Apprenticeship Learning based Spectrum Decision in Multi-Channel Wireless Mesh Networks with Multi-Beam Antennas

Yeqing Wu, Fei Hu, *Member, IEEE*, Sunil Kumar, *Senior Member, IEEE*, John D. Matyjas, *Senior Member, IEEE*, Qingquan Sun, *Member, IEEE*, and Yingying Zhu, *Member, IEEE*

***Abstract*—We propose a novel spectrum decision scheme (i.e., channel selection and handoff) for wireless mesh networks (WMN) which use multiple channels and nodes equipped with multi-beam directional antennas. Our scheme has the following features: (i) It performs spectrum decision by considering various WMN parameters, including the channel quality, beam orientation, antenna-caused deafness and capture effects, and application priority level; (ii) It uses the reinforcement learning (RL)-based spectrum decision process to achieve the optimal quality of multimedia transmission in the long term. However, a newly-joined WMN node could take a long time to make a correct spectrum decision due to the difficult choice of initial RL parameters. Therefore, our scheme uses the apprenticeship learning in conjunction with the RL model, to speed up the spectrum decision process by choosing a suitable neighboring node (called 'expert') to teach a newly-joined node (called 'apprentice'). Our experiments demonstrate that the proposed spectrum decision scheme improves the network performance and multimedia transmission quality.**

***Index Terms*—Spectrum Decision, Channel Selection, Wireless Mesh Networks (WMN), Directional Antennas, Multi-Beam, Apprenticeship Learning, Reinforcement Learning, Multimedia Transmission.**

I. Introduction

The performance of wireless mesh networks (WMN) degrades when their size and number of hops increase. The use of multiple channels can significantly improve WMN throughput due to the reduction in the contending transmissions in the frequency domain [1], [2]. The use of directional antennas can further improve the WMN capacity by enhancing the transmission range and spatial reuse. Lately, the use of multi-beam smart antennas (MBSAs) has been investigated in the literature [3], [4]. However, most schemes select channels simply based on the non-interference principle between neighboring links, without considering other important factors (such as the link quality, antenna deafness, and node capture effect); these factors can considerably degrade the network performance.

In this paper, we address the spectrum decision issues in directional, multi-channel WMNs, consisting of nodes equipped with MBSAs. Here, the spectrum decision process includes two steps: Step 1 is spectrum selection (i.e., selection of a suitable channel with proper bandwidth to suit the application quality of service (QoS); Step 2 is spectrum handoff, which allows a node to switch to a new channel if the existing channel becomes unavailable or has poor quality. The spectrum handoff also includes a special case, i.e., a node may choose to temporarily pause its data transmission until the channel is available again. This would be useful when the spectrum handoff delay is relatively large and the link outage is likely to be short.

It is important to make an intelligent spectrum decision in directional, multi-channel WMNs due to the following reasons. *First*, the spectrum selection should consider the time-varying WMN conditions. *Second*, the use of multi-beam directional antennas complicates the channel switching scheme in WMNs, as widely different traffic patterns may be present in different beams of a node. Also, the nodes equipped with directional antennas suffer from the deafness and hidden terminal problems. *Third*, the spectrum decision should not be executed in a myopic way since a channel selection in the current time slot could adversely impact the long-term optimization goal. For example, if the existing channel has started experiencing high interference, we may not need to immediately switch to a new channel if we know that the current channel will again become available soon (based on the statistical analysis of the past channel usage), and the application allows us to buffer the unsent packets during the link outage period. Therefore, it is important to define a long-term optimization goal to determine the spectrum decision policy based on cumulative reward.

In our previous work [5], we designed a reinforcement learning (RL)-based intelligent spectrum handoff scheme for cognitive radio networks (CRNs), to achieve the maximum cumulative, long-term system rewards. In this paper, we use RL-based spectrum decision scheme for *wireless mesh networks with directional antennas*, which has a different application environment than [5]. In the proposed scheme, we consider the beam orientation, and the antenna-caused deafness and capture effects. A significant contribution of this scheme is the use of a teaching model (called the apprenticeship learning (AL)) for speed-up of the spectrum adaptation process. The proposed intelligent spectrum decision scheme not only

Manuscript received 13 Nov 2014; revised 13 May 2015 and 11 Oct 2015. This work was supported by U.S. Department of Defense under Grant No. FA8750-14-1-0075. Approved for public release; Distribution Unlimited: 88ABW-2016-1345, 21 March 2016.

Y. Wu and F. Hu are with the Electrical and Computer Engineering, University of Alabama, Tuscaloosa, AL 35487, USA (e-mail: ywu40@ua.edu, fei@eng.ua.edu).

S. Kumar is with the Electrical and Computer Engineering, San Diego State University, San Diego, CA 92182, USA (e-mail: skumar@mail.sdsu.edu).

J. D. Matyjas is with Air Force Research Laboratory, Rome, NY 13441, USA (e-mail: john.matyjas@us.af.mil).

Q. Sun is with the Computer Science and Engineering, California State University San Bernardino, San Bernardino, CA, 92407, USA (e-mail: quanqian12345@gmail.com).

Y. Zhu is with the Electrical and Computer Engineering, University of California, Riverside, CA 92507, USA (e-mail: yzhu010@ucr.edu).

utilizes the channel usage history of the node which is making the spectrum decision, but it also exploits the usage history of its neighboring nodes by using the AL algorithm to speed up the optimization of spectrum decision, by shortening the RL algorithm execution time and avoiding the local maxima.

We denote a newly-joined WMN node or an existing node that needs "guidance" from other nodes as an *apprentice node*. The node to be learned from is called an *expert or teacher node*. The AL-based learning model has the following two advantages: (1) In order to speed up its optimization procedure, the apprentice node can initialize its own reward function based on the useful RL parameters (such as critical state-action pairs) from an expert node that has well-adapted channel selection/handoff behavior; (2) AL can largely avoid channel switching errors (i.e., switching to a wrong channel and/or at a wrong time) through future WMN state prediction. By analyzing the past state transition patterns, a node can predict which channel will be available in the next phase.

The above AL-based inter-node learning model needs to solve three critical issues. (1) **When**: the first issue is to figure out when an apprentice node should trigger its learning process. (2) **Who**: the second issue is to find a suitable expert which has the most appropriate information/knowledge for the apprentice node. (3) **How**: the third issue is how to learn from the expert. To reduce the communication overhead, the expert node should select only the most critical, representative data for transfer to the apprentice node. In this paper, we use the Manifold learning to reduce the dimensionality of the data. The apprentice node uses this information (transferred from the expert node) to come up with a complete radio environment adaptation model, in order to handle any unseen, complex WMN conditions.

A. Main Contributions

The main contributions of our work are:

(1) *Spectrum decision with the consideration of a time-varying WMN environment*: We design an efficient spectrum decision scheme with optimized spectrum selection and handoff, and effectively exploit the spatial opportunities provided by a multi-beam directional antenna as well as the frequency opportunities from orthogonal multi-channels. Our scheme considers the channel quality, beam orientation, the antenna-caused deafness and capture effect, and the application QoS. Note that the conventional channel selection schemes only consider the interference between the neighboring links.

(2) *AL enhanced learning model*: Conventional spectrum selection simply picks up the best channel in a myopic manner, without considering its impact on the long-term QoS performance. Our scheme uses RL to overcomes this issue by defining a cumulative reward after each phase of spectrum decision. Besides, our scheme uses the AL model to speed up the spectrum decision process by enabling an apprentice node to learn from an appropriate neighboring expert node. We also solve the above mentioned when-who-how issues in AL.

The rest of this paper is organized as follows. The related work is briefly discussed in Section II. Section III describes the assumptions of directional WMNs used in this paper. Section IV describes the core idea of our proposed spectrum decision scheme. Section V briefly describes the RL-based spectrum decision which is an important part of our proposed scheme. The AL-based enhancement of our proposed spectrum decision scheme is described in Section VI. Simulation results are presented in Section VII, followed by conclusions in Section VIII.

II. Related Work

In this section, we briefly describe the existing literature related to our work.

