POMDP-Based Cell Selection Schemes for Wireless Networks

Po-Hsuan Tseng, *Member, IEEE*, Kai-Ten Feng, *Senior Member, IEEE*, and Chao-Hua Huang

***Abstract*—To acquire higher network capacity, the cell selection problem is formulated based on the partially observable Markov decision process (POMDP). It observes/predicts unavailable cell loading information from those non-serving base stations and takes actions for maintaining active base station set and handover target selection. Various utility functions are designed to consider different factors in the proposed POMDP-based cell selection (POCS) schemes, including system capacity, handover time, and mobility of mobile station. With the considerations of cell load as a hidden Markov model and predicted mobility in the reward function, simulation results show that the proposed POCS schemes can outperform conventional received signal strength based and load balancing methods.**

***Index Terms*—Partially observable Markov decision process, hidden Markov model, cell selection.**

I. Introduction

To guarantee quality of service (QoS), feasible cell selection scheme can enable mobile stations (MSs) to associate with a base station (BS) that possesses the best channel and network conditions. In high-speed downlink packet access (HSDPA), fast cell selection allows MSs to maintain an active set which contains the candidate BSs, and the MS selects the serving BS (SBS) with the strongest pilot signal in the active set. Notice that MS will execute cell selection or reselection procedure to find a suitable cell after power-on or link loss according to the LTE-A system. In 802.16m, either the MS or the SBS can initiate the handover procedure and select the next SBS. The procedure of transferring MS from the connection with its original SBS to the next BS is called handover. Handover helps the MS fulfill its QoS requirements from selecting a better SBS, but the procedure of handover also causes extra time delay and communication overheads. The first stage of handover is recognized as the cell selection scheme.

There are two main categories to initiate the cell selection/handover procedure as follows. *(i) Load Balancing* [1]: The number of connections of each BS is usually unequal due to the non-uniform population distribution. When the total amount of traffic in current cell exceeds the capacity limit (a.k.a. hot spot problem), the goal of this category is to enhance user throughput by off-loading the data transmission to its neighboring cell with the load-balancing cell selection procedure. *(ii) Mobility* [2], [3]: Due to user movement, the received channel quality decreases when the MS moves out of current cell coverage. The goal of this category is to minimize the handover frequency while maintaining system capacity with proper cell selection design.

A partially observable Markov decision process (POMDP) based handoff [4] has been proposed to observe the signal strengths from each BS according to either measurement or prediction to estimate the hidden states as the SBS sequence. In this paper, a POMDP framework is modeled from a different perspective to select the next SBS based on traffic load information as a hidden Markov model (HMM), given that the traffic information of neighbor BSs is not fully known by the MS. Two different versions of the POMDP-based cell selection (POCS) schemes are proposed associated with different utility functions including system capacity and mobility of MS with handover cost. Simulation results validate that the cell load and mobility information benefits the proposed POCS schemes.

II. System Model

There are N BSs in the considered wireless network environment and MSs move randomly in the BSs coverage area. The MSs are in charge of maintaining an active set and selecting the best SBS from its active set based on different objective functions at every predefined timeslot. The number of associative BSs in an active set is M. γ_h represents the ratio of handover time cost per timeslot, i.e., $0 < \gamma_h < 1$. If the new SBS is different from the original SBS, the ratio γ_h timeslot is utilized for the MS to exchange control message with the selected BS without transmitting data. Moreover, at most L MSs are allowed to connect to a BS, and each MS obtains the traffic load information from its active set. Since the MS would not maintain an active set with all the BSs in the network, i.e., $M < N$, the traffic load information for all the BSs will not be known exactly by the MS which induces the adoption of POMDP framework.

The difference between POMDP and Markov decision process (MDP) is that the MDP requires full information while the POMDP can tolerate uncertainties and gain information from observations. In cell selection scheme, cell load for each BS is quantized into units of the MS's number. Specifically, the cell loading state $s_k(t)$ represents the number of connections of the kth BS at timeslot t, where $s_k(t) \in \mathcal{S}$, $\forall k \in \{1, ..., N\}$ and $\mathcal{S} = \{1, ..., L\}$. Since full information of cell loading is not observable, the belief state which stands for the statistical probability of cell loading state is utilized to represent the state uncertainty. The belief state vector of the kth BS is represented as $\boldsymbol{b}[s_k(t)] = \{b_1[s_k(t)] \ \ldots \ b_i[s_k(t)] \ \ldots \ b_L[s_k(t)]\}$ where $b_i[s_k(t)]$ stands for the probability of $s_k(t) = i$ and $\sum_{i=1}^{L} b_i[s_k(t)] = \sum_{i=1}^{L} P[s_k(t) = i] = 1$.

