-based frame structure that attempts to minimize the control overhead and hence maximize the data transmission period. This is in contrast to fixed-frame protocols, such as OP-MAC [3] and the one used in the ECMA-392 standard for opportunistic media streaming [4].

Third, QASPA uses a novel multichannel sensing/probing scheme that exploits the inherent multiuser diversity of a wireless system to maximize the number of discovered opportunities over a given time period. Several schemes have been developed in the literature to improve spectrum utilization while limiting interference onto PUs. These schemes do not exploit multiuser diversity, whereby the quality of a particular

Manuscript received September 1, 2014; revised January 19, 2015; accepted February 16, 2015. This work was supported by the Qatar National Research Fund (a member of Qatar Foundation) under Grant NPRP 4-1034-2-385. An earlier version of this paper was presented at the IEEE International Symposium on Dynamic Spectrum Access Networks, McLean, VA, USA, April 2014. The review of this paper was coordinated by Prof. X. Wang.

M. J. Abdel-Rahman was with the University of Arizona, Tucson, AZ 85721 USA. He is now with the Department of Electrical and Computer Engineering, Virginia Polytechnic Institute and State University, Blacksburg, VA 24061 USA (e-mail: mo7ammad@vt.edu).

H. K. Shankar was with the University of Arizona, Tucson, AZ 85721 USA. He is now with Palo Alto Networks, CA 95054 USA.

M. Krunz is with the Department of Electrical and Computer Engineering, University of Arizona, Tucson, AZ 85721 USA (e-mail: krunz@email.arizona.edu).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TVT.2015.2427803

channel varies from one link to another. Many protocols have been proposed to improve the spatial reuse of a network (e.g., [5]–[7]). However, our problem is further complicated by the need to maximize the network's spatial reuse while guaranteeing the rate demands of high priority (HP) flows and ensuring interference-free PU communications.

Fourth, QASPA adopts an optimal adaptive double-threshold-based spectrum sensing algorithm, which aims at improving the spectrum sensing efficiency while achieving high PU detection accuracy. In contrast to the conventional single-threshold sensing (STS) approach, double-threshold sensing schemes can simultaneously achieve low miss-detection and low false-alarm probabilities within a short sensing time [8]. However, double-threshold sensing involves an *uncertainty region*: If the sensing outcome falls between the two sensing thresholds (i.e., in the uncertainty region), the channel is considered neither busy nor idle [8]. In this paper, we address this problem by adaptively adjusting the channel sensing time, so that the uncertainty region is reduced while, at the same time, the required miss-detection and false-alarm probabilities are met. In contrast to [9], in our protocol, the total sensing time per frame is fixed, and the objective is to maximize the number of discovered opportunities within the allocated sensing time, while maintaining the required detection accuracy. We formulate the optimal adaptive double-threshold-based sensing (DTS) problem using optimal stopping theory [10]. The optimal stopping theory has been previously used in single-threshold-based systems to optimize the *number of sensed channels* [11] and not the *channel sensing time*, as done in this paper. Unlike the cooperative sensing scheme in [12], our scheme is efficient even with a small number of operational SUs. As a result, we do not compromise the sensing accuracy, even when the number of operational SUs is small.

Finally, using the outcome of the channel sensing/probing process, we design a centralized spectrum assignment scheme for QASPA that supports channel bonding and aggregation, and that aims at maximizing the number of concurrently active flows. Rather than focusing on the performance of an individual link, we are interested in optimizing the overall efficiency of the entire opportunistic network.

The remainder of this paper is organized as follows. In Section II, we define our network model. We provide an overview of QASPA in Section III. The adaptive frame structure (AFS) is introduced in Section IV, followed by the proposed parallel sensing/probing design in Section V. In Section VI, we present the channel allocation scheme. We evaluate the proposed design in Section VII. Conclusion are provided in Section VIII.

II. Network Model and Assumptions

We consider an opportunistic CR network (CRN) with a centralized controller, called the *master device* (MD) (see Fig. 1), which plays an analogous role to a wireless access point (AP). In contrast to a Wi-Fi-based AP, data communications (i.e., user payload) in our setup can occur directly between any two nodes (two *slave devices*, or a slave device and the MD), whereas control information (e.g., scheduling information, channel quality reporting, etc.) can be exchanged only between a slave device and the MD. The conventional viewpoint of SU–PU coexistence is preserved, wherein SUs strive to communicate over channels that are not used by PUs. An arbitrary number of peer-to-peer (P2P) links, comprised of CR-enabled Tx–Rx pairs, become associated with the MD through appropriate signaling that occurs before initiating any data communication. Each device is equipped with a half-duplex transceiver, operating at constant transmission power. Hence, a node can only listen to or transmit over one channel at a time. This setup is the same as the one used in the ECMA-392 architecture [4].

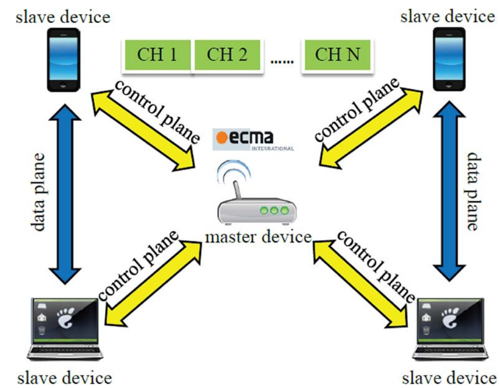

Fig. 1. Network topology used in our design.

We classify the traffic over active P2P links into two classes: bandwidth-intensive HP flows and best-effort low-priority (LP) flows. Specifically, flows with a rate demand greater than R_{th} Mb/s are classified as HP, and all other flows are deemed as LP flows. The flow type is indicated using a single bit in the association packet sent to the MD. Let $\mathcal{L}$ represent the set of HP flows in the network ($|\mathcal{L}| = L$). HP flows are assumed to be long-lived, with a stringent rate demand of $R_d^{(j)}$ for flow $j \in \mathcal{L}$. Once the MD allocates channels for links, data exchanges take place between the respective peer devices without any intervention from the MD. All peer devices are assumed within the communication range of the MD, which is typically the case in indoor environments.

Let $\mathcal{N}$ ($|\mathcal{N}| = N$) denote the set of orthogonal channels in the UHF band that can be used opportunistically. We require that no two links can transmit over the same channel at the same time. The channel quality, obtained through probing, is assumed stationary during the channel's coherence time, denoted by $\tau_c^{(i)}$ for channel $i \in \mathcal{N}$. Channel quality is assumed link dependent. PU activity over channel i is modeled as a continuous-time Markov process, which alternates between busy and idle states, with average busy duration and idle duration of $T_{\text{on}}^{(i)}$ and $T_{\text{off}}^{(i)}$, respectively.

III. Protocol Overview

Our QASPA uses a predefined frame length T_{frame}. Synchronization among various SUs is achieved by disseminating control packets over *a dynamically assigned* control channel (CC), which is determined on the fly by the MD based on the estimated PU channel-usage profile. Similar to ECMA-392, each frame consists of a number of medium access slots

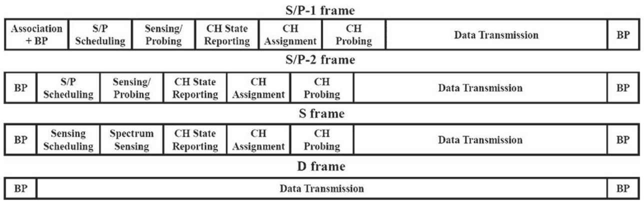

Fig. 2. Frame types used in AFS.

(MASs), reserved for various operations, including network-wide synchronization, parallel sensing/probing, concurrent data communication, etc.

QASPA follows an AFS design, whereby fields used to support various protocol operations may vary from one frame to another, although the overall frame length is still fixed at T_{frame}. This design is intended to eliminate redundant operations, resulting in drastic reduction in the protocol overhead.

Association: In line with [13], we assume the existence of an intelligent AP discovery mechanism for associating prospective P2P links with the MD and establishing a synchronized network. Once associated, links carrying HP flows provide their rate demands to the MD. Rate demands play a key role in the operation of other elements of the protocol.

Beacon Period: Beacons are transmitted by the MD in two instances. First, at the beginning of each frame, the type of the TDMA frame to be used (as determined by the AFS algorithm) is broadcast to all links in the network. Second, toward the end of the frame, the newly designated CC for the following frame period is broadcast to the entire network. Allowing the CC to be adjusted on a per-frame basis gives great flexibility and robustness against fast PU dynamics.

Sensing/Probing Scheduling: The scheduling of concurrent sensing/probing processes forms the crux of the QASPA design. The MD schedules the channels that various links need to sense/probe during the spectrum discovery phase. Such scheduling takes into account the rate demands of HP flows and the estimated PU profiles, and attempts to maximize the number of discovered spectrum opportunities (hence, the number of active links in the network). Unlike the schemes in [2], our sensing/probing scheduling approach takes into account the link-dependent channel quality obtained from previous probing instances. In [2], the quality of one channel over a given link is used to infer the quality of other channels, based on simple path-loss models. This is not accurate for indoor settings.

Reporting: After sensing/probing a specified set of channels, each link reports the PU state of these channels along with their measured qualities (if detected idle) to the MD. This information is used by the MD for channel assignment.