A. WMNs with Directional Antennas:

Use of directional antennas in WMN has been investigated in several schemes. The authors in [3] proposed a MAC protocol to address beam-synchronization, beam-overlapping, mobility, and receiver blocking (deafness problem). A polling-based MAC protocol for a WLAN with multi-beam access point was discussed in [4]; it supports the QoS and power conservation for individual mobile users. A unified MAC layer in ad hoc networks with smart antennas was explored in [6]. Other MAC protocol designs which exploit directional antennas to improve the network performance can be found in [7]–[10]. However, these schemes select a channel for transmission based on the idle status of the selected channel. They seldom investigate other factors that can affect the system throughput, such as the link load, channel quality, and antenna-caused deafness and capture effects.

B. Multi-channel WMNs:

Several channel selection schemes have been proposed for WMNs. For example, the authors in [11] investigated the unique constraints and issues of channel assignment in WMNs, and proposed a channel assignment scheme which incorporates the traffic pattern and connectivity issues to minimize the interference in WMNs. A multi-channel network architecture that integrates multiple channels and directional antennas to improve the network capacity was proposed in [12]. Other channel assignment schemes in WMNs can be found in [13]–[16]. However, these schemes choose the channel for transmission in a myopic manner (i.e., they maximize the immediate rewards without considering the impact of the current action on the future node state). In fact, a greedy scheme may not be able to achieve the optimal rewards in the long-term in a time-varying WMN.

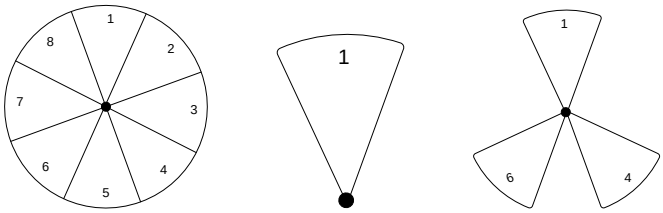
C. Learning-based Node Adaptation:

The concept of an old node teaching a new node was initially proposed in [17]; it can save energy during the startup and learning process. It has been successfully used for interference management in Femtocell Networks [18]–[21]. However, these schemes simply assume that the nearest neighbor node is the expert node. In directional WMNs, the neighbor nodes may have totally different traffic patterns or application requirements, compared to a newly-joined node. Therefore, a similarity comparison scheme needs to be used to compare the node status.

III. Network Model

A hybrid WMN architecture is assumed in this paper, in which the mesh clients can access the network through mesh routers or by directly meshing with other clients [22]. We assume the use of an efficient directional MAC scheme, called circular directional RTS MAC (CDR-MAC) protocol proposed in [10]. The CDR-MAC introduces a circular directional transmission of the RTS control packet to discover and track their neighbor nodes. By caching the neighbor information, CDR-MAC protocol can reduce the hidden-terminal and the deafness problems.

A. Antenna Model



(a) Omni beam (b) Single beam (c) Multiple beams

Fig. 1. The antenna model.

The antenna model of multiple beam smart antennas (MBSAs) in this paper follows the antenna model of [3], [4]. As shown in Fig. 1, the antenna system consists of 8 non-overlapping narrow beams, where each beam has a beamwidth of $\frac{360^\circ}{8}$. All beams together cover the entire (360°) azimuth plane. The antenna can implement various multi-beam patterns by turning on/off some subset of the beams. When all beams of the antenna are turned on, it acts as an omnidirectional antenna. An idle node uses the omnidirectional mode to listen to all its beams for detecting the signals. When a signal arrives or a communication is set up, the node turns the antenna into the directional mode by using its beam in the direction that has the maximum signal strength.

IV. Spectrum Decision Scheme

In this section, we describe our spectrum decision scheme, followed by the description of channel quality, packet drop rate, and the utility function, which measures the overall benefit of selecting a candidate channel for transmission. The spectrum decision scheme described in this section is myopic but it considers a comprehensive WMN metric, including the link load, channel quality, node position, beam orientation, antenna-caused deafness and capture effects, interference to neighboring nodes, and application priority level (i.e., QoS). It serves as the baseline for our proposed AL-based scheme which is described in Section VI.

A. Channel Selection

Proposed spectrum decision scheme effectively exploits the spatial separation from the directional antennas and the frequency separation in the frequency domain. A node chooses the channel according to a utility function, which measures the the impact of interference among nodes, traffic load, channel quality (i.e., packet error rate, PER), and the position of the interfering node(s), on the system performance.

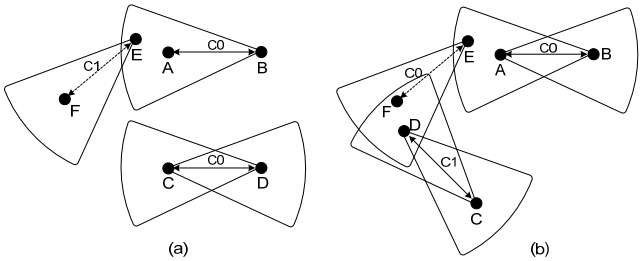


Fig. 2. Channel assignment schemes. Double arrows with dash line show the nodes that are trying to set up their connections. Double arrows with solid line show the nodes that have already set up a link.

The node needs to know the channel usage information in its vicinity in order to select a channel with the best utility value. The node also needs to know the position and the beam direction of its neighbors in order to determine whether it is located in their interference range. Each node will periodically broadcast its local channel usage information to its neighbors.

An example explaining the scheme is shown in Fig. 2. We consider the scenario with two channels ($N = 2$) which are ranked based on the proposed utility function. The channel is denoted as C_k, where $k = 0, ..., N-1$, and C_0 (C_{N-1}) has the best (worst) quality. There are six nodes with directional antennas, and two links ($A \longleftrightarrow B$ and $C \longleftrightarrow D$) are in communication. Assume that node E, which is located in the interference range of a beam of B, wants to set up its connection with node F. E first finds a free channel with the best utility value from the channel list. Since channel C_0 is already used by the link ℓ_{AB} and E is in the interference range of the link ℓ_{AB}, E needs to check the availability of the second best channel. In Fig. 2(a), our proposed scheme selects channel C_1 for communication between E and F. However, the link ℓ_{CD} between C and D can select channel C_0 because C and D are not located in the interference range of the link ℓ_{AB}. Thus, with the directional antennas in A, B, C, D, the links ℓ_{AB} and ℓ_{CD} can simultaneously select the same best channel C_0. Therefore, our proposed scheme can effectively exploit the spatial separation from directional antennas as well as the frequency separation from multiple channels to improve the network performance.

If E and F cannot find a free channel, our proposed scheme uses the best available channel as determined by its utility function. For example, in Fig. 2(b), channel C_0 and channel C_1 are used by the links ℓ_{AB} and ℓ_{CD}, respectively. Since, no free channel is available for E and F, our algorithm uses channel C_0 for E and F which has the best utility value. After, the best channel is selected, the node performs the RTS/CTS handshake to set up its connection. Meanwhile, the beams not used for transmissions are turned off to reduce the interference from the sidelobes and backlobes.

Thus our proposed scheme can effectively avoid the interference from sidelobes by prohibiting a node from selecting the channel(s) which are being used by its interfering neighbor nodes. This is very important in directional communication [23]. Moreover, we can use the the exchanged node positions and beam directions to track neighbors' directions. Therefore,

it can alleviate the deafness as well as the hidden-terminal problems.

B. Channel Quality

We use PER to represent the channel quality. Let $PER_{ij}^{(k)}$ denote the PER of channel C_k for the link ℓ_{ij}. For a given SINR, the PER can be approximated by a sigmoid function [5], [24], [25] as

$$PER_{ij}^{(k)} = \frac{1}{1+e^{\eta(SINR_{ij}^{(k)}-\sigma)}}, \tag{1}$$

where σ and η are constants corresponding to the coding and modulation schemes for a given packet length.

