Leveraging Diverse Propagation and Context for Multi-Modal Vehicular Applications

Pengfei Cui, Hui Liu, Jialin He, Onur Altintas, Rama Vuyyuru,
Dinesh Rajan, and Joseph Camp

***Abstract*—Vehicular wireless channels have a high degree of variability, presenting a challenge for vehicles and infrastructure to remain connected. The emergence of the white space bands for data usage enables increased flexibility for vehicular networks with distinct propagation characteristics across frequency bands from 450 MHz to 6 GHz. Since wireless propagation largely depends on the environment in operation, a historical understanding of the frequency bands' performance in a given context, could speed multiband selection as vehicles transition across diverse scenarios. In this paper, we leverage knowledge of in-situ operation across frequency bands with real-time measurements of the activity level to select the optimal band for the particular application in use. To do so, we perform a number of experiments in typical vehicular topologies. With two models based on machine learning algorithms and an in-situ training set, we can predict the throughput on a free channel. We can then consider the activity level per band to compute the resulting performance one could expect on context information to guide protocol. In the field, we exploit the propagation differences experienced per band to show that training on a repeatable route can yield vast performance improvements from prior schemes. We show that minimal amounts of training can provide such improvements and that a simple scheme that can allow multiband adaptation gains when there is insufficient levels of training.**

I. Introduction

Drivers can benefit from a wide array of vehicular applications ranging from real-time traffic monitoring and safety applications to *infotainment* applications spanning news, weather, audio, and video streams. However, the continuous use of such applications is limited due to the challenge of transmitting over highly-dynamic vehicular wireless channels. In such networks, the increasing availability of different frequency bands with correspondingly diverse propagation characteristics could allow flexibility and robustness of vehicular links. Even with spectral flexibility, links are extremely tenuous, demanding instantaneous decisions to remain connected, motivating an algorithm that can find the appropriate frequency band quickly and according to the current application.

Another motivation for studying multiband adaptation is the recent availability of white space frequency bands. Prior work has considered a number of challenges in leveraging these white space frequencies including spectrum sensing, frequency-agile operation, geolocation, solving stringent spectral mask requirements, and providing reliable service in unlicensed and dynamically changing spectrum [1]. In particular, there has recently been an acceleration in spectrum sensing work [2]–[4]. Based on these works, protocols have been built for multi-channel and/or multiband wireless operation [5]–[7]. Other works have presented a method for searching the most efficient transmission channel [8], discovering channel information from limited measurements [2], [7], and estimating channel quality through limit information [5].

While these works have considered spectral activity and developing protocols and algorithms to find spectral holes, less of a focus has been on coupling such information with contextual knowledge of the propagation environment. In this paper, we present a multiband adaptation protocol which couples the prior knowledge of performance of various bands in a given context with the instantaneous knowledge of various measurable attributes of each band to arrive at a decision on the optimal band to transmit. To do so, we use an off-the-shelf platform that allows direct experimentation across four different wireless frequency bands simultaneously from 450 MHz to 5.8 GHz while maintaining the same physical and media access layers.

The main contributions of our work are as follows:

- We first develop a framework for multiband adaptation using both historical information and instantaneous measurements. This framework is broad enough to study adaptation across licensed, and unlicensed band including white space frequency bands.
- We propose two different machine-learning-based adaptive algorithms. The first machine learning algorithm, which we refer to as the *Location-based Look-up Algorithm*, is based on the idea of $k-$nearest-neighbor classification. The second machine-learing-based algorithm uses *decision trees* for classification. For comparison, we also create two baseline adaptation algorithms which attempt to make the optimal band selection based on only: *(i)* historical performance data, and *(ii)* instantaneous SNR measurements across various bands.
- We perform extensive outdoor V-2-V experiments to evaluate the proposed algorithms. Our results indicate that the proposed machine learning algorithms outperform these baseline methods in throughput by up to 49.3%.

II. Multiband Adaptation

In this section, we first formulate the multiband adaptation problem in vehicular wireless links and introduce the context information that we use. We then propose two machine-learning-based multiband adaptation algorithms for vehicular channels. For comparison, we also propose two baseline adaptation methods based on existing solutions.