Channel Assignment: The MD strives to maximize the number of admitted flows with satisfied rate demands (HP and LP) by incorporating channel bonding and aggregation techniques. We introduce a second round of probing to further increase the admission probability. The motivation behind this design is to support multiple flows simultaneously to increase the spatial reuse of the network.

Data Transmission: After channel assignment, links communicate in a P2P fashion over the assigned channels for a duration T_{data}, which depends on the frame type.

It is to be instilled that the beacon periods (BPs) at the start and end of a frame are the only two recurring fields in a frame. The occurrence of other operations in the frame is in accordance with the AFS algorithm, which is discussed in Section IV.

Estimation of PU Dynamics: Unpredictable PU dynamics result in intermittent connectivity and high channel switching rates for SUs. The observed correlations of PU activity over TVWS (demonstrated in [14]) permit us to estimate the PU profile based on past observations. This minimizes the time to identify an idle channel Subsequently, incorporating the estimated PU profile while designing spectrum sensing sequences leads to increased discovery of spectrum opportunities [15]. In contrast, random scheduling of sensing events can lead to inefficient sensing as SUs may end up sensing channels that are more likely to be busy. To account for PU dynamics, we employ an exponentially weighted moving average-based estimation approach, wherein the weight given to the recent sample is appropriately adjusted to cope with PU dynamics.

We assume that the MD has an initial estimate of the PU profile over all channels. Subsequent estimates are obtained through sensing. A sliding window of size T_{est} is used to estimate $T_{\mathrm{on}}^{(i)}$ and $T_{\mathrm{off}}^{(i)}$. These estimates, which are denoted by $\hat{T}_{\mathrm{on}}^{(i)}$ and $\hat{T}_{\mathrm{off}}^{(i)}$, are used in computing the probability that channel i is idle at time t, which is denoted by $P_{\mathrm{idle}}^{(i)}(t)$, i.e.,

$$P_{\mathrm{idle}}^{(i)}(t) = \frac{\hat{T}_{\mathrm{off}}^{(i)}}{\hat{T}_{\mathrm{off}}^{(i)} + \hat{T}_{\mathrm{on}}^{(i)}} + \frac{\hat{T}_{\mathrm{on}}^{(i)}}{\hat{T}_{\mathrm{off}}^{(i)} + \hat{T}_{\mathrm{on}}^{(i)}} \; e^{-\left(\frac{1}{\hat{T}_{\mathrm{on}}^{(i)}} + \frac{1}{\hat{T}_{\mathrm{off}}^{(i)}}\right)t}. \quad (1)$$

IV. Adaptive Frame Structure Design

To reduce the control overhead, our protocol adaptively selects one of four frame types: S, S/P-1, S/P-2, and D frames. The formats of these frame types are shown in Fig. 2.

Four parameters determine the decision process for the AFS design: $\hat{T}_{\mathrm{off}}$, τ_c, L, and the arrival rate of new SU connection requests (τ_{new}). To simplify the notation, we drop the superscript i when we are not referring to any specific channel. Without loss of generality, we assume that each link transports only one flow at a time, which can be either HP or LP. The rationale behind the design of AFS is as follows. If $\hat{T}_{\mathrm{off}}^{(i)} > T_{\mathrm{frame}}$ and $P_{\mathrm{idle}}^{(i)}(t) > \beta$, where β is a design parameter, then channel i does not need to be sensed in the current frame, as this channel is expected to remain idle throughout the current frame. Similarly, if $\tau_c^{(i)} > T_{\mathrm{frame}}$, then the quality of channel i is not expected to vary throughout the current frame duration, obviating the need to carry out channel probing. We interrupt the operation of the AFS algorithm every τ_{new} seconds to accommodate new traffic requests and to check for any changes in the rate demands of existing flows. This step is very crucial to cope with the dynamic nature of applications in today's mobile computing platforms. To jointly accommodate the arrival of new flows and changes in the rate demands of existing ones, we use S/P-1 frame, which enforces mandatory channel probing to obtain the link-dependent channel quality information.

To cope with PU dynamics and fluctuating channel quality, we use the S/P-2, S, and D frame types. The quantities whose values are not expected to expire within T_{frame} are termed *valid* entries. We continuously monitor the validity of each entry by using a timer at the MD. Initially, all channels for which $\tau_c \geq T_{\mathrm{frame}}, \hat{T}_{\mathrm{off}} \geq T_{\mathrm{frame}}$, and $P_{\mathrm{idle}}(t) > \beta$ are grouped into a list called List_1. The remaining channels are grouped in List_2. If the channels in List_1 can guarantee the rate demands of all HP flows in $\mathcal{L}$, then we use a D frame as we do not need to sense or probe any additional channels. The computation of the number of satisfied flows over a given channel set requires knowledge of the channel quality, which varies from one link to another within the same network. On the other hand, LP flows can be transported over idle channels whose $\hat{T}_{\mathrm{off}} > T_{\mathrm{frame}}$, $P_{\mathrm{idle}}(t) > \beta$, and $\tau_c \leq T_{\mathrm{frame}}$. This way, we exploit the predicted values of PU usage profiles in minimizing the control overhead.

If the sum of the rate demands of HP flows cannot be satisfied by channels in List_1, we defer from using a D frame. In this case, we first compute the maximum number of HP flows whose demands can be satisfied by the channels in List_1. Denote the set of links that transport these flows by $\mathcal{L}' \subseteq \mathcal{L}$. The objective now is to maximize the number of flows transported by the links in $\mathcal{L} \setminus \mathcal{L}'$ whose rate demands can be met by the channels in List_2. By employing this methodology, we obviate the need to probe the channels in List_1 and meet the rate demands of flows carried by links in $\mathcal{L}'$ without incurring any additional overhead.

To minimize the control overhead, we check the validity of the coherence time τ_c for channels in List_2. Channels in List_2 that have a valid τ_c are grouped into a new list, which is called List_3. We check if the channels in List_3 can satisfy the rate demands of all flows transported by the links in $\mathcal{L} \setminus \mathcal{L}'$ (by solving Problem 1 in Section V-B). If so, we switch to an S frame, as we do not need to probe the channels in List_3 due to the validity of their channel quality information. If the demands of all flows transported over the links in $\mathcal{L} \setminus \mathcal{L}'$ cannot be satisfied, we switch to an S/P-2 frame. Before scheduling the sensing/probing processes, we determine the set of links whose demands can be met using List_3 channels. Denote this set by $\mathcal{L}''$, where $\mathcal{L}'' \subseteq \mathcal{L} \setminus \mathcal{L}'$. This way, we avoid the control overhead incurred by probing the channels whose τ_c is valid throughout T_{frame}. The remaining links in $\mathcal{L} \setminus \{\mathcal{L}', \mathcal{L}''\}$ are considered for scheduling joint sensing and probing processes over the channels in List_2 with $\tau_c < T_{\mathrm{frame}}$ and $\hat{T}_{\mathrm{off}} > T_{\mathrm{frame}}$ (we group these channels into a new list, called $\mathrm{List}_4 \stackrel{\mathrm{def}}{=} \mathrm{List}_2 \setminus \mathrm{List}_3$). If the receiver experiences any interference from a PU, the corresponding channel is reported to the MD during the next BP and not used again for data transmission until it is sensed to be idle, thereby restricting the maximum interference duration to T_{frame}. A pseudocode of the AFS design can be found in our technical report [16] (see Algorithm 1).

Algorithm 1 Optimal ADTS Algorithm

Input: $T_{s,\max}$, M_j, δ, and c
Output: $\tau_{s_k}, k \in \mathcal{M}_j = \{1, 2, \ldots, M_j\}$
1: Set the maximum sensing time per channel to $T_{s,\max}/M_j$ (i.e., $\tau_{s_i,\max} = T_{s,\max}/M_j, \forall\, i \in \mathcal{M}_j$)
2: **for** $i \in \mathcal{M}_j$ **do**
3: Set the sensing time, τ_{s_i}, to $c\ \delta$
4: Compute the corresponding ϵ_l and ϵ_h from (4) and (5)
5: Sense the channel for τ_{s_i} seconds
6: **while** the state of the channel is uncertain and $\tau_{s_i} < \tau_{s_i,\max}$ **do**
7: Determine the optimal sensing decision (stop/continue)
8: **if** the optimal decision is to continue sensing **then**
9: $\tau_{s_i} = \tau_{s_i} + \delta$
10: Go to line 4
11: **else**
12: Go to line 15
13: **end if**
14: **end while**
15: **if** $i < M_j$ **then**
16: $\tau_{s_{i+1},\max} = \tau_{s_{i+1},\max} + (\tau_{s_i,\max} - \tau_{s_i})$
17: **end if**
18: **end for**

Avoiding Stale Entries in the Database: A natural question is what should be done with the links in $\mathcal{L}'$ during the sensing/probing phase of S, S/P-1, and S/P-2 frames, and how can links in $\mathcal{L}' \cup \mathcal{L}''$ be better used during the probing operation of S/P-1 and S/P-2 frames? This issue arises because we wish to maintain synchronization among the associated links in the network and improve the accuracy of estimated PU profiles by maximizing the number of sensing samples collected from a given channel per unit time.