TABLE I
PARAMETERS FOR QUEUEING ANALYSIS

Symbol	Meaning
$\lambda_m^{(k)}$	Arrival rate of a user with priority m at channel k
$\mu_m^{(k)}$	Service rate of a user with priority m at channel k
$E[X_m^{(k)}]$	First moment of service time for a user with priority m at channel k
$E[(X_m^{(k)})^2]$	Second moment of service time for a user with priority m at channel k
$E[N_m^{(k)}]$	Average number of users with priority m in queue $Q_m^{(k)}$ at channel k
$\rho_m^{(k)}$	Normalized load of channel k due to a user with priority m, where $\rho_m^{(k)} = \lambda_m^{(k)} E[X_m^{(k)}]$
$E[W_m^{(k)}]$	Average waiting time of priority m at channel k
$R^{(k)}$	Average residual service time of the ongoing connection at channel k

C. Non-preemptive M/G/1 Queueing Model for Delay Analysis

Multimedia applications have different delay deadlines and Quality of Experience (QoE) requirements. Based on these requirements, we categorize the multimedia applications in M different priorities, where priority m = 1 (m = M) is the highest (lowest) priority. We assume that the arrival process of different data flows follows independent Poisson process and their service time is also independent.

We use a non-preemptive M/G/1 queueing model to characterize the spectrum decision behavior of each node, in which a user with lower priority is allowed to complete its service without being interrupted by a higher priority user. However, when the channel becomes idle, the higher priority user is served first. Moreover, in order to avoid the head-of-line blocking effect [26], [27], each channel maintains a separate queue for each prioritized user group. The main parameters of non-preemptive M/G/1 queueing model are listed in Table I.

Assume that the maximum transmission rate of a user with priority m over channel k is T_m^k. The first and second moments of the service time can be obtained as [25]:

$$E[X_m^{(k)}] = \frac{L_m}{T_m^{(k)}(1-PER_m^{(k)})}, \tag{2}$$

$$E[(X_m^{(k)})^2] = \frac{(L_m)^2(1+PER_m^{(k)})}{(T_m^{(k)})^2(1-PER_m^{(k)})^2}. \tag{3}$$

where L_m is the average packet length for priority m.

According to the non-preemptive M/G/1 queueing model in [27] and the arrival rate $\lambda_m^{(k)}$, we can obtain the expected queue waiting time $E[W_m^{(k)}]$ and the average delay $E[D_m^{(k)}]$ of a user with priority m using channel k as

$$E[W_m^{(k)}] = \frac{R}{2(1-\sum_{b=1}^{m-1}\rho_b^{(k)})(1-\sum_{b=1}^{m}\rho_b^{(k)})}. \tag{4}$$

$$E[D_m^{(k)}] = \frac{1}{\mu_m^{(k)}} + E[W_m^{(k)}]. \tag{5}$$

where $R^{(k)}$ is the average residual service time of the ongoing connection at channel k, and can be represented as [27]

$$R^{(k)} = \frac{1}{2}\sum_{b=1}^{M}\lambda_b^{(k)} E[(X_b^{(k)})^2]. \tag{6}$$

A data packet with priority m will be dropped if its delay exceeds its deadline d_m. It can be shown that a packet is more likely to be dropped when it is transmitted over a channel with a higher traffic load. Let $PDR_m^{(k)}$ be the probability of a packet being dropped during its transmission. It equals the probability of the average delay $E[D_m^{(k)}]$ being larger than $d_m - Delay_m^{(k)}$, where $Delay_m^{(k)}$ is the current delay of the packet. Then we can calculate $PDR_m^{(k)}$ as [28], [29]

$$PDR_m^{(k)} = \begin{cases} \rho_m^{(k)} exp(-\frac{\rho_m^{(k)} \times (d_m - Delay_m^{(k)})}{E[D_m^{(k)}]}) & \text{if } \rho_m^{(k)} < 1 \\ 1 & \text{if } \rho_m^{(k)} \geq 1 \end{cases}, \tag{7}$$

where $\rho_m^{(k)}$ is the normalized load of channel C_k caused by priority m applications.

D. Utility Function

We represent the overall metric used for selecting a channel for a link between nodes i and j, as a utility function, which is defined as,

$$U_{ijm}^{(k)} = \omega_1 * PER_{ijm}^{(k)} + \omega_2 * PDR_{ijm}^{(k)}. \tag{8}$$

Where the weight ω_1 represents the relative importance of channel quality, and ω_2 represents the channel traffic load and delay. Different users may have different preferences for values of ω_1 and ω_2 as some applications may have stringent delay constraints whereas others may need higher network throughput (and hence care more about the channel quality than delay). In our experiments, we set $\omega_1 = 0.3$ and $\omega_2 = 0.7$ since our video streaming applications are delay sensitive. More discussion about the optimal values of ω_1 and ω_2 based on the applications can be found in [30].

In order to evaluate the utility function of a channel, each node broadcasts the following information to its neighbors: 1) The channel status, which specifies whether a channel is used or not; 2)The node position and its beam direction denoted as the vector Υ, which is used to determine whether a node is located in the interference range of its neighbor nodes; 3)The channel quality represented by its PER; 4) The traffic load $\Re$. These attributes are stored at each node for future channel assignments.

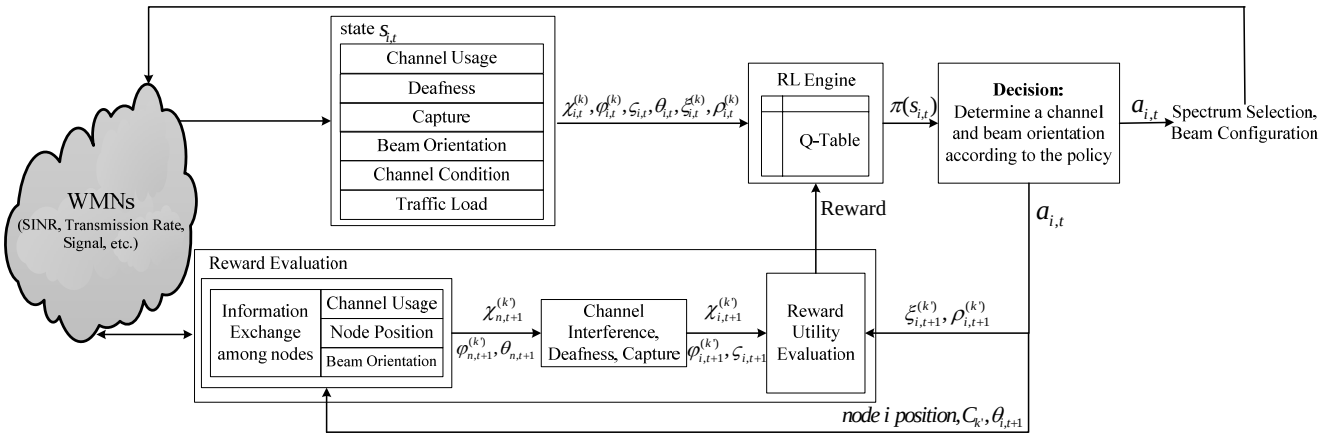


Fig. 3. The system diagram of the proposed RL-based spectrum decision scheme for directional WMNs (algorithm viewpoint).

Next, we describe a video quality metric to evaluate the impact of spectrum decision on the video quality at the receiver. The peak signal-to-noise ratio (PSNR) is widely used to measure the video quality on the receiver side. However, computing the video PSNR requires complete decoding of the real-time video at the receiver. Therefore, the use of PSNR as the spectrum selection metric is not realistic due to its heavy computational complexity and the resulting delay. We use a low-complexity and widely-used QoE metric (known as the mean opinion score (MOS)) to represent the video quality [5], [31]. The value of MOS is in the range of 1 to 5. In general, the higher the MOS, the higher the PSNR is. The MOS can be calculated as [31]

$$MOS_{j,i}^{(k)} = \frac{\tau_1 + \tau_2 FR + \tau_3 ln(SBR)}{1 + \tau_4 (TPER_j^{(k)}) + \tau_5 (TPER_j^{(k)})^2}. \quad (9)$$

where $TPER_j^{(k)} = PER_j^{(k)} + PDR_j^{(k)} - PER_j^{(k)} \cdot PDR_j^{(k)}$ is the estimated total PER of channel k for node j. The coefficients τ_1 , τ_2 , τ_3 , τ_4 , τ_5 can be obtained by a linear regression analysis [31]. In our scheme, we focus on the analysis of MOS as a function of the expected spectrum decision, while assuming that other parameters of MOS including sender bitrate (SBR) and frame rate (FR) are fixed.