A. Problem Formulation

Consider a system with n frequency bands, represented by an index set $\{1, 2, \ldots, n\}$. The objective is to select the optimal band, b_{best}, to transmit at each time instant that maximizes a desired objective function such as throughput. The throughput r_i on band i depends on several factors such as received signal power P_R^i, noise power P_N^i, the activity/occupancy level B^i, the velocity of the transmitter, v_{tx}, the velocity of the receiver, v_{rx}, and other factors such as location and contextual information. This relationship is represented in general as $r_i = f(P_R^i, P_N^i, B^i, v_{tx}, v_{rx},\text{context-information})$. The objective can be stated as

$$b_{best} = \arg\max_i r_i \tag{1}$$

The framework allows us to separate the interference from other nodes using the same technology via the busy time and the interference from nodes using other technologies in the same and via the noise level P_N^i. For instance, an 802.11 node can observe the packets of other 802.11 nodes but only the increase noise levels from other Zigbee/Bluetooth nodes. The existing pattern embedded in the historical data connecting the performance of different bands and collected context information e.g., v_{tx}, v_{rx}, B^i, P_N^i, P_R^i could be extracted and help make decisions for multiband adaptation in a similar context [9].

To represent the utilization level of the channel, we define *busy time*, B, as the percentage of time when the channel is occupied by all competing sources $x_j (j = 1, 2, 3, ...)$ other than the intended transmitter y. For 802.11-based transmissions, the busy time on band i is defined as:

$$B^i = \frac{\sum_j \sum_k \frac{L_k^{x_j}}{R_k^{x_j}}}{\sum_k \frac{L_k^y}{R_k^y} + \sum_j \sum_k \frac{L_k^{x_j}}{R_k^{x_j}} + S\sigma} \tag{2}$$

where $L_k^{x_j}$ and $R_k^{x_j}$ represent, respectively, the packet length in bits and the data rate at which that packet is transmitted, for external sources x_j; S and σ are the idle slots and the slot duration. When considering the activity level of non-802.11 users (e.g., the bands currently licensed to TV and other non-802.11 devices), we use the received signal level from non-802.11 interference sources P_N^i on band i as an input to our algorithms.

B. Multiband Adaptation Algorithms

In order to evaluate the proposed multiband adaptation algorithms, we construct two baseline multiband adaptation methods: (1) We search the most commonly selected band as the best band in the historical data and configure the most common band as the final decision. (2) For each band, we build a lookup table for throughput T_{ideal} in the ideal channel given $RSSI$ and obtain the best band according to following:

$$\max_i T_{ideal}^i \times (1 - B^i), \tag{3}$$

The throughput T_{ideal} is measured with Azimuth ACE-MX channel emulator [10]. The details of system setup and data collection are discussed in Section III.

Machine learning has been introduced as an important tool in wireless communication [11]. When the user enters an area, the machine learning algorithm can learn from the historical data and train a mapping function to select the potential optimal band given the input, e.g., P_R^i, v and P_N^i. We propose two multiband adaptation algorithms based on two machine learning methods: the k-nearest neighbor (KNN) and the decision tree.

Location-based Look-up Algorithm. KNN is a machine learning method based on searching closest training data points in the feature space and various modified versions have been applied successfully for classification [12]. In the Location based Look-up Algorithm, we search the closest neighbors of a testing point by using each parameter one by one in the input set. The performance of the selected training data points is averaged to generate an estimate of the performance at each band. Then the band with the highest throughput performance is selected as the b_{best}. The Location based Look-up Algorithm additionally involves geographic information for band selection.

For the Location-based Look-up Algorithm, context information involves location information g from GPS, v, P_R^i, P_N^i and B^i. To make a band prediction, we have four look-up blocks to narrow down the training data points which are similar to the testing data point. First, we search the historical data which is closest to the testing data based on GPS location, if the number of found historical data points is less than a predefined threshold, θ_{AArea}, we increase the distance range (the actual threshold is discussed in Section III). Then based on the filtered historical data, we search the close data points based P_R^i with a threshold for number of data points, θ_{ARSSI}. A similar process is repeated based on P_N^i and v, respectively. After deciding the final data set, we average the throughput of data points at each band. The key steps of this algorithm are presented as Algorithm 1.