Note that channels without a valid $\hat{T}_{\mathrm{off}}$ entry are not considered for sensing/probing, which creates a "starvation" condition for these channels, i.e., such channels end up having an undetermined PU state, although they could exhibit good quality and/or low PU occupancy. To avoid this situation, we allow links in $\mathcal{L}'$ and links carrying LP flows to sense channels whose PU state is undetermined (List_4) along with channels in List_1 during the sensing phase of S frame. We also allow links in $\mathcal{L}' \cup \mathcal{L}''$ along with links transporting LP flows to sense List_1, List_3, and List_4 channels during the sensing/probing phase of S/P-1 and S/P-2 frames. It is to be noted that no two links are made to sense the same channel at any given point in time, to increase the number of discovered opportunities. The scheduler implemented in QASPA excludes channels in List_1 and links in $\mathcal{L}'$. In turn, links in $\mathcal{L}'$ are made to sense maximum number of channels in List_1 and List_4 during the spectrum discovery period. Moreover, links contained in $\mathcal{L}''$ are made to sense channels in List_3, during the spectrum discovery phase of S/P-1 and S/P-2 frames. After sensing the scheduled channels in List_3, links contained in $\mathcal{L}''$ use the remaining time in the discovery phase (if any) to sense additional channels in List_4.

V. Quality-of-Service-Aware Parallel Sensing/Probing Architecture Design

QASPA encompasses two functional blocks: an adaptive DTS mechanism and a parallel sensing/probing scheduling mechanism.

A. Adaptive Double-Threshold-Based Sensing Algorithm

In QASPA, we resort to a DTS approach instead of the conventional STS (see Fig. 3). As shown here, for the same

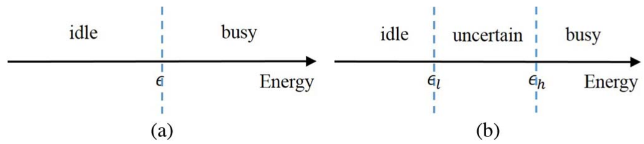

Fig. 3. Single- and double-threshold sensing approaches. (a) STS. (b) DTS.

sensing accuracy, DTS results in several orders of magnitude less sensing time than STS. DTS uses two thresholds (ϵ_l and ϵ_h) in contrast to only one threshold (ϵ) in STS. In DTS, the sensed channel is considered idle if the received energy over this channel is below ϵ_l, busy if the received energy exceeds ϵ_h, and uncertain if the received energy falls between ϵ_l and ϵ_h.

Spectrum sensing accuracy is characterized by the probabilities of miss detection and false alarm, denoted by $P_{\rm md}$ and $P_{\rm fa}$, respectively. For a given sensing threshold ϵ, $P_d \triangleq 1 - P_{\rm md}$ and $P_{\rm fa}$ can be expressed as [17]

$$P_d = Q\left(\left(\frac{\epsilon}{\sigma_n^2} - \gamma - 1\right)\sqrt{\frac{U}{2\gamma+1}}\right) \quad (2)$$

$$P_{\rm fa} = Q\left(\left(\frac{\epsilon}{\sigma_n^2} - 1\right)\sqrt{U}\right) \quad (3)$$

where $Q(\cdot)$ is the complementary distribution function of a standard Gaussian random variable, γ is the SNR of the received PU signal, U is the number of sensed samples ($U = \tau_s f_s$, where f_s is the sampling frequency), and σ_n^2 is the variance of the additive white Gaussian noise.

To simultaneously achieve low $P_{\rm md}$ and low $P_{\rm fa}$, the STS approach requires a considerably large sensing time (denoted by τ_s) [17]. Instead, DTS can simultaneously achieve low $P_{\rm md}$ and low $P_{\rm fa}$ using a much smaller sensing time. In DTS, ϵ_l is selected such that the required $P_{\rm md}$ (which is equal to $1 - \bar{P}_d$) is satisfied. On the other hand, ϵ_h is selected to satisfy the required $P_{\rm fa}$ (which is equal to $\bar{P}_{\rm fa}$). The relations between ϵ_l and $\bar{P}_d$ and ϵ_h and $\bar{P}_{\rm fa}$ are given by

$$\epsilon_l\left(\tau_s, \bar{P}_d\right) = \sigma_n^2\left[\sqrt{\frac{2\gamma+1}{U}}Q^{-1}\left(\bar{P}_d\right) + \gamma + 1\right] \quad (4)$$

$$\epsilon_h\left(\tau_s, \bar{P}_{\rm fa}\right) = \sigma_n^2\left[\sqrt{\frac{1}{U}}Q^{-1}\left(\bar{P}_{\rm fa}\right) + 1\right]. \quad (5)$$

The required sensing time to satisfy $\bar{P}_d$ and $\bar{P}_{\rm fa}$ for STS and DTS schemes, which are denoted by $\tau_{s,\rm single}$ and $\tau_{s,\rm double}$, respectively, can be easily expressed as [17]

$$\tau_{s,\rm single} = \frac{1}{\gamma^2 f_s}\left[Q^{-1}\left(\bar{P}_{\rm fa}\right) - Q^{-1}\left(\bar{P}_d\right)\sqrt{2\gamma+1}\right]^2 \quad (6)$$

$$\tau_{s,\rm double} = \frac{1}{f_s}\left[\frac{\epsilon_h\sqrt{2\gamma+1}Q^{-1}\left(\bar{P}_d\right) - \epsilon_l Q^{-1}\left(\bar{P}_{\rm fa}\right)}{\epsilon_l - \epsilon_h(\gamma+1)}\right]^2. \quad (7)$$

Let ρ denote the probability of uncertainty (i.e., the probability that the received energy over a given channel falls between ϵ_l and ϵ_h). Fig. 4 compares $\tau_{s,\rm single}$ and $\tau_{s,\rm double}$ for different $\bar{P}_d$ and $\bar{P}_{\rm fa}$ values. It shows that, for a given $\bar{P}_d$ and $\bar{P}_{\rm fa}$ and with a small value of ρ, $\tau_{s,\rm double}$ is three to four orders of magnitude less than $\tau_{s,\rm single}$.

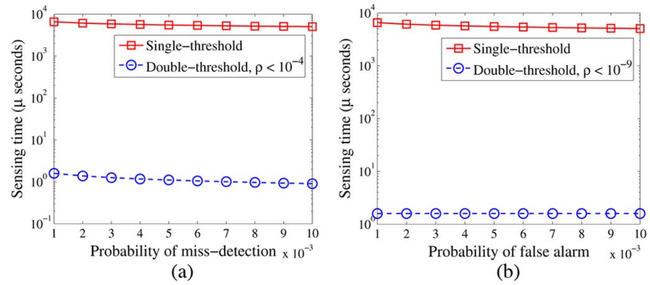

Fig. 4. Sensing time versus (a) $P_{\rm md}$ and (b) $P_{\rm fa}$ ($P_{\rm idle} = 0.5$, $f_s = 6$ MHz, and $\gamma = -15$ dB). (a) $P_{\rm fa} = 0.001$. (b) $P_d = 0.999$.

Although DTS exhibits a smaller sensing time than STS, it faces the problem of the uncertain region. DTS can certainly determine the state of the channel only if its received energy is below ϵ_l or above ϵ_h. Otherwise, the state of the channel cannot be determined with certainty. To overcome this problem, we propose an adaptive DTS-based (ADTS) algorithm.

The idea behind the ADTS algorithm comes from (4) and (5). For a fixed $\bar{P}_d$ and $\bar{P}_{\rm fa}$, if the sensing time increases (hence, $U = \tau_s f_s$ increases) ϵ_l increases and ϵ_h decreases, which causes ρ to decrease. Therefore, starting from a reasonable value of U, if the sensing outcome is uncertain, gradually increasing U may eventually result in having a certain outcome. However, since the total sensing time in each frame cannot exceed $T_{s,\max}$ (obtained from the sensing–throughput tradeoff; see, e.g., [17]), and each link is required to sense a certain number of channels (M_j for link $j \in \mathcal{L}$), the sensing time of a channel over link j cannot exceed $T_{s,\max}/M_j$. Instead of increasing U, it might be better to stop sensing the current channel and use the saved sensing time for sensing the next channel(s).

We formulate the optimal ADTS problem as an optimal stopping rule with horizon $a \stackrel{\rm def}{=} \tau_{s,\max}/\delta$, where $\tau_{s,\max}$ is the maximum allocated sensing time for the given channel, and δ is the increment of the sensing time. An optimal stopping problem with horizon a is defined by two components [10]:

- a sequence of random variables, $X_1, X_2, \ldots, X_a$, whose joint distribution is assumed to be known;
- a sequence of real-valued reward functions, $y_0, y_1(x_1)$, $y_2(x_1, x_2), \ldots, y_a(x_1, x_2, \ldots, x_a)$.