Manuscript received December 11, 2013. The associate editor coordinating the review of this paper and approving it for publication was Z. Lei.

P.-H. Tseng is with the Department of Electronic Engineering, National Taipei University of Technology, Taipei, Taiwan (e-mail: phtseng@ntut.edu.tw).

K.-T. Feng is with the Department of Electrical and Computer Engineering, National Chiao Tung University, Hsinchu, Taiwan (e-mail: ktfeng@mail.nctu.edu.tw).

C.-H. Huang is with MediaTek Inc. (e-mail: chhuang.cm98g@nctu.edu.tw).

Digital Object Identifier 10.1109/LCOMM.2014.031414.132731

With N BSs in the network, the dimension of belief state L^N grows exponentially as the number of BSs increases, which is infeasible for practical implementation. A sub-optimal strategy is considered in this paper that each BS keeps track of its cell load and estimates the transition probabilities separately. The belief state comprises sufficient statistical information for the past history. A linear state vector denotes the probability that each BS contains number of MSs at the beginning of a slot, i.e., $\boldsymbol{\psi}[s_k(t)] = \{\psi_1[s_k(t)], \ldots \psi_i[s_k(t)], \ldots \psi_L[s_k(t)]\}$. With sufficient statistic $\boldsymbol{\psi}[s_k(t)]$, the belief is evolved as $P\{s_k(t)\ |\boldsymbol{\psi}[s_k(t)]\}$ for the kth BS. Therefore, when N BSs evolve independently, $\boldsymbol{b}[s_k(t)]$ can be obtained from $\boldsymbol{\psi}[s_k(t)]$ $\forall k \in \{1, ..., N\}$. This strategy reduces the statistical dimension of belief states from L^N to N. The sufficient statistical information and the performance loss due to dimension reduction have been discussed in [5].

In the POMDP framework, the objective for an MS is to choose the action for both the SBS selection and active set maintenance in order to maximize the reward. An MS maintains its active set $\mathcal{A}$, where the element of $\mathcal{A}$ is the index of the BS in the active set. The SBS is selected among the BSs in the active set as $a_1(t)$, where $a_1(t) \in \mathcal{A}$. For the ease of representation, the action is denoted as a vector $\mathbf{a}(t) = [a_1(t)\ \mathbf{a}_2(t)]$, where $\mathbf{a}_2(t)$ is a $1 \times M$ vector and it contains all the elements of $\mathcal{A}$ to provide active set maintenance. Based on the current cell load state $s_k(t)$ at current timeslot t and historical information of cell load $\boldsymbol{s}_k(1:t-1)$ from time 1 to $t-1$, the transition probabilities are estimated by the kth BS to describe the cell load transition. T_k represents the cell loading state transition probability in which current cell loading state of the kth BS only depends on the previous state and historical actions, i.e., $T_k[s_k(t) = i \mid s_k(t-1) = j] = P[s_k(t) = i \mid s_k(t-1) = j, \mathbf{a}(1:t-1)]$, where $\mathbf{a}(1:t-1)$ denotes actions from time 1 to $t-1$. Each BS periodically broadcasts its transition probability and only the latest T_k of the kth BS should be kept by the MS.

The observation $o_k(t)$ represents the observed number of connections for the kth BS at timeslot t, and $o_k(t) \in \mathcal{O}$, where $\mathcal{O} = \mathcal{S}$. The expression $\Omega[s_k(t) = i] = P[o_k(t) = i\ |\boldsymbol{s}_k(1:t), \mathbf{a}(1:t-1)]$ represents the likelihood of observation $o_k(t)$ from the resulting historical state $s_k(1:t)$ and action $\mathbf{a}(1:t-1)$. Providing that $s_k(t)$ is available, the observation is described as a Dirac delta function as $\Omega[s_k(t) = i] = \delta[i - s_k(t)]$ $\forall k \in \mathcal{A}$ and $\forall i \in \mathcal{S}$. On the other hand, if $s_k(t)$ is not available, the belief is predicted based on the transition model while the observation is described as $\Omega[s_k(t) = i] = 1/L$, $\forall k \notin \mathcal{A}$ and $\forall i \in \mathcal{S}$. For example, $M = N$ stands for the availability of the entire cell loading information and the POMDP model becomes equivalent to an MDP model. The belief update can be acquired by Baye's Theorem as,