The above approach uses a myopic spectrum decision strategy since it greedily chooses a channel with the maximum utility value, without considering the time-varying radio environment in WMNs and the impact of the current decision on future states. RL [5] can overcome this drawback by allowing a node to adaptively select a channel under dynamic channel conditions in each phase. Thus, it can achieve asymptotically optimal long-term reward by considering the maximum value of the cumulative rewards for all phases of spectrum decision. We use RL to define specific states, actions, and rewards, in order to solve the spectrum decision issue. A brief introduction of RL is given in the next section.

V. Overview of RL-based Spectrum Decision

The RL-based spectrum decision scheme used in this paper is based on our previous work [5]. The RL-based spectrum decision model uses the Markov decision process (MDP) [32], which can be described by a tuple of parameters: (S, A, T, R). Here S denotes the set of system states; A is the set of candidate actions at each state; $T = \{P_{s,s'}(a)\}$ is the set of state transition probabilities, where $P_{s,s'}(a)$ is the state transition probability from state s to s' when taking action a in state s. $R : S \times A \mapsto \Re$ is the reward function, which specifies the reward or cost at state $s \in S$ when taking action $a \in A$.

1) States: For a node i, its state before the $(t+1)^{th}$ spectrum assignment can be represented as $s_{i,t} = \{\chi_{i,t}^{(k)}, \varphi_{i,t}, \zeta_{i,t}, \theta_{i,t}, \xi_{i,t}^{(k)}, \rho_{i,t}^{(k)}\}$, here k is the channel ID, $\chi_{i,t}^{(k)}$ represents of the idle status of channel k, and $\varphi_{i,t}$ reflect the deafness status. $\chi_{i,t}^{(k)} = 0$ means node i is NOT in the deafness direction of its next hop node, whereas $\chi_{i,t}^{(k)} = 1$ means node i is in the deafness direction of its next hop node. $\zeta_{i,t}$ tells whether node i is in the capture range of other nodes. The remaining parameters have the same meaning as before.

2) Actions: We denote $a_{i,t} = \{\alpha_{i,t}^{(k)}, \beta_{i,t}^{(k)}\} \in \mathcal{A}$ as the candidate actions of node i in state $s_{i,t}$ at its $(t+1)^{th}$ channel assignment. $\alpha_{i,t}^{(k)}$ represents which beams will be turned on for transmission after the $(t+1)^{th}$ channel assignment. $\beta_{i,t}^{(k)}$ represents the channel selection parameter, which determines the probability of selecting channel k as the transmission channel after the $(t+1)^{th}$ channel assignment.

3) Rewards: The reward R of an action is defined as the predicted reward function of multimedia transmission, for a certain channel assignment. If the transition and reward models of MDP are known, we can obtain the optimal action for each node. It enables a node to find an optimal policy $\pi^*(s) \in A$, i.e., a sequence of actions $\{a_1, a_2, a_3, ...\}$ for state s, to maximize the total expected discounted reward in the long run. The Bellman optimality equation [33] takes into account the discounted long-term reward of taking an action.

The system diagram of the proposed RL-based spectrum decision scheme is shown in Fig. 3. The node i first observes the current state $s_{i,t}$ at its t^{th} spectrum decision. Based on the current state, the node will choose an action for spectrum decision. After the selected action is performed, the node transits to a new state. Meanwhile, the node can calculate the reward based on the exchange information from its neighbors and its current state. For example, based on the neighbors' position, beam direction, and channel usage, the node can determine the idle status of a channel. From the traffic load of the channels, the node can calculate the waiting time before

being served again. The node uses the calculated reward to update its policy. Then it repeats the same process.

VI. Apprenticeship Learning based Performance Speed-Up

Due to the complex and dynamic nature of channel assignment in directional WMNs, the learning process of RL is slow, especially in the startup phase or when a node experiences unexpected changes. Moreover, it is difficult to define an explicit reward function to exactly represent the different action-state pairs due to the large state space of directional WMNs. In this section, we describe how to use the knowledge from expert nodes to expedite the learning process of the apprentice node. Besides, the apprentice node is able to learn from the expert node via expert demonstration with the unknown reward function.

The AL algorithm is adapted from [34], but we have used it for a different application and also defined different states, actions, and rewards function. The issues of ***HOW*** to choose the teachers (i.e., expert nodes) and ***WHEN*** to learn from the teachers were not addressed in [34]. These issues have been addressed in our paper as discussed below.

A. When to Learn?

We use the AL for only the following two cases by taking into account the tradeoff between the learning rewards and the overhead. (*i*) A new node joining the network can learn from its neighboring expert node which has already acquired the optimal polices. (*ii*) The performance of a node may degrade (such as decrease in the link SNR below a threshold) due to changes in the network (such as the node mobility or changes in the beams). As a result, the current channel occupied by the node may become unavailable because this node may move into the interference range of other nodes currently using this channel. Our scheme can be used by this node to learn from its neighboring expert node about how to act in its new environment.

B. How to Select a Teacher or Expert Node?

The node should select the most suitable expert node, based on the level of expertise and the impact that its actions may have on the environment. We assume that the expert node is already stable and has a maximum-reward function, which is a linear combination of known features. We use three types of information (i.e., channel condition, node statistics, and application information), shown in Table II to evaluate the similarity between the expert and apprentice nodes.

Since the log database that stores the expert node's radio condition data could be huge (because many records are buffered and there are multiple attributes for each channel), we need to deal with high-dimensional data. We propose to use the manifold learning (as shown in Fig. 4) to extract the past channel features from each beam's records. The manifold learning uses Bregman Ball models (Fig. 4) to extract different types of patterns from the database. We define a Bregman ball [35]–[37] as the minimum manifold with a central μ_k, and a

TABLE II
Parameters for Searching the Expert Node

Layer	Symbol	Parameters
Channel Statistics	y_1	Bandwidth
	y_2	RSSI
	y_3	CINR
	y_4	Code Rate
	y_5	BER
Node Statistics	y_6	Modulation Schemes
	y_7	Available Data Rate
Application Statistics	y_8	Application Data Rate
	y_9	Delay Constraint

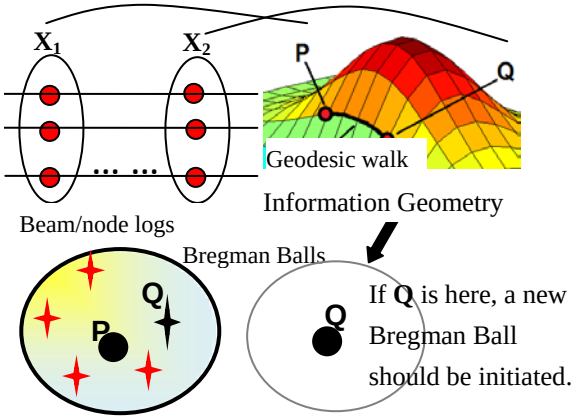


Fig. 4. Information Geometry based node-to-node teaching.

radius R_k. Any data point X_t at time t, which is inside this ball, has a strong statistical similarity (or, small signal distortion) with the central μ_k. That is:

$$B(\mu_k, R_k) = \{X_t \in X : D_\Phi(X_t, \mu_k) \leq R_k\}. \tag{10}$$

Here $D_\Phi(p, q)$ is well-known *Bregman divergence*, which is defined as the manifold distance between two signal points p and q (both are probability distribution mass values in a manifold space of X). Such a distance calculation is associated with a strictly convex and differentiable generator function $\Phi()$:

$$D_\Phi(p, q) = \Phi(p) - \Phi(q) - \langle \nabla\Phi(q), p - q \rangle. \tag{11}$$

where $\nabla\Phi = [\frac{\partial\Phi}{\partial x_1}, \frac{\partial\Phi}{\partial x_2}, \ldots]$ is a gradient operator. Here $< *, * >$ is the inner product operation. An interesting aspect of Bregman divergence is that many types of distances could be generated from its different generator functions. For example, if $\Phi(x) = x\log(x) - x$, Bregman divergence becomes Kullback-Leibler (K-L) divergence [35]–[37].

The Manifold learning analyzes each segment (called "window") of high-dimensional data and detects similar or dissimilar data points based on their statistical distances. Such a distance is based on a geodestic distance model, which is a special K-L divergence value between two statistical variables. The central μ_k is the cluster center in each data window from the apprentice node's network context statistics. The radius R_k is the pre-set threshold of geodestic distance. It cannot be set too small since the teacher node's context may have much dissimilarity from the apprentice node. Also, it should not be too large, otherwise different contexts would be classified in the same Bregman ball.