In our formulation, $X_i, i = 1, \ldots, a$, is defined as the state of the sensed channel at the end of the ith sensing slot. X_i equals 1 if the state is idle, -1 if it is busy, and 0 if it is uncertain. The distribution of X_i is given by $P_{1,i} \stackrel{\rm def}{=} \Pr\{X_i = 1\} = \Pr\{E_i \leq \epsilon_{l,i}\}$, $P_{-1,i} \stackrel{\rm def}{=} \Pr\{X_i = -1\} = \Pr\{E_i \geq \epsilon_{h,i}\}$, and $P_{0,i} \stackrel{\rm def}{=} \Pr\{X_i = 0\} = \Pr\{\epsilon_{l,i} < E_i < \epsilon_{h,i}\}$, where E_i is the output of the energy detector at the end of the ith slot, and $\epsilon_{l,i}$ and $\epsilon_{h,i}$ are the values of ϵ_l and ϵ_h at the ith slot, respectively. For any i and j such that $i > j$, $\epsilon_{l,i} > \epsilon_{l,j}$ and $\epsilon_{h,i} < \epsilon_{h,j}$. E_i is assumed to follow a central (noncentral) chi-square distribution in the absence (presence) of the PU signal, with U_i degrees of freedom [18].

Consider the jth sensed channel in a frame. The reward of stopping at the end of the ith sensing slot, i.e., $y_i(x_1,\ldots,x_i)$, depends only on x_i and is defined as follows. Irrespective of the value of x_i, stopping to sense at the end of the ith slot will save $(a-i)\delta$ seconds, which can then be used in sensing the $j+1$th channel in the frame. If $x_i=0$, stopping sensing will waste $i\delta$ seconds consumed in sensing the jth channel without concluding its state. For the jth sensed channel, we associate a quality metric, which is denoted by q_j. Channels are sensed based on their qualities such that the best channel is sensed first (i.e., $q_j > q_{j+1}\ \forall\, j$). q_j is defined as a linear combination of $P_{\rm idle}$ and the normalized expected rate of the channel. Therefore, $y_i(x_1,\ldots,x_i)=y_i(x_i)$ can be expressed as

$$y_i(x_1,\ldots,x_i) = \begin{cases} (a-i)\delta q_{j+1}, & \text{if } x_i=\pm 1 \\ (a-i)\delta q_{j+1} - i\delta q_j, & \text{if } x_i = 0 \end{cases} \tag{8}$$

where $y_0 \stackrel{\rm def}{=} 0$. Recall that our objective is to obtain the state of the sensed channel while satisfying $\bar{P}_d$ and $\bar{P}_{\rm fa}$. Hence, if $x_i=\pm 1$, we stop sensing. The optimal decision to stop or continue sensing is only taken when $x_i=0$. When $x_i=0$, stopping the sensing process will cause losing $i\delta q_j$, whereas continuing to sense will either result in returning $i\delta q_j$ but losing δq_{j+1} (if $x_{i+1}=\pm 1$) or losing an additional δq_j (if $x_{i+1}=0$). Let $V_i^{(a)}(x_1,\ldots,x_i)$ be the maximum return that can be obtained starting from stage i, having observed $X_1=x_1,\ldots,X_i=x_i$. Then, $V_i^{(a)}(x_1,\ldots,x_i)=V_i^{(a)}(x_i)=\max\{y_i(x_i),\mathbb{E}[V_{i+1}^{(a)}(X_{i+1})|X_1=x_1,\ldots,X_i=x_i]\}$. Therefore, it is optimal to stop sensing at the end of the ith slot if $V_i^{(a)}(x_i)=y_i(x_i)$; otherwise, it is optimal to continue sensing. Note that $V_a^{(a)}(x_a)=y_a(x_a)$. The value of the stopping rule problem is given by $V_0^{(a)}$. Accordingly, the optimal stopping sensing problem can be written as follows.

Problem 1:

$$V_0^{(a)} = \underset{j\in\{1,2,\ldots,a\}}{\text{maximize}}\ \mathbb{E}\left[V_1^{(a)}(X_1)\right] \tag{9}$$

$$\text{subject to } j \le (1-|x_i|)\,a + |x_i| i \ \ \forall\, i\in\{1,2,\ldots,a\}. \tag{10}$$

Recall that $V_0^{(a)}$ is given by $\max\{y_0,\mathbb{E}[V_1^{(a)}(X_1)]\}$ and $y_0=0$. Constraint (10) ensures that a link will stop sensing a given channel once the sensing outcome of that channel is certain (i.e., $x_i=\pm 1$). Therefore, if $x_i=\pm 1$, then the stopping time j needs to be less than or equal to i. If $x_i=0$, constraint (10) reduces to $j\le a$, which is redundant.

Because we need to stop sensing at slot a, using the given equation of $V_i^{(a)}(x_1,\ldots,x_i)$, we can obtain the optimal rule at slot $a-1$. Then, knowing the optimal rule at slot $a-1$, we can derive the optimal rule at slot $a-2$, and so on, back to slot 1. A more detailed description of the backward induction procedure can be found in [16].

Next, we describe our proposed algorithm for adaptively adjusting the sensing time and the thresholds ϵ_l and ϵ_h. Consider the jth link. The maximum sensing time allocated for each channel is set initially to $T_{s,\max}/M_j$. Our adaptive sensing algorithm takes M_j rounds. During each round, only one channel is sensed (we index the channel sensed during the kth round by k and denote its sensing time by τ_{s_k}). In each round, after sensing the corresponding channel, the maximum sensing time allocated for the next channel is updated. We denote the maximum sensing time allocated for channel i during the kth round by $\tau_{s_i,\max}^{(k)}$. A pseudocode of our optimal adaptive sensing mechanism (as executed by link j) is shown in Algorithm 1. In [16], we propose another ADTS scheme, which we refer to as optimal ADTS with enhanced accuracy (ADTSEA). The objective of ADTSEA is to maximize the total number of sensed channels that result in a certain outcome and minimize $P_{\rm md}$ and $P_{\rm fa}$.

B. Scheduling Sensing/Probing Processes for HP Flows

The sensing/probing processes for links transporting HP flows are scheduled while probabilistically guaranteeing that the aggregate bandwidth of the discovered opportunities by each link j is equal to $R_d^{(j)}+\kappa$, where $\kappa < R_d^{(j)}$. The rationale behind this approach is twofold: to account for the impact of fluctuations in channel quality and to ensure serving the links that carry LP flows in the network using the excess discovered opportunities. We assume that channel quality does not vary drastically in a short interval of time. Hence, for any link j with rate demand $R_d^{(j)}$, we choose κ such that $\kappa < R_d^{(j)}$.

We formulate our scheduling problem as a constrained optimization problem whose objective is to maximize the number of HP flows with satisfied rate demands. Let $\tilde{\mathcal{N}}$ ($|\tilde{\mathcal{N}}|=\tilde{N}$) be the set of channels that are considered for sensing/probing scheduling, and let $\tilde{\mathcal{L}}$ ($|\tilde{\mathcal{L}}|=\tilde{L}$) be the set of links that will participate in the sensing/probing process. Let $y_i^{(j)}, i\in\tilde{\mathcal{N}}, j\in\tilde{\mathcal{L}}$, be a binary variable that equals 1 if channel i is scheduled to be sensed/probed by link j, and 0 otherwise. Let $R_i^{(j)}$ be the rate supported by channel i over link j and $\Theta_i^{(j)} \stackrel{\rm def}{=} y_i^{(j)} R_i^{(j)} P_{\rm idle}^{(i)}(t)$. The optimization problem can be formulated as follows.

Problem 2:

$$\underset{\substack{y_i^{(j)}\\ i\in\tilde{\mathcal{N}},j\in\tilde{\mathcal{L}}}}{\text{maximize}} \left\{ \sum_{j\in\tilde{\mathcal{L}}} \mathbf{1}_{\left\{\sum\limits_{i\in\tilde{\mathcal{N}}}\Theta_i^{(j)} > R_d^{(j)}+\kappa\right\}} + \frac{\sum_{\substack{i\in\tilde{\mathcal{N}}\\ j\in\tilde{\mathcal{L}}}} y_i^{(j)} R_i^{(j)}}{\sum_{\substack{i\in\tilde{\mathcal{N}}\\ j\in\tilde{\mathcal{L}}}} y_i^{(j)} R_{\max}} \right\} \tag{11}$$

$$\text{subject to } \sum_{j\in\tilde{\mathcal{L}}} y_i^{(j)} \le 1 \ \ \forall\, i\in\tilde{\mathcal{N}} \tag{12}$$

$$\sum_{i\in\tilde{\mathcal{N}}} y_i^{(j)} \le M_j \ \ \forall\, j\in\tilde{\mathcal{L}} \tag{13}$$

where $R_{\max}$ is the maximum supported data rate by any channel $i\in\tilde{\mathcal{N}}$, and $\mathbf{1}_{\{.\}}$ is the indicator function. The first term in the objective function reflects the desire to maximize the number of satisfied HP flows through parallel sensing/probing. The second term is intended to resolve the tie in the case of multiple optimal solutions. The second term always has a value ≤ 1. Constraint (12) ensures that no channel can be sensed/probed by more than one link in a given frame. Constraint (13) restricts the maximum number of channels that can be sensed/probed by link j to a predefined value M_j. Inoperable links whose requested rate demands cannot be met are made to sense the channels whose PU state is undetermined to prevent the occurrence of a "starvation" condition. Constraints (12) and (13) can be written in matrix form as $A\vec{y} \leq \vec{b}$, where $A = (a_{ij})_{1\leq i\leq \tilde{N}+\tilde{L}, 1\leq j\leq \tilde{N}\tilde{L}}$, $\vec{y} = (y_i^{(j)})_{i\in\tilde{\mathcal{N}}, j\in\tilde{\mathcal{L}}}$, and $\vec{b} = (1,\ldots,1,M_1,\ldots,M_{\tilde{L}})^T \in \mathbb{R}_{\tilde{N}+\tilde{L}}$.