$$b_i[s_k(t)] = P\{s_k(t) = i|\mathbf{o}_k(1:t), \mathbf{a}(1:t-1)\} = \frac{\Omega[s_k(t) = i]\sum_{j=1}^{L} T_k[s_k(t) = i \mid s_k(t-1) = j] b_j[s_k(t-1)]}{P[o_k(t)|\mathbf{o}_k(1:t-1), \mathbf{a}(1:t-1)]}.$$

Since the number of reduced belief state is equal to the number of BSs as N and each of the state is updated by a specific equation, it can be concluded that the time complexity is significantly reduced to $O(N)$. Furthermore, $r[s_k(t), \mathbf{a}(t)]$ is the expected reward by taking the action $\mathbf{a}(t)$ at state $s_k(t)$. The expected reward over the states is derived as

$$R\{\mathbf{b}[s_k(t)], \mathbf{a}(t)\} = \sum_{i=1}^{L} b_i[s_k(t)] \cdot r[s_k(t) = i, \mathbf{a}(t)]\ . \quad (1)$$

Consequently, the optimal policy to maximum expected reward for cell selection problem is to select the action as

$$\hat{\mathbf{a}}(t) = \mathop{\arg\max}_{\forall k \in \{1 \ldots N\}, \mathbf{a}(t)} R\{\boldsymbol{b}[s_k(t)], \mathbf{a}(t)\}\ . \quad (2)$$

III. Proposed POMDP-based Cell Selection Schemes

A. Capacity-based POCS (C-POCS) Scheme

The reward function of proposed capacity-based POCS (C-POCS) scheme is defined to achieve high system capacity:

$$r[s_k(t) = i, \mathbf{a}(t)] = \mathcal{I}_k[a_1(t)] \frac{W}{i} \log_2(1 + \Gamma_k(t))\ , \quad (3)$$

where

$$\mathcal{I}_k[a_1(t)] = \begin{cases} 1, & a_1(t) = k. \\ 0, & a_1(t) \neq k. \end{cases} \quad (4)$$

W is the total network bandwidth, and $\Gamma_k(t)$ represents the SNR for the kth BS-to-MS link at timeslot t. Notice that compared to conventional capacity objective function, the term $\frac{W}{i}$ includes the bandwidth allocation by considering the cell loading. It helps the MS avoid selecting the BS with higher load such as to decrease the occurrence of hot spot problem.

B. Mobility and Handover-based POCS (MH-POCS) Scheme

Considering the mobility of MS, the received SNR would not be fixed when the MS moves fast in the network. Therefore, not only the current reward such as capacity at timeslot t but also the future reward predicted by the mobility of MS should be taken into consideration. Assuming that beyond the close-in distance d_0, the path loss model of distance $d_k(t)$ for the kth BS-to-MS link at timeslot t [6], can be written as

$$PL[d_k(t)] = A + 10\alpha \log_{10}\left[\frac{d_k(t)}{d_0}\right] + n(t), \quad \forall d_i \geq d_0, (5)$$

where $A = 20\log_{10}\left(\frac{4\pi d_0}{\lambda}\right)$ is the decibel path loss at distance d_0; λ is the carrier wavelength and α is the path loss exponent. $n(t)$ represents the shadowing effect with normal distribution. According to the experimental results from Gudmundson model [3], the fading process of two consecutive observation samples can be expressed as

$$E[n(t-1)n(t)] = \sigma_n^2 \cdot \exp\left(-\frac{|d_k(t-1) - d_k(t)|}{\kappa}\right)\ , \quad (6)$$

where $d_k(t-1)$ and $d_k(t)$ are the distances of these two consecutive observation samples and κ represents how fast the correlation is decayed with the distance. Based on (6), when $d_k(t-1) \approx d_k(t)$, the signal variance σ_n^2 demonstrates that the fading effect between $n(t-1)$ and $n(t)$ are strongly correlated as $d_k(t-1)$ and $d_k(t)$ of two observation samples are close to each other. As a result, the path loss model for