We now describe the process to find the similarity among nodes: First, we take all nodes' information as multi-variable random signals, and map them to a manifold that uses Fisher Information (here Bregman divergence is used to approximate its value) to measure a Geodesic walking distance (it is the shortest path distance between any two manifold points). Each manifold point (Fig. 4, P or Q) could follow a particular probability distribution. Therefore, the Geodesic distance (between P and Q) reflects the similarity between two probability distributions. Such a similarity level could be strictly defined as follows:

For any two manifold points P and Q, we define a distance, called symmetric Bregman divergence (SBD) [35]–[37]:

$$Dis(P,Q) = [D_{\Phi-L}(P,Q) + D_{\Phi-R}(Q,P)]/2. \quad (12)$$

Where $D_{\Phi-L}$ is left-type Bregman divergence, and $D_{\Phi-R}$ is right-type Bregman divergence [35]–[37]. If SBD is less than a pre-defined threshold, we say P and Q are similar to each other. The reason for using such a SBD-based distance metric is because general *single-type* Bregman divergence does not meet the conditions of *Symmetry* (i.e., $dis(y,x) = dis(x,y)$) and *Triangle Inequality* (i.e., $dis(x,y) \leq dis(x,z) + dis(z,y)$).

Next, based on the node similarity level between a new node point (say, Q), and a central point (say, P), Q is regarded as a part of the Bregman ball of P if they are similar to each other; otherwise, Q is used as a center and a new Bregman ball is initiated. Fig. 4 illustrates such an idea. Each ball actually represents a "cluster" from classification viewpoint. Thus we have used the information geometry theory to achieve both segmentation and clustering.

C. How to Learn?

To avoid the overhead, only those state-action pairs of the Q-table that occur frequently should be sent to the apprentice node. In AL, an expert node has the optimal policy π_E and its state-action pairs. This optimal policy can be found according to some unknown reward function R^*. The goal of AL is to have the apprentice node learn this reward function by sampling the trajectory of the expert node. In this paper, we adopt the AL [34] to learn a policy $\widetilde{\pi}$ that can achieve as good performance as the expert policy π_E under the unknown reward function R^*.

We assume that, in directional WMNs, there exists a vector of features $\phi(s)$ over states and the "true" reward functions $R^*(s)$ which belong to some hypothesis space $\mathcal{H} = \{(w^*)^T\phi(s), w \in \mathbb{R}^k\}$. We impose the constraint $w^* < 1$ as in [34], to ensure that the rewards are bounded by 1. With the vector w*, we can specify the relative weights among different system states, which may have different effect on the network performance. A policy π denotes the probability of an agent taking the action a given the state s. Therefore, we can redefine the value function $V^\pi(s)$ of a policy π as [34]:

$$\begin{aligned} V^\pi(s) &= E[\sum_{t=0}^{\infty} \gamma^t R(s_t)|\pi] \\ &= E[\sum_{t=0}^{\infty} \gamma^t w^T \phi(s_t)|\pi] \\ &= w^T E[\sum_{t=0}^{\infty} \gamma^t \phi(s_t)|\pi]. \end{aligned} \quad (13)$$

where the expectation is achieved by starting from an initial state s_0, and taking the actions according to the policy π for a random state s_i. We express the second term of the value function in (13) as $\mu(\pi)$, the so-called feature expectations. Therefore, the function can be rewritten as $V^\pi(s) = w^T\mu(\pi)$. From the above equation, we can see that the apprentice node in directional WMNs can perform as well as the expert node with an optimal policy π_E if we find a policy $\widetilde{\pi}$ for the apprentice node such that $\|\mu_E - \mu(\widetilde{\pi})\|_2 \leq \epsilon$, where ϵ is a small value with $\epsilon > 0$.

The use of a high-similarity value would require more calculations in the apprentice node. We set ϵ to be 0.01 in order to achieve a good balance between the computational complexity and the accuracy. The AL algorithm used in our channel assignment scheme is described in Algorithm 1 (Part 2) [34]. It was shown in [34] that AL algorithm can converge in a small number of iterations, and the apprentice node can find a policy that can achieve the performance close to that of the expert node, even when the algorithm cannot correctly recover the expert's reward function.

The proposed AL-based spectrum decision scheme is discussed in Algorithm 1. The apprentice node broadcasts its state information through the HELLO messages in the all-beam-on mode. Once the neighbor nodes within communication range receive the broadcast messages, they will determine their similarity with the apprentice node by using the manifold learning as described in Section VI-B. Every neighbor node sends the similarity score to the apprentice node. The apprentice node ranks its neighbor nodes according to their similarity scores and chooses the one with the highest score as the expert node. Then AL is performed for this apprentice node to learn the policy for its spectrum decision.

VII. Experimental Results

In this section, we compare the performance of our proposed spectrum decision scheme with the CI-based scheme which uses only the channel interference (CI) as the metric for spectrum decision. Both schemes also use the multi-beam antenna and multiple channels. However, the CI-based scheme does not consider the traffic priority. Four types of multimedia applications are considered in our experiments, which are prioritized according to their delay deadline requirements.

1) Video with very low delay constraint of 500ms is assigned the highest priority. The 420p resolution, whale_show video sequence is used in this application, encoded at 10 frames per second and 512Kbps bit rate. The video PSNR is 31.4dB.

2) Live streaming with delay constraint of 2s is assigned the 2nd highest priority. The 720p resolution, park_joy video sequence is used in this application, encoded at 30 frames per second and 2Mbps bit rate. The video PSNR is 34.9dB.

3) Video on demand with delay constraint of 4s is assigned the 3rd highest priority. The 720p resolution, old_town sequence is used in this application, encoded at 30 frames per second and 2Mbps bit rate. The video PSNR is 34.3dB.

4) File download application, which is best effort, is assigned the lowest priority. No packet is dropped due to the long delay deadline.

Algorithm 1 The AL-based spectrum decision scheme.

Part 1:

Input: channel statistics, node statistics, application statistics

Output: The best policy $\pi(s,a)$ for node i

1) **if** node i is a new node
2) Perform AL algorithm in Part 2.
3) **elseif**
4) Determine the status of channel k used by node i based on node's position and beam orientation.
5) Calculate channel PER using (1).
6) Calculate the expected queue waiting time $E[W_j^{(k)}]$ using (4).
7) Calculate the average delay $E[D_j^{(k)}]$ using (6).
8) Calculate the PDR using (7).
9) //Calculate the utility function
10) $U_{ij}^{(k)} = \omega_1 * PER_{ij}^{(k)} + \omega_2 * PDR_{ij}^{(k)}$.
11) **if** $U_{ij}^{(k)}$ is less than a predefined threshold
12) //The performance of node i is worse
13) Perform AL algorithm in Part 2.
14) After learning from the expert node, node i performs RL by itself.

Part 2:

Input: channel statistics, node statistics, application statistics

Output: The learning policy of node i

1) Initialize $Q(s,a)$ arbitrarily.
2) Exchange info. among node i and its neighbors.
3) Use manifold learning to find the expert node.
4) Transfer the frequently occurring state-action pairs from expert node to node i.
5) **Repeat**
6) Node i randomly selects an initial policy $\pi^{(0)}$.
7) Compute $\mu(\pi^{(0)})$ and set $j = 1$.
8) Solve $max_{t,w} t^{(j)}$ to find w, $s.t.$ $w^T\mu_E \geq w^T\mu^{(k)} + t^{(j)}, k = 0,\ldots,j-1$ and $\|w\|_2 \leq 1$.
9) **if** $t^{(j)} \leq \epsilon$
10) Terminate the AL learning process.
11) Select a policy from a set of policies $\{\pi^{(0)},\ldots\pi^{(j-1)}\}$ which is closet to the expert policy μ_E as in [34].
12) **elseif**
13) Slove the MDP by using RL algorithm and the reward function $R^{(j)}(s) = (w^{(j)})^T\phi(s)$.
14) Obtain the optimal policy $\pi^{(j)}$.
15) Compute the feature expectation $\mu^{(j)} = \mu(\pi^{(j)})$.
16) Set $j \leftarrow j+1$.