Proposition 1: Matrix A in Problem 2 is totally unimodular (TU) [19]. This means that the decision variables $y_i^{(j)}$ can be relaxed to continuous variables, and the resulting solution to Problem 2 is still optimal.

Proof: It can be seen that 1) $a_{ij} \in \{0,1\}$, 2) each column in A contains at most two nonzero coefficients (i.e., $\sum_{i=1}^{\tilde{N}+\tilde{L}} |a_{ij}| \leq 2$), and 3) there exists a partition ($M_1 = \{1,\ldots,\tilde{N}\}, M_2 = \{\tilde{N}+1,\ldots,\tilde{N}+\tilde{L}\}$) of rows such that each column j containing two nonzero coefficients satisfies $\sum_{i\in M_1} a_{ij} - \sum_{i\in M_2} a_{ij} = 0$. Since the given three conditions are satisfied, matrix A is TU [19]. ■

Additional Constraints: In channel aggregation systems, a device cannot aggregate two channels if the frequency separation between these channels exceeds a certain threshold. In here, we modify the sensing/probing scheduling scheme formulated in Problem 2 to account for this additional hardware constraint. The channel aggregation constraint can be stated as follows: $\forall\, l,k \in \tilde{\mathcal{N}}$ and $\forall\, j \in \tilde{\mathcal{L}}$, if $|l-k| > \mathcal{B}$, then $y_l^{(j)} + y_k^{(j)} \leq 1$, where $\mathcal{B}$ is the maximum allowable separation between the aggregated frequency channels. This constraint is reformulated as a set of linear constraints in [16]. This scheduling scheme is evaluated in Section VII.

VI. Allocation of Data Channels

Here, we present an efficient data channel assignment algorithm, which incorporates channel aggregation and bonding techniques. Our objective is to minimize the number of channels allocated to any HP link, in an attempt to maximize the number of admitted flows in the network.

A. Channel Assignment Algorithm

Several wireless protocol designs have been proposed to improve the spatial reuse of the network (e.g., [5]–[7]). In addition to maximizing the network's spatial reuse, our problem is further complicated by the need to meet the rate demands of the HP flows and guarantee interference-free communication for PUs.

Given the sensing/probing outcomes provided by active links, the MD initially computes the feasibility of supporting the rate demands of each HP flow using the channels, which were sensed to be idle by the corresponding link. Let $\mathcal{K}_j$ ($|\mathcal{K}_j| = K_j$) be the set of idle channels discovered by link j. The MD checks if $\sum_{i=1}^{K_j} R_i^{(j)} > R_d^{(j)}$. If this condition is satisfied, the MD executes Problem 3 to compute the optimal number of channels needed to support the rate demand on that link. The links that do not meet the earlier condition are classified as unsatisfied links. Let $x_i^{(j)}, i \in \mathcal{K}_j$ be a binary variable, which equals 1 if channel i is assigned to link j, and 0 otherwise. Our optimization problem is stated as follows.

Problem 3:

$$\underset{x_i^{(j)}, i\in\mathcal{K}_j}{\text{minimize}} \left\{ \sum_{i\in\mathcal{K}_j} x_i^{(j)} + \frac{\sum_{i\in\mathcal{K}_j} x_i^{(j)} R_i^{(j)}}{\sum_{i\in\mathcal{K}_j} x_i^{(j)} R_{\max}} \right\} \tag{14}$$

$$\text{subject to } \sum_{i\in\mathcal{K}_j} x_i^{(j)} R_i^{(j)} \geq R_d^{(j)}. \tag{15}$$

Problem 3 is a nonlinear binary program. The objective function aims at minimizing the number of channels to be allocated for a given link while meeting the requested rate demand. The second term in the objective function is used to break the ties among multiple optimal solutions. When multiple optimal solutions exist, our formulation ensures the selection of channels with minimum aggregated data rate, such that the remaining channel(s) can be used during the second round of channel probing to support unsatisfied links.

Additional Constraints: We also modify the channel assignment scheme in Problem 3 to account for two additional constraints: 1) A link cannot aggregate two channels if the frequency separation between these channels exceeds $\mathcal{B}$, and 2) it cannot bond/aggregate more than $\mathcal{C}$ channels.

B. Second Round of Channel Probing

To meet the rate demand of an unsatisfied link, we first compute the maximum data rate that can be supported by its channel set. Consequently, we schedule this link for a second round of probing over the channels that have been deemed excess by the satisfied links in the network. More specifically, we allocate a total of three MASs before the start of a data transmission in S, S/P-1, and S/P-2 frames for the second-round probing, wherein a maximum of one channel can be probed by any unsatisfied link in the network. If the rate demand of the HP flow can be satisfied by combining the channels in its own channel set and the newly probed channel, the receiver sends a positive feedback to the transmitter, after which the transmitter starts sending data by aggregating/bonding these channels.

The drastic increase in the number of discovered opportunities brought in by QASPA always guarantees fair share of bandwidth for the LP flows. Note that LP flows have no stringent bandwidth requirement, which permits us to assign channels to these flows, regardless of their exhibited quality. The idle channels that are not assigned to HP flows are used by LP flows.

As an example of the second-round probing, consider two links l_1 and l_2 with rate demands of 10 and 7 Mb/s, respectively. Assume that the amount of bandwidth discovered by these two links during their sensing/probing phase are 7 and 5 Mb/s, respectively. Hence, their rate demands cannot be met using the discovered opportunities, making these links inoperable. Assume that h_5 and h_7 are the channels that have been deemed excess by the satisfied links in the network. We randomly assign h_5 to l_1 and h_7 to l_2 for probing. We allocate channels in a random fashion because we do not know the link-dependent channel quality information ahead of time. Assume that h_5 can support 4 Mb/s over l_1 and h_7 can support 2 Mb/s over l_2. Hence, we are able to satisfy the rate demand of l_1, whereas the rate demand of l_2 remains unsatisfied. This way, we are able to increase the number of satisfied links at the cost of a second round of probing.

Common Control Channel Selection: We exploit the centralized architecture to maintain a time-varying common CC (CCC) using estimated PU profiles. The selection of a CCC depends on its PU state rather than its quality as control packets are transmitted at the lowest rate. In the current frame, let $\mathcal{H}$ be the set of channels whose $\hat{T}_{\text{off}} > T_{\text{frame}} + \mu$, where μ ($< T_{\text{frame}}$) is a small quantity that accounts for discrepancy in the estimated values of T_{off}. Within $\mathcal{H}$, the idle channel with the highest probability of remaining idle during the next frame is selected as a CCC. The selected CCC is broadcast by the MD during the BP at the end of the current frame.

Adaptive Reporting Strategy: Motivated by the high switching speeds of the analog-to-digital (A/D) converters deployed in software defined radio front ends, we propose an adaptive reporting strategy wherein the time allocated to the reporting phase T_{rep} is decided dynamically based on the number of channels scheduled for sensing/probing. Assume that r channels were scheduled by the MD. We equally divide T_{rep} into r subslots, one for each channel. In addition, the MD broadcasts a hopping sequence to rendezvous with the devices to gather the report packets. This hopping sequence information is encapsulated in the sensing/probing scheduling packet. Assume that the hopping sequence is $C_1, C_2, \ldots, C_r$. The MD tunes to channel C_i during the ith subslot. Suppose that channel i is scheduled for sensing/probing by link j. If channel i is sensed to be idle, it will be probed. The destination device of link j will then send a control packet to the MD, containing the PU state of the channel along with its quality (in case of successful probe packet exchange) during the ith subslot. Note that, if a channel was sensed to be busy by link j, no control packet is sent during the ith subslot. Consequently, the MD interprets this channel to be occupied and excludes it when assigning channels for data communication. The major accomplishment in this reporting strategy is the elimination of channel contention delay, which might hinder the timely exchange of vital control information. The MD waits for $t_{\text{rep}} \stackrel{\text{def}}{=} T_{\text{rep}}/r$ at each subslot for hearing any control packets from the associated SUs. If this time expires, MD moves to the next subslot in T_{rep}.

VII. Performance Evaluation

A. Evaluation Setup

We consider an area of 50×50 m^2 and 40 channels in the UHF band, each of 6-MHz bandwidth. Each channel can support one of five rates: 2, 4, 8, 12, and 16 Mb/s, each with a probability of 0.2. This randomness in the channel data rate captures wireless phenomena such as fading, shadowing, and RF interference. We use the approximation in [20] to capture τ_c in our simulations. Specifically, $\tau_c(t) = 9\lambda/(16\pi v(t))$, where $v(t)$ represents the velocity of the transmitter toward the intended receiver along the line of sight, and $\lambda = c/f_c$ represents the wavelength of the signal corresponding to carrier frequency f_c. We vary τ_c uniformly between 1 and 10 s. $T_{\text{on}} = T_{\text{off}} = 5$ s.