Region-based Decision Tree Algorithm. Decision trees are a widely-used machine learning algorithm due to its low complexity and stable performance [13]. A decision tree can model the relationship in the training data between the context information and the optimal band as a set of tree-like deduction structures. Before implementing the training process, we prepare a training set including a group of training data points of $\{v_{tx}, v_{rx}, P_R^1, ..., P_R^n, B^1, ..., B^n, P_N^1, ..., P_N^n, b_{best}\}$ based on the collected measurements. We obtain b_{best} by comparing the throughput performance of all available bands and selecting the band with the highest throughput. We choose C4.5 algorithm [14], a widely-used algorithm based on the information entropy gain to build the decision tree used in our system. At each intermediate node in the decision tree, the learning algorithm calculates the information entropy gain of splitting the remaining training data points based on each parameter in the input set, e.g., P_R^i, v or P_N^i. Then, it compares and selects the parameter with the highest entropy gain to decide the test condition at each intermediate node until all training data points are classified. The leaf nodes indicate the optimal band for prediction in our application. Then, the trained decision structure is integrated into the transmitter protocol stack. With the collected context information, the decision structure can suggest the band with the best throughput performance.

The relationship between the context information and the best band could differ at different locations because of diverse propagation environment characteristics. To reduce the heterogeneity of training data from different locations, we split the vehicular route into several regions and implement the training process based on the historical data collected in each region.

Algorithm 1 Location-based Look-up Algorithm

Input:

g: Location information of multiband node
θ_{Area}: Threshold of a location
θ_{RSSI}: Threshold of RSSI
$\theta_{Velocity}$: Threshold of velocity
θ_{AArea}: Threshold of data amount for a location
θ_{ARSSI}: Threshold of data amount for RSSI
$\theta_{ANon802.11SI}$: Threshold of data amount for non-802.11 interference
$\theta_{AVelocity}$: Threshold of data amount for velocity
$D^i \in \{D^1, D^2, \ldots, D^n\}$: Historical look-up data

Output:

b_{best}: Optimal transmission band

for $i <= n$ **do**
  Initialize $Data_{Location}, Data_{RSSI}, Data_{Velocity}$ to zero matrix;
  while $Amount(Data_{Location,i}) < \theta_{AArea}$ **do**
    $Data_{Location,i} \leftarrow f_{Lookup}(D^i, g, \theta_{Area})$: Find data in D^i whose distance less than θ_{Area};
    $\theta_{Area} = \theta_{Area} \times 1.1$;
  end while
  while $Amount(Data_{RSSI,i}) < \theta_{ARSSI}$ **do**
    $Data_{RSSI,i} \leftarrow f_{Look-up}(D_{Location,i}, P_R^i, \theta_{RSSI})$: Find data in $D_{Location}$ the RSSI similar to P_R^i in range θ_{RSSI};
    $\theta_{RSSI} = \theta_{RSSI} \times 1.1$;
  end while
  while $Amount(Data_{P_N,i}) < \theta_{ANon802.11SI}$ **do**
    $Data_{RSSI,i} \leftarrow f_{Look-up}(D_{Location,i}, P_N^i, \theta_{RSSI})$: Find data in $D_{Location}$ the RSSI similar to P_N^i in range θ_{RSSI};
    $\theta_{ANon802.11SI} = \theta_{ANon802.11SI} \times 1.1$;
  end while
  while $Amount(Data_{Velocity,i}) < \theta_{AVelocity}$ **do**
    $Data_{Velocity,i} \leftarrow f_{Lookup}(D_{RSSI,i}, v, \theta_{Velocity})$: Find data in D_{RSSI} the RSSI similar to v in range θ_{RSSI};
    $\theta_{Velocity} = \theta_{Velocity} \times 1.1$;
  end while
  $T_{a,i} = avr(Data_{Velocity,i})$;
  $T_{e,i} = T_a^i \times (1 - B^i)$;
end for
$b_{best} = \max_i\{T_e^1, \ldots, T_e^i, \ldots, T_e^n\}$;

Then, the trained decision structure is integrated in our system for multiband adaptation in each region. The granularity of region division is one parameter that affects the training set as well as the performance of the resulting decision tree. We evaluate granularity of these divisions in Section III.