The video sequences were encoded using H.264/AVC JM reference software [38] at the transmitter.

We carried out our simulations using Matlab. We assume that IEEE802.11 MAC layer and a proactive routing scheme (such as dynamic source routing or DSR) are used for searching the teacher (or expert) nodes in the neighborhood. Physical layer uses OFDM-based multi-carrier modulation. The WMN has 30 nodes, which are randomly deployed in a 5000mx5000m area. Each node is equipped with directional antenna and has mobility ($< 10m/s$). The antenna beamwidth is $45°$, and ON (OFF) duration of each beam is 1s (3s). The number of channels is 10 and the maximum transmit rate of each channel is 3Mbps. The PER of a channel is chosen randomly from 2% to 10%.

1) Effect of the Traffic Load: In this section, we study the performance of both spectrum decision schemes for various traffic loads. Fig. 5 shows the average traffic delay as a function of per node traffic load. The traffic delay (when the application priority is ignored) in our proposed scheme is lower than the the CI-based scheme. When the applications are prioritized in four levels, the applications with higher priority have more strict delay constraints. Our proposed scheme

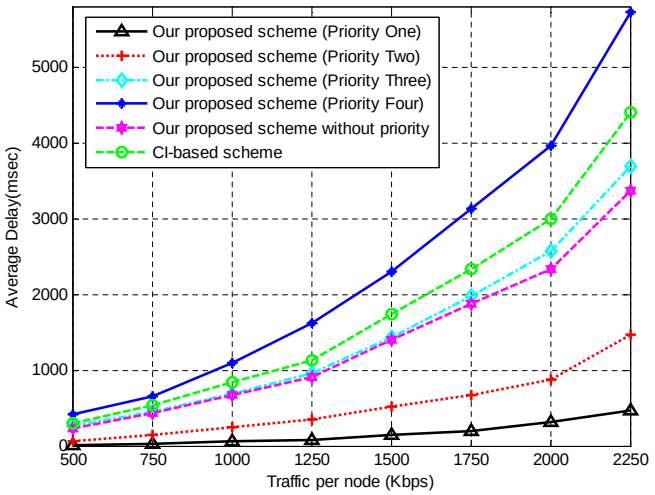


Fig. 5. The effect of traffic load on the average delay of directional WMNs for spectrum decision schemes.

is able to satisfy the delay requirements of different video applications. The higher priority applications in our scheme have much better performance than the CI-based scheme. This is because the non-preemptive M/G/1 queueing model in our scheme can provide more channel access opportunities for the higher priority applications. Moreover, our scheme considers the effect of link load, channel quality, node position, beam orientation, deafness, capture, and application priority levels, rather than selecting a channel with the lowest interference between neighboring links in the CI-based scheme.

In Fig. 5, the average delay increases with the increasing traffic load. Since the probability of a channel being busy increases for a higher traffic load, the packets have to wait for longer duration in the queue before being served. However, the delay for the higher priority (lower priority) applications in our scheme increases at slower rate (faster rate) than the CI-based scheme. Since the applications are prioritized according to their delay deadlines, our proposed scheme allocates more network resources for the higher priority applications at the cost of increasing delay of the lower priority applications.

TABLE III
COMPARISON OF TPER VALUES FOR DIFFERENT TRAFFIC LOAD PER NODE (BPS) FOR BOTH SPECTRUM DECISION SCHEMES.

	traffic	500K	750K	1M	1.25M	1.5M	1.75M	2M	2.25M
P1	**Ours**	3.6%	3.7%	3.9%	3.9%	4.1%	4.2%	4.4%	4.5%
	CI	7.5%	7.7%	8.2%	8.9%	10.1%	11.2%	12.4%	14.0%
P2	**Ours**	3.0%	3.2%	3.4%	3.8%	4.4%	4.8%	5.2%	5.6%
	CI	7.0%	7.2%	7.5%	8.0%	8.7%	9.7%	10.8%	11.9%
P3	**Ours**	2.3%	2.4%	2.7%	3.3%	3.9%	5.5%	7.3%	10.0%
	CI	6.4%	6.5%	6.8%	7.1%	7.5%	8.0%	8.4%	9.0%
P4	**Ours**	2.0%	2.0%	2.0%	2.0%	2.0%	2.0%	2.0%	2.0%
	CI	5.6%	5.6%	5.6%	5.6%	5.6%	5.6%	5.6%	5.6%

In Table III, our proposed scheme achieves much lower TPER values than the CI-based scheme. This is because the CI-based scheme does not consider the priority of the applications and selects an idle channel for the transmission without considering the channel quality and the traffic load of that channel. Another interesting observation in Table III is

that the applications with the second priority (having longer delay deadline) have a lower $TPER$ value than the applications with the first priority (having smaller delay deadline) when the traffic load is less than 1250Kbps in our proposed scheme. Similarly, the applications with the third priority have lower $TPER$ values than the applications with first and second priority when the traffic load is less than 1500Kbps. This is because less packets of a lower priority application are dropped due to their longer deadlines when the traffic load is lower. However, the $TPER$ value of a lower priority application increases faster than a higher priority application for a higher traffic load, because (i) more packets are generated by the application, leading to longer delay in the queue before being transmitted. If their waiting time exceeds their delay deadline, these packet are dropped; and (ii) our scheme allocates the best channel(s) for a higher priority application by considering the channel PER, load, etc. Also, the TPER values of the applications with the 4th priority are always very low because they do not have a delay deadline requirement.

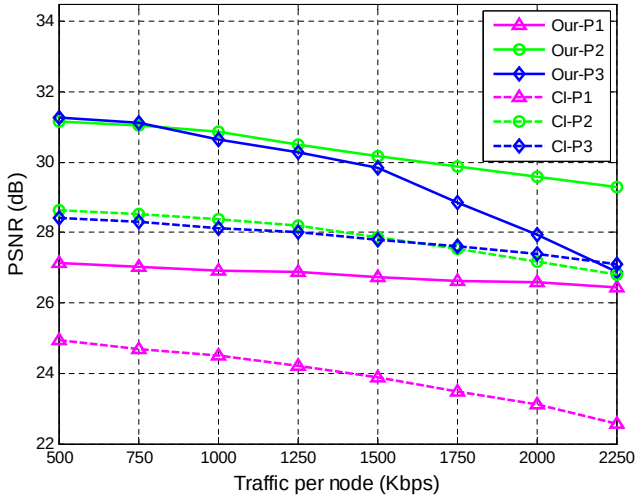


Fig. 6. The effect of traffic load on video PSNR of directional WMNs for spectrum decision schemes.

Fig. 6 shows the video PSNR of different priority applications as a function of the traffic load. The PSNR drops when the traffic increases, because many packets are dropped due to their delay deadline. Our scheme has much better PSNR performance than the CI-based scheme for the reasons described above.

2) Effect of Node Density: The maximum traffic per node is 2Mb/s and the PER of a channel is picked randomly from 2% to 10%. Fig. 7 shows the average delay as function of node density achieved by the proposed and the CI-based schemes, for the traffic of the same priority and different priorities. As expected, the average delay increases with the increasing number of nodes in the network as more traffic is generated which leads to more contention for the channel access among these nodes. However, our scheme achieves much better performance than the CI-based scheme because it considers the effect of the node deafness, capture and the traffic load. Moreover, the delay of the higher-priority applications increases at a much slower rate in our scheme than the the CI-based scheme, because the non-preemptive M/G/1 queueing model in our scheme provides more channel access to the packets of the higher-priority applications.

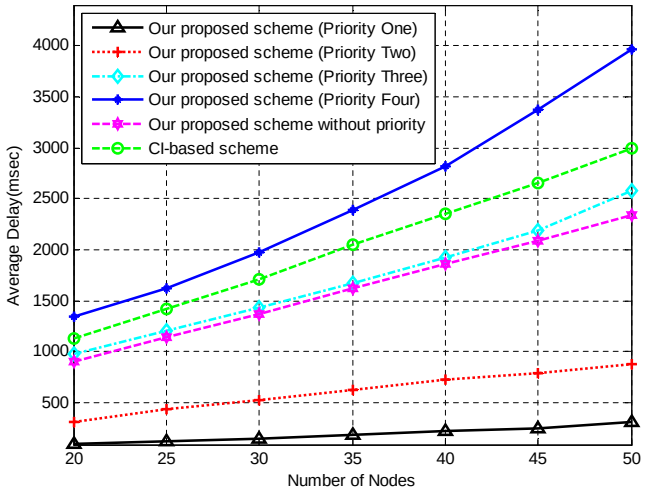


Fig. 7. Effect of node density on the average delay of directional WMNs for the spectrum decision schemes.