As elucidated earlier, we resort to a frame-based time-slotted protocol. The time slots can be described at the medium-access level using MASs, with the duration of each MAS defined to be 1 ms in our setup. The fixed number of MASs allocated for each operation in our TDMA frame is shown in Table I. $T_{\text{frame}} =$ 250 MASs. Each link carrying HP flow is randomly assigned a rate demand between 5 and 15 Mb/s. The rate demands of the HP flows are updated once they complete transferring their video flows of predefined duration, as dictated by video trace files. P_d and P_{fa} are set to 0.9 and 0.25, respectively. β in Section IV is set to 0.75. All LP flows are composed of constant bit rate (CBR) traffic with constant data-packet size of 1 KB. We use a 500-B control packet. All simulations are carried out for 10 000 frames using CSIM (a C-based process-oriented discrete-event simulation package), and reported values are the average of 20 runs. The optimization problems are solved using MATLAB. Our simulations are conducted to evaluate the performance gain resulting from the interactions between the various proposed components: QASPA, AFS, and the adaptive channel assignment strategy.

TABLE I
Durations of Various Operations in the TDMA Frame

Operation	Duration (MAS)	Operation	Duration (MAS)
Association	5	Channel Reporting	25
Beacon Period	2	Channel Assignment	2
S/P Scheduling	2	Second-round Probing	3

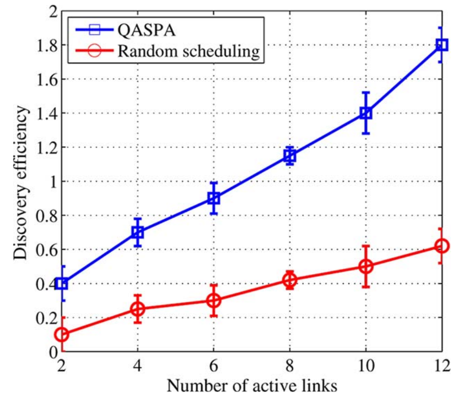

Fig. 5. Discovery efficiency versus active links.

B. Evaluation of QASPA

1) Impact of QASPA on Discovery Efficiency: We study the discovery efficiency of QASPA, defined as the number of idle channels discovered per unit time, and contrast it with a random scheduling scheme that does not make use of the estimated PU profiles and prior channel quality information when scheduling links for sensing/probing. This random scheduling is very similar to the scheme in [2]. Note that we allow unique sets of channels to be sensed/probed by individual links even under the random scheduling scheme to prevent probe packet collisions during the spectrum discovery phase. $T_{s,\max} = 20$ ms and $\kappa = 2$ Mb/s. $R_d = 3$ Mb/s for all HP flows. Fig. 5 shows that the number of operational links has a profound impact on the discovery efficiency. Operating sensing/probing processes in parallel boosts the discovery efficiency significantly. In contrast, random scheduling, despite being facilitated with parallel

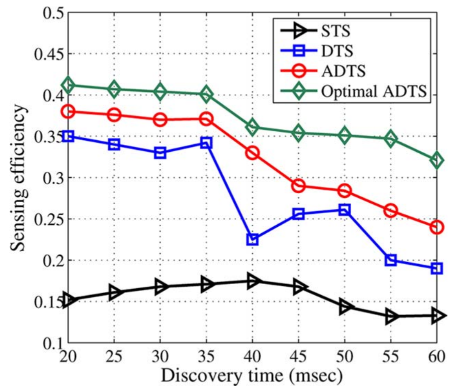

Fig. 6. Sensing efficiency versus discovery time.

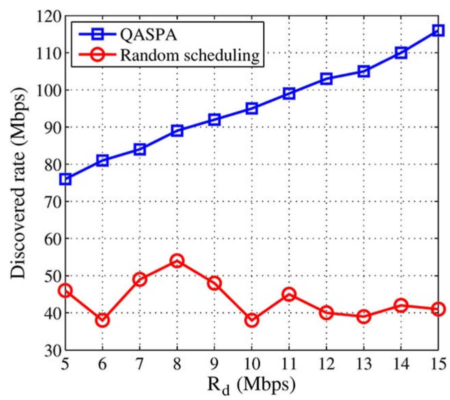

Fig. 7. Bandwidth discovered versus the rate demand.

sensing/probing, fails to match the efficiency of QASPA due to its lack of knowledge on PU profiles.

It is to be noted that the increase in discovery efficiency is also heavily attributed to the optimal ADTS scheme. To verify this, we compare the discovery efficiency of STS, DTS, and the proposed ADTS and optimal ADTS schemes for a single-link scenario under various values of $T_{s,\max}$. The DTS scheme uses a fixed per-channel sensing duration. The uncertain region in DTS prevents us from accurately determining the PU state of the channel, and as a conservative approach, we deem these channels to be busy. This results in lower discovery efficiency as several channels, which might be idle, are deemed occupied. We break this norm in our ADTS scheme, wherein we use an adaptive per-channel sensing duration, which helps us to accurately identify the PU state of the channel, thereby leading to improved sensing efficiency (defined as the ratio between the number of idle channels and the total number of channels sensed during $T_{s,\max}$), as shown in Fig. 6. Our optimal ADTS scheme further improves the sensing efficiency by optimizing the sensing time. In contrast to STS, ADTS, and optimal ADTS, the sensing efficiency of DTS changes abruptly with the discovery time. This happens because of the uncertainty region. With slight changes in the discovery time, several channels that had "certain" states might fall in the uncertainty region, and *vice versa*. Although the uncertainty region also exists in the ADTS and "optimal ADTS" schemes, the adaptability of the maximum per-channel sensing time in these schemes makes them immune to such abrupt changes. In our technical

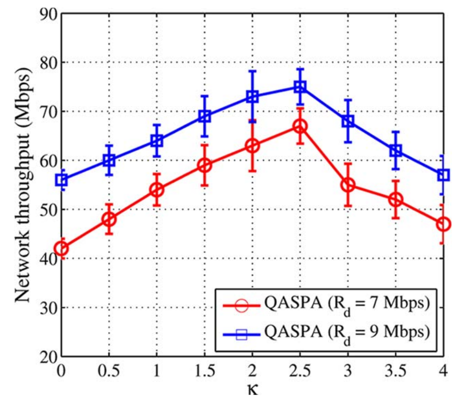

Fig. 8. Network throughput versus κ (10 HP and 5 LP flows; $T_{s,\max} =$ 20 ms).

report [16], we provide a more detailed comparison between the various sensing schemes.

Next, we compute the bandwidth discovered by QASPA to illustrate the impact of considering the channel quality in the sensing/probing scheduling. In this experiment, we fix the number of HP flows to 10 and vary R_d of each HP flow from 5 to 15 Mb/s. Fig. 7 shows that the bandwidth discovered by random scheduling is much less than that of QASPA.

2) Impact of κ on QASPA: κ plays a very important role in mobile wireless networks, wherein the quality of the channel changes frequently due to node mobility. To evaluate the impact of κ on network throughput, we consider 10 HP and 5 LP flows. $T_{s,\max}$ is set to 20 ms. Fig. 8 shows the throughput versus κ for two different rate demands. It is clear that the network throughput exhibits an upward trend for increasing κ until it reaches a breaking point, beyond which, it starts decreasing.

Increasing κ beyond a certain limit prevents the scheduler from scheduling channels for the links to discover the high bandwidth requirement of $R_d + \kappa$ in a restricted spectrum discovery duration, leading to the links being dropped from data communication. In turn, these links sense unique channels during the spectrum discovery phase to improve the accuracy of the estimated PU usage profiles. Subsequently, these dropped links get associated during the occurrence of the next S/P-1 frame. This experiment illustrates the need to choose an optimal value of κ during the operation of the protocol. Based on our experiment, we recommend κ to be $\sim$2.5 Mb/s to overcome the impact of channel quality fluctuation in a relatively low mobile environment. It is interesting to note that positive values of κ might not be suitable in all cases as the channel quality can improve from the previously measured instant, in which case we may choose negative values for κ. However, our protocol is designed in a way that the excess opportunities can be used to serve unsatisfied HP flows via the second-round probing process.

C. Evaluation of AFS

1) Reduction in Control Overhead: Control overhead is the ratio of control packets to the total network packets exchanged over the entire simulation time. In our protocol, fast PU dynamics and low τ_c induce more control packets. To evaluate the

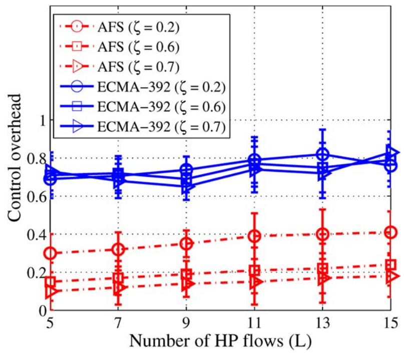

Fig. 9. Control overhead versus L ($\tau_c = 5$ s, $R_d = 7$ Mb/s, and $\kappa = 2$ Mb/s).