III. Experimental Analysis for Prediction Algorithms

As discussed in Section 3, some algorithms are more appropriate in certain scenarios. To study these algorithmic differences, we have developed indoor and in-field experiments on an off-the-shelf wireless platform. To ensure the results are broadly applicable, we employ a Linux-based 802.11 testbed [15]. The platform includes a Gateworks 2358 motherboard with Ubiquiti XR series radios (XR9 at 900 MHz, XR2 at 2.4 GHz, XR5 at 5.2 GHz) as well as a DoodleLabs DL475 radio at 450 MHz [16], [17]. Another instrument involved in the experimentation is an Azimuth ACE-MX channel emulator, allowing controllable propagation and fading characteristics with a broad range of industry-standard channel models from 450 MHz to 5.9 GHz [10].

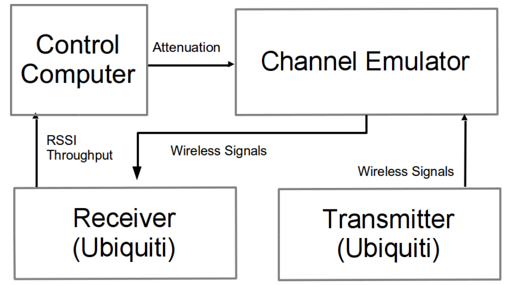

Fig. 1. Experimental setup for channel emulator.

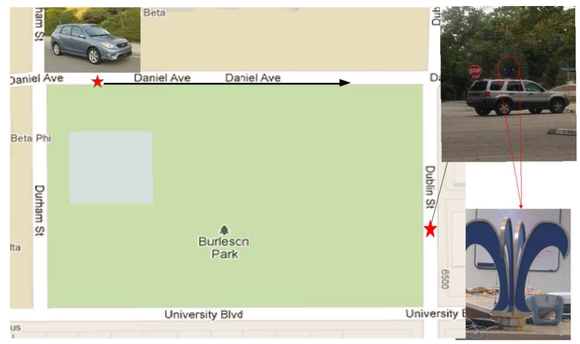

Fig. 2. In-field Experiment Setup

A. In-lab Experiments for Radio Characterization

To establish an SNR to throughput relationship for the *SNR-based Throughput Look-up Algorithm*, we use an experimental setup where two wireless nodes communicate across repeatable emulated channels generated by Azimuth ACE-MX channel emulator (Fig. 1). For a given band and card, we measure the throughput of a fully-backlogged UDP flow using the *iperf* traffic generator. We use constant attenuation over an idealized channel condition (i.e., a bypass channel) and repeat the experiment to produce various RSSI values. Despite the same physical and media access layers of the radios, there are slight differences in the throughput achieved per radio at the same attenuation level. Thus, we normalize these throughput values to have the same maximum throughput across radio types for a fair comparison of the frequency bands.

B. Experimental Design for In-field Data Collection

We now describe the in-field experimental design to obtain a data set for evaluating our multiband algorithms. Two Gateworks boards, each containing the aforementioned four radios are installed on two cars. One node is always the receiver and at a fixed location. The other node is always the transmitter and traverses around the block of a public park as shown in Figure 2, one loop of which will be used as a unit of training in the next section.

During each loop, the transmitter sends a fully-backlogged UDP flow using *iperf* over each of the four radios simultaneously. To focus on band selection, we disable autorate and use a fixed data rate of 6 Mbps. The receiver continually performs a *tcpdump* of all received 802.11 packets [18]. Additionally, a QH 400 Quad Ridge Horn antenna (shown in Figure 2) is connected to a Rhode & Schwarz FSH8 mobile spectrum analyzer at the receiver's location to monitor spectral activity. Then, based on the time stamps, 802.11 packets can

