Dynamic User Demand Driven Online Network Selection

Zhiyong Du, *Student Member, IEEE,* Qihui Wu, *Senior Member, IEEE,* and Panlong Yang, *Member, IEEE*

***Abstract*—Network selection plays a key role in reaping the potential benefit of heterogeneous wireless networks. Aiming at improving the user's quality of experience, we study the network selection problem with time-varying user demand and non-uniform network handoff costs in a dynamic environment. One appealing solution in converging the time-vary user demand and the diverse network performance is dynamic network selection, which, however, poses the dilemma between satisfying user demand and controlling the network handoff cost. To get around this problem, we propose an online network selection algorithm to learn the optimal network selection policy with network handoff cost consideration. In addition, we exploit the inherent dependency in the problem and derive another two algorithms with much faster convergence speed. Simulations reveal that the proposed algorithms can achieve $10\% \sim$ performance gain over existing methods.**

***Index Terms*—Network selection, user demand, network handoff cost, online learning.**

I. Introduction

THE proliferation of wireless networking and new services has brought about the emergence of heterogeneous wireless networks, where networks with different radio access technologies and/or different coverage range coexist. The heterogeneous wireless networks can significantly improve system coverage and capacity [1]. While reaping the potential benefit of different wireless networks heavily depends on users' network access schemes, where network selection policy plays a key role. One of the challenges in network selection is network selection decision algorithms. Existing methods include Multi Attribute Decision Making (MADM) [2], fuzzy logic theory [3], neural network [4] and Markov decision process (MDP) [7]. Another challenge lies on network selection criterion. Commonly, various QoS related parameters can be the network selection criterion, such as the received signal strength [5], the available bandwidth [6] and delay [7]. Compared to the conventional QoS centered network selection, a new network selection criterion, users' quality of experience (QoE), is proposed in [9], recently.

Nowadays users commonly have various wireless applications with diverse user demand. For instance, users are sensitive to packet loss in video, while the throughput is emphasized for file transfer service [10]. In particular, the user demand can vary with the change of running applications types, which turns the time-varying user demand into a universal phenomenon. Therefore, how to satisfy the time-varying user demand poses new challenges in network selection, which is neglected in current literature.

Considering the diversity in network performance, different user demand may prefer different networks. Hence, satisfying the time-varying user demand means dynamic network selection, resulting in excessive network handoff cost. In response to this dilemma, we propose an online network selection algorithm. The proposed algorithm can learn to match different user demand with respectively optimal network selections in balancing the user QoE and network handoff cost. Moreover, we explore the inherent dependence of our problem and utilize it to derive another two online network selection algorithms with substantial improvement in convergence speed.

II. System Model and Problem Formulation

We assume that a user locates in the overlapping area of N wireless networks $\mathcal{N} = \{1, 2, ..., N\}$. In a slotted system with epoch duration l seconds, the user can dynamically change its access network at the beginning of each epoch. We seek for a network selection algorithm with the following considerations: 1)**Time-varying user demand**: Dependent on the applications, user preference, etc., the user may pose different performance requirements. Even the dominating QoS parameters in user perception may vary with time. We highlight the dynamics of user demand, which is differentiated from the fixed and general user utility functions in current literature [6][7]. 2)**Non-uniform network handoff costs**: Network handoffs definitely incur handoff cost, in terms of signaling, energy, or time. Specially, depending on the involved wireless networks' radio interfaces, network handoff costs between two networks could be different [8]. For example, the handoff cost in a handoff between two WLANs with the same standard is clearly smaller than that between a WLAN and a cellular network. 3)**Dynamic environment**: Due to the dynamics in wireless channel condition and network traffic load, networks' performance is uncertain and dynamic to users.

Obviously, consideration 1) requires that the selected network is user demand matched, i.e., selecting the optimal network corresponding to specific user demand. However, such requirement may incur frequent network handoffs due to the dynamic of user demand. Hence, consideration 1) and 2) indicate that there is a tradeoff between satisfying the user demand and controlling the handoff cost. Further, constraint of 3) implies that direct optimization may be infeasible due to the dynamic network performance.

To model this problem, we denote the user demand type set as $\mathcal{S} = \{s_1, ..., s_L\}$. We assume that the user demand is invariant during one epoch and may change between two successive epochs. Specifically, the dynamic of user demand between two successive epochs follows some unknown Markov transition

Manuscript received November 25, 2013. The associate editor coordinating the review of this letter and approving it for publication was G. Lazarou.

This work was supported by the NSF of China under Grant No. 61172062, 61003277, 61232018, 61272487, and in part by the Jiangsu Province NSF of China under Grant No. BK2011116.

The authors are with the College of Communications Engineering, PLAUST, Nanjing, 210007, China (e-mail: duzhiyong2006@163.com).