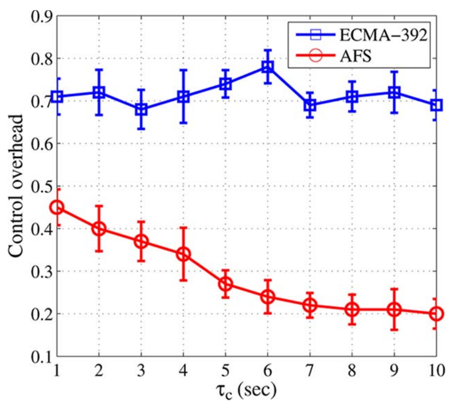

Fig. 10. Control overhead versus τ_c (10 HP flows, $R_d = 7$ Mb/s, $\kappa = 2$ Mb/s, $\zeta = 0.5$).

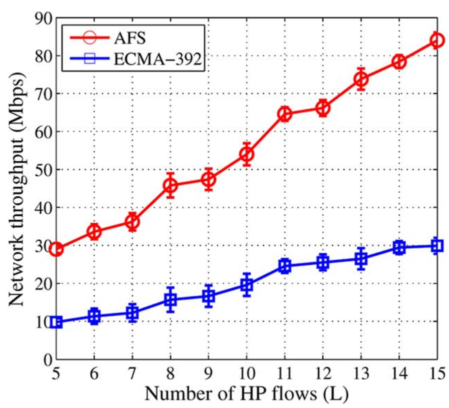

Fig. 11. Network throughput versus L $\tau_c = 5$ s ($R_d = 7$ Mb/s, and $\kappa = 2$ Mb/s).

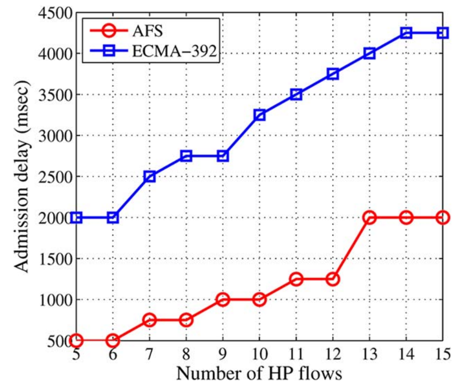

Fig. 12. Admission delay versus L ($\tau_c = 5$ s, $R_d = 9$ Mb/s, and $\kappa = 2$ Mb/s).

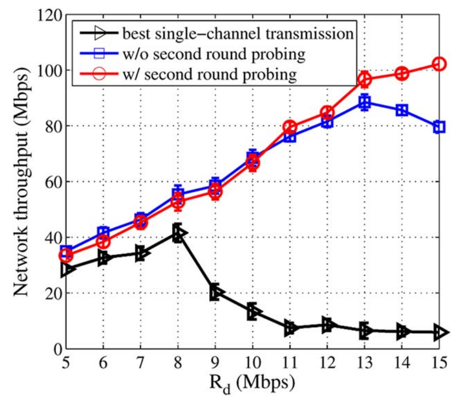

Fig. 13. Network throughput versus R_d (10 HP and 5 LP flows and $\zeta = 0.5$).

reduction in control overhead achieved by AFS, we conduct two experiments. First, we steadily increase the channel vacancy factor $\zeta \overset{\text{def}}{=} \hat{T}_{\text{off}}/(\hat{T}_{\text{on}} + \hat{T}_{\text{off}})$ over all channels in the system while fixing τ_c at 5 s. The rate demands of all HP flows is set to 7 Mb/s with κ set to 2 Mb/s. Fig. 9 shows the reduction in control overhead achieved by AFS over the ECMA-392, which uses a fixed framing structure. Note that the fixed framing structure used in our comparison is comparable with the OP-MAC protocol proposed in [3]. Increasing the number of HP flows increases the exchange of control packets in the network, as shown in Fig. 9. The adaptive nature of AFS helps in reducing the control overhead by $\sim 126\%$ when compared with the ECMA-392 frame design.

Next, we study the impact of τ_c on the control overhead by simulating 10 HP flows, each with $R_d = 7$ Mb/s and $\kappa = 2$ Mb/s. ζ is set as 0.5. A small value of τ_c implies high mobility and hence high channel quality variations, which invariably induce higher control packet exchanges. Fig. 10 shows the drastic reduction in the control overhead achieved by AFS.

2) Impact on Network Throughput: The reduction in the control overhead achieved by AFS allows us to increase the network throughput by up to 460% due to the increased use of D frames. For a large number of HP flows, Fig. 11 shows that AFS helps in improving the network throughput significantly, thereby making our protocol tailor-made for supporting multiple HP flows simultaneously. Fig. 12 shows the significant reduction in the admission delay achieved by our protocol compared with ECMA-392. It is to be emphasized that the entire proposed architecture is simulated to obtain the results in Figs. 9–12.

D. Evaluation of Adaptive Channel Assignment Strategy

1) Impact of Channel Bonding/Aggregation: The efficiency of channel bonding/aggregation in supporting HP flows is shown in Fig. 13, wherein we simulate 10 HP and 5 LP flows. We set ζ to 0.5. In this experiment, we compare three schemes: best single-channel transmission, assignment based on channel bonding/aggregation without using a second round of prob-

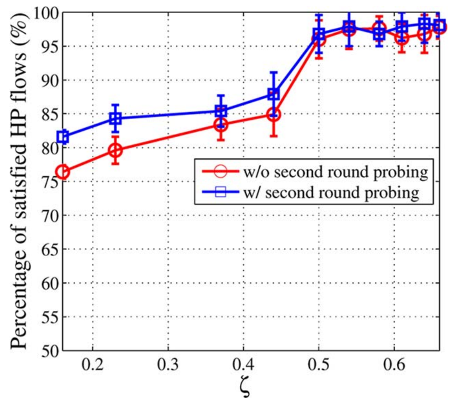

Fig. 14. Impact of second-round probing (10 HP and 5 LP flows; $R_d = 5$ Mb/s).

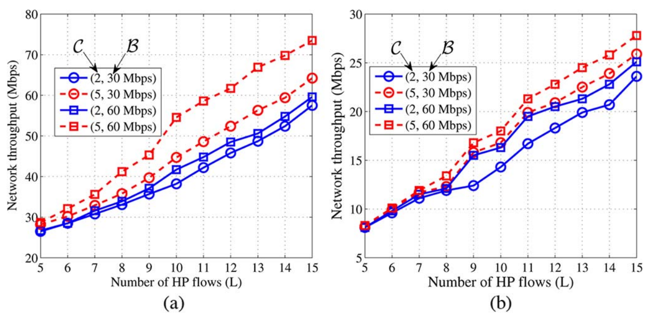

Fig. 15. Network throughput versus L ($\tau_c = 5$ s, $R_d = 9$ Mb/s, and $\kappa = 2$ Mb/s).

ing, and assignment based on channel bonding/aggregation augmented with a second round of probing. The best single-channel transmission scheme tries to accommodate the HP flow over the channel that exhibits the highest data rate among the ones discovered through channel probing. This scheme tries to emulate the "high preference" channel maintained at each node in multichannel MAC protocol [21]. However, in our case, the "best single channel" is chosen based on the supported data rate rather than channel occupancy. The best single-channel transmission scheme closely follows the proposed channel assignment strategy when the rate demands of the HP flows are relatively low. We can see a sharp dip in throughput with increase in R_d as the increased bandwidth requirement can no longer be accommodated in a single narrow-band channel, leading to the flow being dropped. This observation is very unappealing given the requirements imposed by current high-bandwidth applications. Channel bonding/aggregation helps in significantly improving the throughput as the value of R_d increases by bundling the available resources. One interesting observation made in Fig. 13 is the minor reduction in throughput by using second-round channel probing when the value of R_d is relatively low. This demonstrates the slight increase in overhead introduced by a second-round channel probing process leading to a minor reduction in throughput. However, Fig. 13 also demonstrates the increase in throughput brought in by deploying second-round probing when the value of R_d is high. As the value of R_d increases, the efficiency of second-round channel probing becomes prominent, as observed in Fig 13.

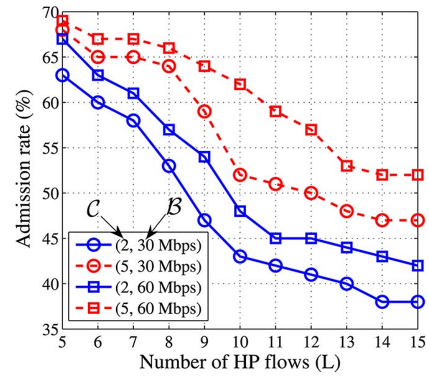

Fig. 16. Admission rate versus L ($\tau_c = 5$ s, $R_d = 7$ Mb/s, and $\kappa = 2$ Mb/s).

2) Impact of Second-Round Channel Probing: Second-round channel probing often acts as a decider when it comes to supporting multiple HP flows with high demands. When the PU activity is high over all channels, the number of idle channels at any given instant is small. As a result, the demands of several links might not be met by solely relying on the opportunities discovered by them. In such a scenario, second-round channel probing proves to be beneficial, as shown in Fig. 14. Thus, channel assignment based on aggregation/bonding augmented with second-round channel probing can be very useful when multiple HP flows need to be supported over a CRN exhibiting high PU activity.