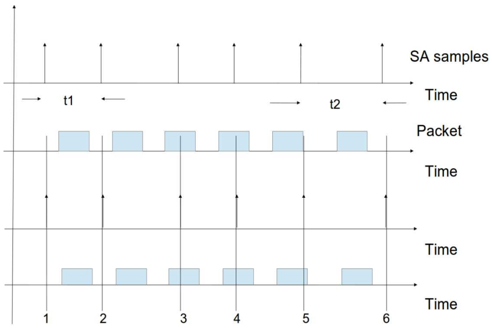

Fig. 3. Spectrum Analyzer Data Processing

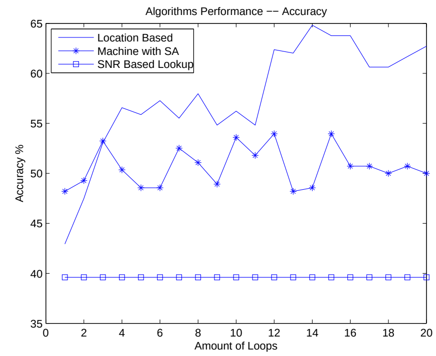

Fig. 4. Accuracy of the four multiband algorithms.

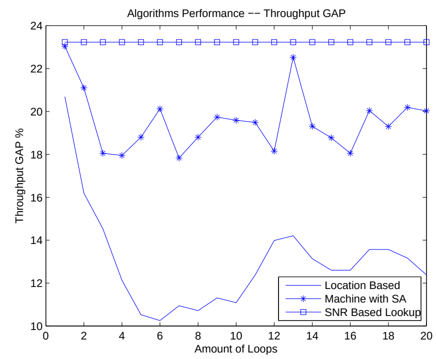

Fig. 5. Throughput Gap of the four multiband algorithms.

be removed from the spectral trace so that only non-802.11 interference will contribute to P_N^N for our algorithms.

Figure 3 shows how we find the value of the non-802.11 interference, P_N^N. We delete the spectrum analyzer (SA) samples which overlap in time with the dumped 802.11 packets, such as packet 3, 4, and 5 shown. Then, the reported interference value will not contain received power from 802.11 packets which have already been considered in the busy time, B.

The in-field data is processed offline where data from all instruments involved is synchronized based on the GPS time stamps of each. As discussed in III-A, the throughput of each radio is normalized based upon the emulator experiment to account for any manufacturing differences.

C. Performance Analysis of Algorithms

We now investigate the performance of our proposed multi-band adaptation algorithms with data collected from Sections III-A and III-B. To investigate the performance, the metrics of *Accuracy* and *Throughput Gap* are used in the evaluation. To do so, we consider each second of the in-field trace. In each second, we can observe the frequency band that had the highest throughput. The *Accuracy* is defined as the percentage of time the prediction of the best band for that second matches the observed best band. Conversely, the *Throughput Gap* is defined as the difference between the throughput observed on the best band and the throughput achieved by the predicted best band over the throughput of the observed best band.

Since the *SNR-based Throughput Look-up Algorithm* requires only emulator-based training, the *Accuracy* and *Throughput Gap* can be calculated for all loops of the in-field trace. However, the *Location-based Look-up Algorithm* and the *Region-based Decision Tree Algorithm* requires in-field training. Thus, the data set must be divided into a training set and testing set for evaluation.

In Figure 4, we show the aforementioned *Accuracy* of the four multiband algorithms in selecting the band with the highest throughput. The x-axis represents the number of loops the mobile transmitter traversed around the block shown in Figure 2 which will be used by the machine-learning-based algorithms. We use the same training and testing set to compare the *Location-based Look-up Algorithm* and the *Region-based Decision Tree Algorithm*. From the results, we observe the following:

- At each loop, the first baseline algorithm, *Most Commonly-Selected Band*, uses the band with the greatest long-term average of the percentage of time that band yields the highest throughput over the previous loops. The accuracy ranges from 36.1% to 42.9%.
- The second baseline algorithm, *SNR-based Throughput Look-up*, maintains a 39.2% across all the data since it relies only on emulator-based training.
- The *Region-based Decision Tree Algorithm* has an accuracy ranging from 48.2% to 54.0% but contains many dips due to the relationship between the context information and the distribution of the best band choice changing on a loop-by-loop basis. Additional training data slightly improves the decision structure overall but primarily induces additional noise in the training process.
- The *Location-based Look-up Algorithm* begins with an accuracy of 42.5% but improves the most out of any algorithm to finish with an accuracy 62.5% with a highest accuracy of 65.0% occurring after loop 14. Additional in-field training loops are likely to further improve the multiband selection accuracy.