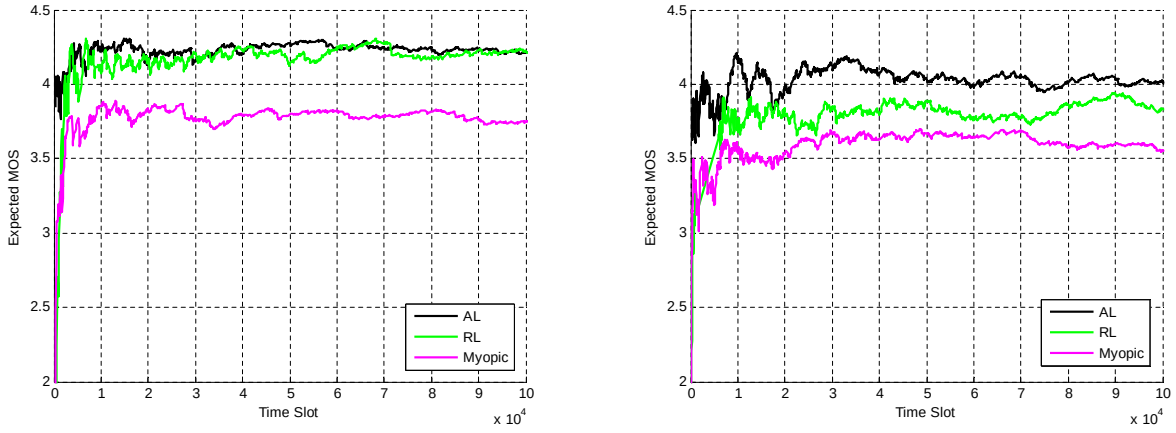
A. Comparison of AL, RL, and Myopic-based Spectrum Decision Schemes

In this experiment, we compare the performance of the AL-based spectrum decision scheme with the myopic and the RL-based schemes when (1) the mobility of the newly-joined node is very low (i.e., it does not experience widely different channel conditions); (2) The mobility of the newly-joined node is relatively high (i.e., it experiences different channel conditions with time). The AL-based scheme searches for the expert node(s) in the 1-hop neighborhood of the newly-joined node. If it cannot find an expert node, the routing or neighbor discovery scheme is used to extend the search area to find a suitable expert node.

We use the MOS metric, instead of PSNR, to represent the QoE. Fig. 8(a) shows the performance of different spectrum decision schemes for a very low mobility node. As expected, the RL-based spectrum decision scheme outperforms the myopic scheme. Furthermore, the AL-based spectrum decision scheme achieves slightly better performance than the RL-based scheme during the start-up stage, as it speeds up the learning process by allowing the newly-joined node to learn from its expert neighbor node(s). Fig. 8(b) shows the performance of different spectrum decision schemes for a relatively higher mobility node. The AL-based scheme outperforms the RL-based scheme during the start-up as well as the other time slots. This is because the AL-based scheme allows the expert nodes to transfer their policy and their frequently occurring state-action pairs to the apprentice nodes when they join the network or experience a different channel condition, which speeds up their learning process. In our experiments, we assume a different channel condition when the utility function drops below a predefined threshold.

VIII. Conclusions

We proposed an AL-based spectrum decision scheme for directional WMNs, which use multiple channels and nodes



(a) Low Mobility (b) Higher Mobility

Fig. 8. Comparison of the myopic-based, RL-based, and AL-based spectrum decision schemes.

equipped with multi-beam directional antennas. The proposed scheme performed the channel selection and handoff by considering various WMN parameters, such as the link load, channel quality, node position, beam orientation, capture, deafness, video priority, and interference among nodes. In order to achieve the optimal quality of multimedia transmission in the long term, our scheme used the RL algorithm for channel assignment. However, RL may be slow and complex due to the distributed and time-varying features of the directional WMNs. Therefore, we proposed an AL algorithm by allowing an apprentice node to learn form the expert node to expedite the learning process. In the AL algorithm, we addressed the vital questions of When/Who/What to learn from the expert nodes. Manifold learning was also used to deal with the high-dimensional data available at the WMN nodes. Our proposed learning-based channel assignment and handoff scheme achieved better performance than other schemes.

IX. ACKNOWLEDGMENT OF SUPPORT AND DISCLAIMER

The authors acknowledge the U.S. Government's support in the publication of this paper. This material is based upon work funded by AFRL, under AFRL Grant No. FA8750-14-1-0075. Any opinions, findings and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of AFRL.

REFERENCES

[1] A. Raniwala and T. Chiueh, "Architectures and Algorithms for an IEEE 802.11-Based Multi-Channel Wireless Mesh Network," in *IEEE INFOCOM*, Mar. 2005, pp. 2223–2234.

[2] R. Draves, J. Padhye, and B. Zill, "Routing in Multi-Radio Multi-Hop Wireless Mesh Networks," in *ACM MobiCom*, Sept. 2004, pp. 114–128.

[3] J. Wang, Y. Fang, and D. Wu, "Enhancing the Performance of Medium Access Control for WLANs with Multi-beam Access Point," in *IEEE Trans. Wireless Comm.*, vol. 6, Feb. 2007, pp. 556–565.

[4] Z. Chou, C. Huang, and J. Chang, "QoS Provisioning for Wireless LANs With Multi-Beam Access Point," in *IEEE Trans. Mobile Computing*, vol. 13, Sept. 2013, pp. 2113–2127.

[5] Y. Wu, F. Hu, S. Kumar, Y. Zh, A. Talari, N. Rahnavard, and J. D. Matyjas.

[6] K. Sundaresan and R. Sivakumar, "A Unified MAC Layer Framework for Ad-Hoc Networks With Smart Antennas," in *IEEE/ACM Trans. Networking*, vol. 15, June 2007, pp. 546–559.

[7] R. R. Choudhury, X. Yang, N. H. Vaidya, and R. Ramanathan, "Using Directional Antennas for Medium Access Control in Ad Hoc Networks," in *ACM MobiCom*, Sept. 2002.

[8] M. Takai, J. Martin, R. Bagrodia, and A. Ren, "Directional Virtual Carrier Sensing for Directional Antennas in Mobile Ad Hoc Networks," in *ACM MobiHoc*, June 2002, pp. 39–46.

[9] T. Korakis, G. Jakllari, and L. Tassiulas, "A MAC Protocol for Full Exploitation of Directional Antennas in Ad Hoc Wireless Networks," in *ACM MobiHoc*, June 2003.

[10] ——, "CDR-MAC: A Protocol for Full Exploitation of Directional Antennas in Ad Hoc Wireless Networks," in *IEEE Trans. Mobile Computing*, vol. 7, Feb. 2008, pp. 145C–155.

[11] H. Skalli, S. Ghosh, S. Das, L. Lenzini, and M. Conti, "Channel Assignment Strategies for Multiradio Wireless Mesh Networks: Issues and Solutions," in *IEEE Comm. Mag.*, vol. 45, 2007, pp. 86–95.

[12] H. Dai, K. Ng, R. C. Wong, and M. Wu, "On the Capacity of Multi-Channel Wireless Networks Using Directional Antennas," in *Proc. IEEE INFOCOM*, Apr. 2008.

[13] P. Dutta, S. Jaiswal, D. Panigrahi, and R. Rastogi, "A New Channel Assignment Mechanism for Rural Wireless Mesh Networks," in *Proc. IEEE INFOCOM*, Apr. 2008.

[14] A. Capone, I. Filippini, and F. Martignon, "Joint Routing and Scheduling Optimization in Wireless Mesh Networks with Directional Antennas," in *Proc. IEEE ICC*, May 2008, pp. 2951–2957.

[15] W. Zhou, X. Chen, and D. Qiao, "Practical Routing and Channel Assignment Scheme for Mesh Networks with Directional Antennas," in *Proc. IEEE ICC*, May 2008, pp. 3181–3187.