E. Evaluation of Constrained Sensing/Probing Scheduling and Channel Assignment Strategies

Here, we evaluate the constrained scheduling and assignment strategies. $\mathcal{C}$ is the maximum number of channels that can be bonded/aggregated, and $\mathcal{B}$ is the maximum separation between the aggregated channels.

Fig. 15 shows the significant throughput gain achieved by our protocol, compared with ECMA-392, even when the constrained scheduling and assignment schemes are used. Fig. 16 shows the significant reduction in the admission rate as L increases when the constrained scheduling and assignment schemes are used.

VIII. Conclusion

Supporting bandwidth-intensive traffic over opportunistic radio networks has always allured the research community. We proposed a framework to guarantee the rate demands of multiple high-bandwidth flows in a centralized setup by designing several efficient spectrum exploration and exploitation strategies. We advocated a technique to drastically reduce the control overhead in frame-based systems to improve the network throughput. We also proposed an efficient channel allocation strategy, employing channel bonding and aggregation, which proved to be fruitful in increasing the throughput. Finally, we designed an optimal adaptive double-threshold based sensing policy, which was shown to drastically improve the spectrum sensing efficiency.

ACKNOWLEDGMENT

The statements made herein are solely the responsibility of the authors.

REFERENCES

[1] M. J. Abdel-Rahman, H. K. Shankar, and M. Krunz, "Adaptive cross-layer protocol design for opportunistic WLANs over TVWS," in *Proc. IEEE DySPAN Conf.*, Apr. 2014, pp. 519–530.

[2] A. Giannoulis, P. Patras, and E. W. Knightly, "Mobile access of wide-spectrum networks: Design, deployment and experimental evaluation," in *Proc. IEEE INFOCOM*, Apr. 2013, pp. 1708–1716.

[3] D. Xue, E. Ekici, and X. Wang, "Opportunistic periodic MAC protocol for cognitive radio networks," in *Proc. IEEE GLOBECOM*, Dec. 2010, pp. 1–6.

[4] *MAC and PHY for Operation in TV White Space*, ECMA-392, 2nd ed., Jun. 2012.

[5] R. Choudhury, X. Yang, R. Ramanathan, and N. Vaidya, "On designing MAC protocols for wireless networks using directional antennas," *IEEE Trans. Mobile Comput.*, vol. 5, no. 5, pp. 477–491, May 2006.

[6] H. Bany Salameh, M. Krunz, and O. Younis, "Cooperative adaptive spectrum sharing in cognitive radio networks," *IEEE/ACM Trans. Netw.*, vol. 18, no. 4, pp. 1181–1194, Aug. 2010.

[7] T.-S. Kim, H. Lim, and J. C. Hou, "Improving spatial reuse through tuning transmit power, carrier sense threshold, and data rate in multihop wireless networks," in *Proc. ACM MobiCom*, 2006, pp. 366–377.

[8] J. Zhu, Z. Xu, F. Wang, B. Huang, and B. Zhang, "Double threshold energy detection of cooperative spectrum sensing in cognitive radio," in *Proc. Int. Conf. Cogn. Radio Oriented Wireless Netw. Commun.*, 2008, pp. 1–5.

[9] J. Jafarian and K. Hamdi, "Throughput optimization in a cooperative double-threshold sensing scheme," in *Proc. IEEE WCNC*, 2012, pp. 1034–1038.

[10] T. S. Ferguson, Optimal Stopping and Applications. [Online]. Available: http://www.math.ucla.edu/~tom/Stopping/Contents.html

[11] N. Chang and M. Liu, "Optimal channel probing and transmission scheduling for opportunistic spectrum access," *IEEE/ACM Trans. Netw.*, vol. 17, no. 6, pp. 1805–1818, Dec. 2009.

[12] T. Bansal, B. Chen, and P. Sinha, "DISCERN: Cooperative whitespace scanning in practical environments," in *Proc. IEEE INFOCOM*, Apr. 2013, pp. 719–727.

[13] P. Bahl, R. Chandra, T. Moscibroda, R. Murty, and M. Welsh, "White space networking with Wi-Fi like connectivity," in *Proc. ACM SIGCOMM*, Aug. 2009, pp. 27–38.

[14] A. Kumar, K. G. Shin, Y. J. Choi, and D. Niculescu, "On time-domain coexistence of unlicensed and licensed spectrum users," in *Proc. IEEE DySPAN*, Oct. 2012, pp. 223–234.

[15] H. Kim and K. Shin, "Fast discovery of spectrum opportunities in cognitive radio networks," in *Proc. IEEE DySPAN Conf.*, Oct. 2008, pp. 1–12.

[16] M. J. Abdel-Rahman, H. K. Shankar, and M. Krunz, "Parallel sensing/probing architecture and adaptive protocol design for opportunistic networks," Dept. ECE, Univ. Arizona, Tucson, AZ, USA, Tech. Rep. TR-UA-ECE-2015-1, Mar. 2015. [Online]. Available: http://ece.arizona.edu/~mjabdelrahman/publications.html

[17] Y.-C. Liang, Y. Zeng, E. C. Y. Peh, and A. T. Hoang, "Sensing-throughput tradeoff for cognitive radio networks," *IEEE Trans. Wireless Commun.*, vol. 7, no. 4, pp. 1326–1337, Apr. 2008.

[18] F. Digham, M.-S. Alouini, and M. Simon, "On the energy detection of unknown signals over fading channels," in *Proc. IEEE ICC*, May 2003, pp. 3575–3579.

[19] S. O. Krumke, Integer Programming: Polyhedra and Algorithms. [Online]. Available: http://staff.guilan.ac.ir/staff/users/salahi/fckeditor_repo/file/ip-lecture-new.pdf

[20] V. Kanodia, A. Sabharwal, and E. Knightly, "MOAR: A multi-channel opportunistic auto-rate media access protocol for ad hoc networks," in *Proc. IEEE BroadNets*, Oct. 2004, pp. 600–610.

[21] J. So and N. H. Vaidya, "Multi-channel MAC for ad hoc networks: Handling multi-channel hidden terminals using a single transceiver," in *Proc. ACM MobiHoc*, 2004, pp. 222–233.

Mohammad J. Abdel-Rahman (S'12–M'15) received the Ph.D. degree in electrical and computer engineering from The University of Arizona, Tucson, AZ, USA, in 2014.

He is currently a Postdoctoral Associate with the Department of Electrical and Computer Engineering, Virginia Polytechnic Institute and State University, Blacksburg, VA, USA. His research interests include the areas of wireless communications and networking, with emphasis on dynamic spectrum access networks, wireless security, resource management, adaptive protocols, satellite communications, and wireless sensor networks.

Dr. Abdel-Rahman serves as a Reviewer for several international conferences and journals.

Harish Kumar Shankar received the B.S. degree in electronics and communications engineering from Anna University, Chennai, India, in 2010 and the M.S. degree in electrical and computer engineering from The University of Arizona, Tucson, AZ, USA, in 2013.

He is currently working in the field of enterprise network security with Palo Alto Networks, Santa Clara, CA. His research interests include wireless communications and networking, with emphasis on resource management, and adaptive protocol design for cognitive radio networks.

Marwan Krunz (S'93–M'95–SM'04–F'10) received the Ph.D. degree in electrical engineering from Michigan State University, Lansing, MI, USA, in 1995.

He is currently a Professor with the Department of Electrical and Computer Engineering and the Department of Computer Science, University of Arizona, Tucson, AZ, USA. He previously held other visiting research positions at INRIA-Sophia Antipolis, HP Labs, University of Paris VI, University of Paris V, and U.S. West Advanced Technologies. In January 1997, he joined the University of Arizona, after a brief postdoctoral stint at The University of Maryland, College Park, MD, USA. In 2010, he was a Visiting Chair of excellence with the University of Carlos III de Madrid, Madrid, Spain. He is currently the Site Codirector with the U.S. National Science Foundation Broadband Wireless Access and Applications Center. He is the author of more than 215 journal articles and peer-reviewed conference papers and is a coinventor on five U.S. patents. His research interests include wireless communications and networking, with emphasis on resource management, adaptive protocols, and security issues.

Dr. Krunz served as the General Cochair for the 2012 ACM Conference on Security and Privacy in Wireless and Mobile Networks and as the Technical Program Committee Chair for the 2004 IEEE Conference on Computer Communications; the 2005 IEEE International Conference on Sensing, Communication, and Networking; and the 2006 International Symposium on a World of Wireless, Mobile and Multimedia Networks. He has been a keynote speaker, an invited panelist, and a tutorial presenter at numerous international conferences. He served on the editorial boards for the IEEE/ACM TRANSACTIONS ON NETWORKING, the IEEE TRANSACTIONS ON MOBILE COMPUTING, the IEEE TRANSACTIONS ON NETWORK AND SERVICE MANAGEMENT, *Computer Communications Journal*, and the IEEE COMMUNICATIONS INTERACTIVE MAGAZINE. He currently serves on the editorial board for the IEEE TRANSACTIONS ON COGNITIVE COMMUNICATIONS AND NETWORKING. He received the U.S. National Science Foundation CAREER Award in 1998 and the IEEE Technical Committee on Computer Communications Outstanding Service Award in 2012. He was an Arizona Engineering Faculty Fellow (2011–2014) and a IEEE Communications Society Distinguished Lecturer (2013 and 2014).