Next, we describe Figure 5, which depicts the *Throughput Gap* of the four algorithms we evaluated.

- The *Most Commonly-Selected Band Algorithm* has two different modes of throughput gap based upon which band has the highest long-term percentage. For loops 2-4, the choice is 5.8 GHz, which has a gap of 41.1% using the test set. For all other loops, the choice is 2.4 GHz, which has a gap of 27.7%.
- The *SNR-based Throughput Look-up Algorithm* shows a baseline performance of 23.6% for the throughput gap.
- The *Region-based Decision Tree Algorithm* has some gains with additional training, going from a throughput

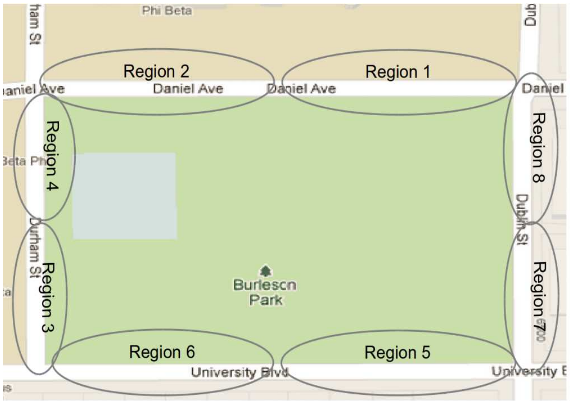

Fig. 6. Spatially splitting experimental area into 8 regions.

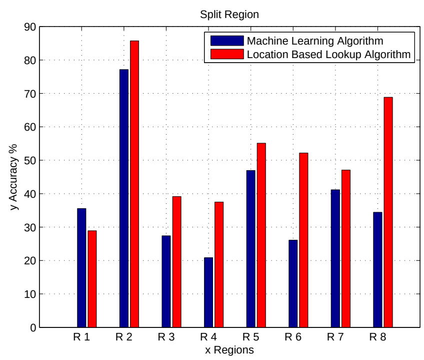

Fig. 7. Accuracy when dividing training set into 8 regions.

gap of 23.0% to 20.0%. As explained before, context information may change spatially and temporally, bringing dips to the curve.

- Finally, the *Location-based Look-up Algorithm* takes only 6 loops of training to reach its lowest value of 10.2% in terms of throughput gap. From loop 1 to loop 20, the throughput gap goes from 20.8% to 12.3%, which could still be improved upon with additional training.

We now consider the effect of further sub-dividing in-field experimental testing data into regions for our *Location-based Look-up Algorithm* and *Region-Based Decision Tree Algorithm*. To do so, we divide the loop around the park into eight regions as shown in Figure 6, which has two competing effects: (1) Smaller regions allow similar experimental data to be used in the training process, potentially improving the decision structure. (2) For a given training set, dividing it into regions reduces the number of training points for the machine learning algorithms, potentially weakening the decision structure. In Figure 7, we observe the *Accuracy* of the eight regions for both algorithms.

- In all but Region 1, the *Location-based Look-up Algorithm* has better performance than *Region-based Decision Tree Algorithm*. The improved accuracy of the former algorithm can be attributed to its ability to distinguish between the middle of a given region to the boundary, considering each point differently. For the *Region-based Decision Tree Algorithm* to capture such a notion, the regions would have to be further sub-divided, increasing the number of trees and reducing the training set per tree.
- For this training set, the reduction in training data caused by the regional divisions had a net loss on the performance of the *Region-based Decision Tree Algorithm*. However, if the training set was much larger for a given area, the net effect of regional divisions could be a promising technique.

IV. Conclusion

In this paper, we investigated the multiband adaptation to leverage the propagation and context for vehicular applications. We did so by proposing two machine-learning-based schemes and compared there performance against two baseline schemes. In our experimental analysis, we evaluated the performance of these algorithms in the field on an off-the-shelf platform. Experimental results demonstrate that the proposed algorithms can achieve up to 49.3% greater throughput than the baseline algorithms with an accuracy up to 65%. Since for mobile networks have limited energy, in future work we plan to examine the energy efficiency of the simultaneous use of multiple, diverse radios.