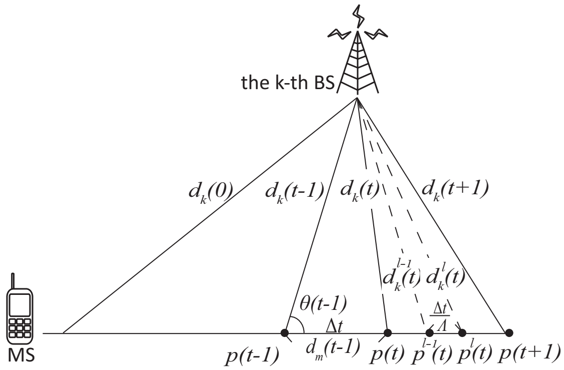

Fig. 1. Schematic diagram for the distance estimation between the MS and its associated BS.

two consecutive points can be derived from (5) as

$$PL[d_k(t)] - PL[d_k(t-1)] = 10\alpha \log\left[\frac{d_k(t)}{d_k(t-1)}\right]. \quad (7)$$

In Fig. 1, $p(t)$ denotes the position of MS at timeslot t and $p^l(t)$ represents the position of MS at the lth simulated time interval after $p(t)$. Δt is the predefined length of a timeslot. Assuming that the referenced distance $d_k(0)$ and the referenced path loss $PL[d_k(0)]$ are acquired as the empirical values while $PL[d_k(t-1)]$ and $PL[d_k(t)]$ are also obtainable at time instants $t-1$ and t. From (7), the distances $d_k(t-1)$ and $d_k(t)$ can be acquired with the known values of $d_k(0)$ and $PL[d_k(0)]$. By the law of cosines, $\theta(t-1)$ in Fig. 1 is calculated based on $d_k(t-1)$, $d_k(t)$ and $d_m(t-1)$. Assuming that the velocity and moving direction from $p(t)$ to $p(t+1)$ are the same as those from $p(t-1)$ to $p(t)$, i.e., $v(t) = v(t-1)$ and $\theta(t) = \theta(t-1)$, the MS's moving distance is also the same as $d_m(t) = d_m(t-1) = v(t-1)\cdot\Delta t$. Therefore, based on $\theta(t-1)$, $d_k(t-1)$, $d_m(t-1)$, the following steps can be utilized to predict the expected capacity from $p(t)$ to $p(t+1)$:

1) ***Define the simulated time interval*** $\frac{\Delta t}{\Lambda}$***:*** To avoid dramatic change of SNR as the MS moves, the continuous movement of MS is modeled by spliting Δt into Λ simulated time intervals, where Λ is selected to be a large number. As $\frac{\Delta t}{\Lambda}$ is small, the value of SNR changes slowly within the simulated time interval.
2) ***Predict the SNR:*** The distance from the MS's position $p^l(t)$ to the kth BS denoted as $d_k^l(t)$ is obtained by

$$[d_k^l(t)]^2 = [d_k(t-1)]^2 + \left[d_m(t-1) + l\cdot\frac{d_m(t-1)}{\Lambda}\right]^2 - 2d_k(t-1)\left[d_m(t-1) + l\frac{d_m(t-1)}{\Lambda}\right]\cos[\theta(t-1)].$$

Since the path loss from the kth BS to $p^l(t)$ as $PL[d_k^l(t)]$ can be calculated in (7) by $d_k^l(t)$, $d_k(0)$, and $PL[d_k(0)]$, the MS can therefore calculate the predicted SNR for the kth BS-to-MS link at the lth simulated timeslot after $p(t)$, i.e., the value of $\Gamma_k^l\left(t + \frac{\Delta t}{\Lambda}\cdot l\right)$.

Based on the predicted SNR $\Gamma_k^l\left(t + \frac{\Delta t}{\Lambda}\cdot l\right)$ from the kth BS, the future reward considering MS's movement can be calculated. To further address the timing cost, the procedure of handover causes transmission delay, and the transmission interrupt can influence the QoS of network communication. By combining the mobility of MS with timing cost, the reward of MH-POCS scheme is defined as

$$r[s_k(t) = i, \mathbf{a}(t)] = \{1 - \gamma_h[a_1(t)]\}\frac{W}{i}\sum_{l=1}^{\Lambda}\frac{1}{\Lambda}\log_2\left[1 + \Gamma_k^l\left(t + \frac{\Delta t}{\Lambda}\cdot l\right)\right]. \quad (8)$$

where γ_h is the penalty by conducting handover procedure as

$$\gamma_h[a_1(t)] = \begin{cases} 0, & a_1(t) = a_1(t-1) \\ \gamma_h, & a_1(t) \neq a_1(t-1) \end{cases} \quad (9)$$

If $a_1(t)$ is the same as $a_1(t-1)$, SBS will not be changed and no handover is needed, i.e., handover cost is equal to 0. On the other hand, if the selected BS is changed, handover time cost will be incurred. Therefore, it can avoid ping-pong effect, i.e., too many unnecessary handover occurrences, when MS moves around the cell edges.