[16] V. Bukkapatanam, A. A. Franklin, and C. S. R. Murthy, "Using Partially Overlapped Channels for End-to-End Flow Allocation and Channel Assignment in Wireless Mesh Networks," in *Proc. IEEE ICC*, June 2008, pp. 1–6.

[17] M. Dohler, L. Giupponi, A. Galindo-Serrano, and P. Blasco, "Docitive Netwroks: A Novel Framework Beyond Cognition," in *IEEE Comm. Society, Multimedia Comm. TC, E-Letter*, Jan. 2010.

[18] A. Galindo-Serrano, L. Giupponi, P. Blasco, and M. Dohler, "Learning from Experts in Cognitive Radio Networks: The Docitive Paradigm," in *Proc. Int. Conf. CrownCom*, June 2010, pp. 1C–6.

[19] A. Galindo-Serrano, L. Giupponi, and M. Dohler, "Cognition and Docition in Ofdma-Based Femtocell Networks," in *Proc. IEEE GlobeCom*, Dec. 2010, pp. 1C–6.

[20] L. Giupponi, A. M. Galindo, P. Blasco, and M. Dohler, "Docitive Network-An Emerging Parading for Dynamic Spectrum Management," in *IEEE Wireless Comm. Magazine*, vol. 17, Aug. 2010, pp. 47–54.

[21] L. Giupponi, A. Galindo-Serrano, and M. Dohler, "From Cognition to Docition: The Teaching Radio Paradigm for Distributed & Autonomous Deployments," in *J. Computer Comm.*, vol. 33, 2010, pp. 2015–2020.

[22] I. F. Akyildiza, X. Wang, and W. Wang, "Wireless Mesh Networks: A Survey," in *Computer Networks*, vol. 47, Mar. 2005, pp. 445–487.

[23] S. M. Das, H. Pucha, D. Koutsonikolas, Y. C. Hu, and D. Peroulis, "DMesh: Incorporating Practical Directional Antennas in Multi-Channel Wireless Mesh Networks," in *IEEE J. Selected Areas Comm.*, vol. 24, Nov. 2006, pp. 2028–2039.

[24] D. Krishnaswamy, "Network-assisted link adaptation with power control and channel reassignment in wireless networks," in *3G Wireless Conf.*, 2002, pp. 165–170.

[25] H.-P. Shiang and M. van der Schaar, "Online Learning in Autonomic Multi-hop Wireless Networks for Transmitting Mission-Critical Applications," in *IEEE J. Selected Areas Comm.*, vol. 28, June 2010, pp. 728–741.
[26] J. Wang, H. Zhai, and Y. Fang, "Opportunistic Packet Scheduling and Media Access Control for Wireless LANs and Multi-hop Ad Hoc Networks," in *IEEE WCNC*, vol. 2, Mar. 2004, pp. 1234–1239.
[27] D. Bertsekas and R. Gallager, *Data Networks*. Upper Saddle River, NJ: Prentice Hall Inc., 1987.
[28] T. Jiang, C. K. Tham, and C. C. Ko, "An Approximation for Waiting Time Tail Probabilities in Multiclass Systems," in *IEEE Commun. Lett.*, vol. 5, Apr. 2001, pp. 175–177.
[29] H.-P. Shiang and M. van der Schaar, "Queuing-Based Dynamic Channel Selection for Heterogeneous Multimedia Applications over Cognitive Radio Networks," in *IEEE Trans. Multimedia*, vol. 10, Aug. 2008, pp. 896–909.
[30] S. Sabri and B. Prasada, "Video Conferencing Systems," in *Proc. IEEE*, vol. 73, Apr. 1985, pp. 671–C688.
[31] A. Khan, L. Sun, E. Jammeh, and E. Ifeachor, "Quality of Experience-Driven Adaptation Scheme for Video Applications over Wireless Networks," in *IET Communications*, vol. 4, Jul. 2010, pp. 1337–1347.
[32] M. L. Puterman, *Markov Decision Process: Discrete Stochastic Dynamic Programming*. John Wiley & Sons, Inc. New York, 1994.
[33] R. S. Sutton and A. G. Barto, *Reinforcement Learning: An Introduction*. Cambridge, MA: MIT press, 1998.
[34] P. Abbeel and A. Ng, "Apprenticeship Learning via Inverse Reinforcement Learning," in *Proc. 21st International Conf. Machine Learning (ICML), ACM*, 2004, p. 1.
[35] A. Banerjee, S. Merugu, I. S. Dhillon, and J. Ghosh, "Clustering with Bregman Divergences," in *Journal of Machine Learning Research*, vol. 6, Jan. 2005, pp. 1705–1749.
[36] F. Nielsen and R. Nock, "Sided and Symmetrized Bregman Centroids," in *IEEE Trans. Inf. Theor.*, vol. 55, June 2009, pp. 2882–2904.
[37] S. Mahmoodi and B. Sharif, "Signal segmentation and denoising algorithm based on energy optimization," in *Signal Processing*, vol. 85, Sept. 2005, pp. 1845–1851.
[38] JVT, H.264/AVC reference software JM14.2, ISO/IEC Std. [Online]. Available: http://iphome.hhi.de/suehring/tml/download/.

Yeqing Wu received his Ph.D. degree from the Department of Electrical and Computer Engineering, University of Alabama, Tuscaloosa, USA, in 2015. He received his M.S. degree from the Department of Electronic Engineering, Shanghai Jiao Tong University, Shanghai, China, in 2008. His research interests include cognitive radio networks, video codec, image processing, machine learning, rateless codes, and multimedia transmissions.

Fei Hu is currently a professor in the Department of Electrical and Computer Engineering at the University of Alabama, Tuscaloosa, USA. He obtained his Ph.D. degrees at Tongji University (Shanghai, China) in Signal Processing (in 1999), and at Clarkson University (New York, USA) in Electrical and Computer Engineering (in 2002). He has published over 200 journal/conference papers, books, and book chapters.Dr. Hu's research has been supported by NSF, Department of Defense (DoD), Cisco, and Sprint. His research interests are in 3S (Security, Signals, and Sensors) and wireless networks.

Sunil Kumar received his Ph.D. from Birla Institute of Technology and Science, Pilani (India) in 1997. Currently, he is a Professor and Thomas G. Pine Faculty Fellow in the Electrical and Computer Engineering Department at San Diego State University, CA, USA. His research interests include the QoS-aware and cross-layer protocols for wireless networks, and robust video compression (including H.264 and HEVC). He has published over 135 peer-reviewed journal/conference papers and four books. His research has been funded by NSF, DOD, DOE, California Energy Commission, and industry.

John D. Matyjas received his Ph.D. in Electrical Engineering from State University of New York at Buffalo in 2004. Currently, he is serving as the Technical Advisor for the Computing and Communications Division at the Air Force Research Laboratory (AFRL) in Rome, NY. His research interests include dynamic multiple-access communications and networking, spectrum mutability, statistical signal processing and optimization, and neural networks. He serves on the IEEE Transactions on Wireless Communications Editorial Advisory Board.

Dr. Matyjas is the recipient of the 2012 IEEE R1 Technology Innovation Award, 2012 AFRL Harry Davis Award for "Excellence in Basic Research," and the 2010 IEEE Int'l Communications Conf. Best Paper Award. He is an IEEE Senior Member, chair of the IEEE Mohawk Valley Signal Processing Society, and member of Tau Beta Pi and Eta Kappa Nu.

Qingquan Sun received his Ph.D. in Electrical and Computer Engineering from The University of Alabama, Tuscaloosa, AL, USA in 2013. He is currently an assistant professor in the School of Computer Science and Engineering at California State University San Bernardino, CA, USA. He has published around 20 journal/conference papers and book chapters. His research has been supported by NSF, DOD, and other sources. His research interests include intelligent sensing, distributed computing, wireless networking, signal processing, and machine learning in cyber physical systems.

Yingying Zhu received her Ph.D. degree from the Department of Electrical Engineering, University of California, Riverside, in 2015. She received her M.S. degrees in Engineering in 2007 and 2010 from Shanghai Jiao Tong University and Washington State University, respectively. Her research interests include computer vision, pattern recognition and machine learning, image/video processing and communication.