References

[1] S. Shellhammer, A. Sadek, and W. Zhang, "Technical challenges for cognitive radio in the tv white space spectrum," in *Information Theory and Applications Workshop, 2009*. IEEE, 2009, pp. 323–333.

[2] S. Rayanchu, V. Shrivastava, S. Banerjee, and R. Chandra, "Fluid: improving throughputs in enterprise wireless lans through flexible channelization," in *Proceedings of the 17th annual international conference on Mobile computing and networking*. ACM, 2011, pp. 1–12.

[3] S. Kim, H. Bertoni, and M. Stern, "Pulse propagation characteristics at 2.4 ghz inside buildings," *Vehicular Technology, IEEE Transactions on*, vol. 45, no. 3, pp. 579–592, 1996.

[4] D. Cabric, S. Mishra, and R. Brodersen, "Implementation issues in spectrum sensing for cognitive radios," in *Signals, Systems and Computers, 2004. Conference Record of the Thirty-Eighth Asilomar Conference on*, vol. 1. Ieee, 2004, pp. 772–776.

[5] V. Kanodia, A. Sabharwal, and E. Knightly, "MOAR: A multi-channel opportunistic auto-rate media access protocol for ad hoc networks," in *Broadnets*, San Jose, CA, October 2004.

[6] D. Raychaudhuri and X. Jing, "A spectrum etiquette protocol for efficient coordination of radio devices in unlicensed bands," in *Personal, Indoor and Mobile Radio Communications, 2003. PIMRC 2003. 14th IEEE Proceedings on*, vol. 1. IEEE, 2003, pp. 172–176.

[7] A. Sabharwal, A. Khoshnevis, and E. Knightly, "Opportunistic spectral usage: Bounds and a multi-band csma/ca protocol," *Networking, IEEE/ACM Transactions on*, vol. 15, no. 3, pp. 533–545, 2007.

[8] J. Mo, H. So, and J. Walrand, "Comparison of multi-channel mac protocols," in *Proceedings of the 8th ACM international symposium on Modeling, analysis and simulation of wireless and mobile systems*. ACM, 2005, pp. 209–218.

[9] R. Meikle and J. Camp, "A global measurement study of context-based propagation and user mobility," in *Proceedings of the 4th ACM international workshop on Hot topics in planet-scale measurement*. ACM, 2012, pp. 21–26.

[10] "Aximuth ACE-MIMO Channel Emulator," http://www.azimuthsystems.com, Mar. 2011.

[11] S. Haykin, "Cognitive radio: brain-empowered wireless communications," *Selected Areas in Communications, IEEE Journal on*, vol. 23, no. 2, pp. 201–220, 2005.

[12] H. Zhang, A. Berg, M. Maire, and J. Malik, "Svm-knn: Discriminative nearest neighbor classification for visual category recognition," in *Computer Vision and Pattern Recognition, 2006 IEEE Computer Society Conference on*, vol. 2, 2006, pp. 2126–2136.

[13] R. Banfield, L. Hall, K. Bowyer, W. Kegelmeyer, *et al.*, "A comparison of decision tree ensemble creation techniques," *IEEE Trans on Pattern Analysis and Machine Intelligence*, vol. 29, no. 1, p. 173, 2007.

[14] M. Hall, E. Frank, G. Holmes, B. Pfahringer, P. Reutemann, and I. Witten, "The weka data mining software: an update," *ACM SIGKDD Explorations Newsletter*, vol. 11, no. 1, pp. 10–18, 2009.

[15] "Openwrt wireless freedom," https://openwrt.org/, Dec. 2011.

[16] "Cambria gw2358-4 network platform," http://www.gateworks.com/, 2007.

[17] "Ubiquiti xtremerange series of radio," http://www.ubnt.com/, 2012.

[18] V. Jacobson, C. Leres, and S. McCanne, “The tcpdump manual page,” *Lawrence Berkeley Laboratory, Berkeley, CA*, 1989.