TABLE 1
SIMULATION PARAMETERS

Parameter Type	Parameter Value
Separated Distance between BSs	1.5 km
Number of BSs N	20
Transmit Power of BS	46 dBm
Carrier Frequency	1.9 GHz
Bandwidth	10 MHz
Antenna Height for BS	30 m
Correlation Factor κ	50 m
Maximum number of connections L	32
Timeslot	1 second
close-in distance d_0	100 m

IV. Performance Evaluation

The simulation parameters of path loss model in (5) adopt the model parameters of terrain type B in [6]. Received signal strength (RSS) is utilized as a quality measure to indicate the link quality between BS and MS. To avoid ping-pong effect, conventional RSS-based scheme selects the new SBS when its RSS is greater than the original SBS with a hysteresis threshold Δ. Two RSS-based schemes are simulated based on different settings of Δ as follows. The parameter Δ of RSS-f is fixed as 3 dB while Δ of RSS-v varies along with the time cost of handover. When the time cost of handover increases, Δ is augmented in order not to spend too much time on handover. The relationship between γ_h and Δ is established as:

$$\log_2(1+\Gamma) = (1-\gamma_h)\log_2(1+\Gamma+\Delta). \quad (10)$$

The left-hand side of equality in (10) means the capacity of connecting to original SBS; while the right-hand side represents the capacity of connecting to the new SBS. Since there are three unknown parameters, $\Delta = 3$ dB when $\gamma_h = 0.3$ are defined to calculate the corresponding SNR values as Γ_r. To obtain the relationships of Δ and γ_h, (10) is derived as

$$\Delta = (1+\Gamma_r)^{\frac{1}{1-\gamma_h}} - 1 - \Gamma_r. \quad (11)$$

RSS-v and RSS-f are the conventional schemes without cell loading information to serve as the performance benchmarks. To validate the gain of cell loading information to the cell selection problem, an MDP case $M = N = 20$ is considered

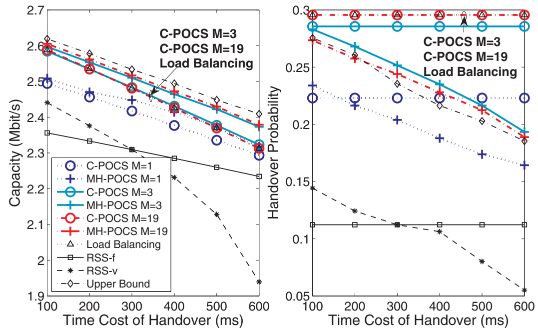

Fig. 2. The capacity and handover probability under different time costs of handover for scenario 1.

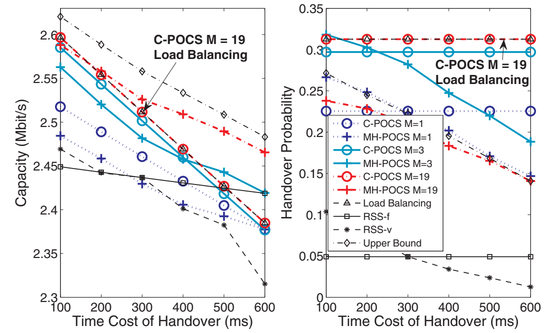

Fig. 3. The capacity and handover probability under different time costs of handover for scenario 2.

the upper bound of all POCS schemes. On the other hand, load balancing (LB) scheme in [1], which considers full information of cell load in capacity-based reward, is also compared. Without mobility prediction, LB scheme can be regarded as an MDP case of C-POCS scheme.

In scenario 1, MS moves at a constant speed $v = 12$ m/s crossing several cell edges. Since scenario 1 considers a straight line movement where the trajectory of MS is predictable, it is shown in the left subplot of Fig. 2 that MH-POCS scheme can provide higher capacity performance. C-POCS scheme outperforms conventional RSS-based methods, which demonstrates the effectiveness of HMM for cell load under a regular movement case. By considering mobility in a regular movement scenario, MH-POCS scheme can outperform LB scheme even with small number of BSs in active set, i.e., $M = 3$, achieving the tradeoff between computational complexity and capacity. Compared to LB scheme, i.e., $M = 20$, POCS schemes with $M = 3$ reduces the signaling cost significantly.

A random walk model is adopted for the MS with velocity between 9 and 15 m/s in scenario 2. Cell loading information has a significant impact on capacity when time cost of handover is small, i.e., all POCS schemes outperform the RSS-based methods in the left subplot of Fig. 3. There is no guarantee that the predicted mobility information would result in better performance. If an MS's movement follows a random walk model, the mobility prediction might even degrade the performance. As shown in the left subplot of Fig. 3, C-POCS scheme outperforms MH-POCS scheme when $M = 1$. When $M = 3$, there is an intersecting point where the MH-POCS scheme is better than C-POCS scheme when the time cost is large. This is attributed to the consideration of handover cost for MH-POCS scheme. When $M = 19$, MH-POCS scheme outperforms C-POCS and LB schemes. This demonstrates that joint considerations of mobility prediction, cell load and handover cost are necessary in the reward function design to achieve higher capacity such as MH-POCS scheme under a random walk scenario. Moreover, M should be carefully chosen based on the time cost of handover of current system.

Handover probability refers to how frequent an MS select new SBS. Ping-pong effect can be observed from handover probability. The right subplots of Figs. 2 and 3 illustrate the handover probability under different time costs of handover. There exists a crossover point at 300ms between the RSS-v and RSS-f methods due to the derivation of Γ_r from (11). C-POCS and RSS-f schemes do not consider the impact of handover time costs as their handover probabilities are not influenced by the time cost of handover. With consideration of handover time cost, handover probabilities of MH-POCS and RSS-v methods decrease as time cost of handover rises, which successfully avoid ping-pong effect. Higher handover cost penalizes the capacity and makes an MS reluctant to perform handover.

V. Conclusion

POMDP-based cell selection (POCS) methods are designed to consider both cell load and mobility to achieve higher network capacity. Capacity and mobility of mobile station with handover time are considered in the design of C-POCS and MH-POCS schemes, respectively. Higher system capacity can be achieved by C-POCS method which validates the impact of predicted information of cell load. The consideration of handover time in MH-POCS scheme decreases handover probability which can effectively alleviate the ping-pong effect. Based on the simulation results, MH-POCS scheme can outperform the other methods. A small number of BSs in the active set, e.g., $M = 3$, is suitable for the MH-POCS scheme in a regular movement case. In a random walk case, M should be designed based on the time cost of handover in the system.

References

[1] A. Sang, X. Wang, M. Madihian, and R. D. Gitlin, "Coordinated load balancing, handoff/cell-site selection, and scheduling in multi-cell packet data systems." in *Proc. 2004 ACM Mobile Computing and Networking*, pp. 302–314.

[2] N. W. Sung, N.-T. Pham, T. Huynh, and W.-J. Hwang, "Predictive association control for frequent handover avoidance in femtocell networks," *IEEE Commun. Lett.*, vol. 17, no. 5, pp. 924–927, 2013.

[3] P.-H. Tseng and K.-T. Feng, "A predictive movement based handover algorithm for broadband wireless networks," in *Proc. 2008 IEEE Wireless Communications and Networking Conference*, pp. 2834–2839.

[4] M. Halgamuge, K. Ramamohanarao, M. Zukerman, and H. Vu, "Handoff optimization using hidden Markov model," *IEEE Signal Process. Lett.*, vol. 18, no. 7, pp. 411–414, 2011.

[5] Q. Zhao, L. Tong, A. Swami, and Y. Chen, "Decentralized cognitive MAC for opportunistic spectrum access in ad hoc networks: a POMDP framework," *IEEE J. Sel. Areas Commun.*, vol. 25, no. 3, pp. 589–600, Apr. 2007.

[6] IEEE802.16m-08/004r5, IEEE 802.16m Evaluation Methodology Document (EMD), The IEEE 802.16 Working Group on Broadband Wireless Access Standards Std., 2009.