Digital Object Identifier 10.1109/LCOMM.2014.011214.132617

matrix $\mathbf{P}(l) = \{p_{s,s'}\}$, $s, s' \in \mathcal{S}$, where $p_{s,s'}$ is the one step transition probability when the user demand changes from s to s' given the epoch duration l. Different epoch durations l may result in different transition matrixes. For each type of user demand $s \in \mathcal{S}$, there is a distinct demand utility function $D_s(\mathbf{v})$ mapping the experienced QoS to users' utility, e.g., satisfaction degree or QoE, where $\mathbf{v}$ is the relevant QoS parameter vector. The network handoff cost matrix is $\mathbf{C} = \{c_{m,n}\}, m, n \in \mathcal{N}$, where $c_{m,n}$ is the cost for the handoff from network m to n. Here, $c_{m,n}$ even can be random variables with bounded mean and variance.

Denote the currently associated network at the beginning of t-th epoch by $n(t)$ and the network selection decision for t-th epoch by $\delta(t)$. Jointly considering the user's QoE and network handoff cost, the user's reward $u(t)$ is

$$u(t) = D_{s(t)}(\mathbf{v}(t)) - \lambda c_{\delta(t),n(t)} \tag{1}$$

where $\lambda \in [0,1]$ is the handoff cost weight determining the tradeoff between the QoE and handoff cost, $\mathbf{v}(t)$ and $s(t)$ are the experienced QoS parameter vector in network $\delta(t)$ and the user demand type in t-th epoch, respectively. We noticed that the relationship of the current associated network and the network selection decision is $n(t+1) = \delta(t)$. Due to the time-varying user demand $s(t)$ and dynamic $\mathbf{v}(t)$, it's reasonable to optimize the cumulative discounted expected reward R. Formally, given an initial user demand type $s(0)$ and network $n(0)$, we have to find a sequence of network selection decisions $\{\delta(0), \delta(1), ...\}$ to maximize

$$R(s(0), n(0)) = E\left[\sum_{t=0}^{\infty} \beta^t u(t) | s(0), n(0)\right] \tag{2}$$

where the discount factor $\beta \in (0,1)$ controls the future rewards' effect on the cumulative reward.

When the network environment $\mathbf{v}(t)$ and user demand $s(t)$ are stationary, we can model the network selection problem in a Markov decision process (MDP) framework. Generally, a MDP can be characterized by state, action, reward and state transition matrix. We map the problem into a MDP as follows,

- State: $\mathbf{x}(t) = [s(t), n(t)]$. The state space is $\mathbf{X} = \mathcal{S} \times \mathcal{N}$.
- Action: $\delta(t) \in \mathcal{N}$.
- Reward: $r(t) = u(t) = u(\mathbf{x}(t), \delta(t))$.
- State transition matrix: $\mathbf{P'} = \{p(\mathbf{x}'|\mathbf{x}, \delta)\}, \mathbf{x}, \mathbf{x}' \in \mathbf{X}$, $p(\mathbf{x}'|\mathbf{x}, \delta) = \begin{cases} p_{s_{\mathbf{x}'}, s_{\mathbf{x}}}, & n_{\mathbf{x}'} = \delta \\ 0, & otherwise \end{cases}$, where $s_{\mathbf{x}}$ and $n_{\mathbf{x}}$ are the corresponding user demand type and associated network in state $\mathbf{x}$, respectively.

The system state $\mathbf{x}(t)$ here is jointly determined by the user demand type and the associated network. As a result, the action can affect the state transition. It is worth noting that the action space is part of the state space, which is different from most existing MDP models. Given the finite states and actions, we define a stationary and deterministic policy $\pi \in \mathbf{\Pi}$ as a function mapping the states to actions $\mathbf{X} \to \mathcal{N}$, where $\mathbf{\Pi}$ is the space of stationary and deterministic policy. Then, the problem in (2) can be changed to find a policy satisfying

$$\pi^* = \arg\max_{\pi \in \mathbf{\Pi}} E_\pi \left[\sum_{t=0}^{\infty} \beta^t r(t) | \mathbf{x}(0)\right] \tag{3}$$

***Algorithm 1**: online network selection algorithm*

Initiate: $t = 0$, $Q(\mathbf{x}, \delta) = 0, \forall \mathbf{x} \in \mathbf{X}, \forall \delta \in \mathcal{N}$.

Loop For each t

Observe state $\mathbf{x}(t)$ and select network $\delta(t)$ according to the following rule:

- With probability ε, randomly select $\delta(t) \in \mathcal{N}$;
- Else, $\delta(t) \in \arg\max_{\delta \in \mathcal{N}} Q(\mathbf{x}(t), \delta)$.

Receive the reward $r(t)$.

The system state changes to $\mathbf{x}(t+1) = [s(t+1), \delta(t)]$

Update

$$Q(\mathbf{x}(t), \delta(t)) = (1 - \alpha_t) Q(\mathbf{x}(t), \delta(t)) + \alpha_t \left[r(t) + \beta \max_{\delta \in N} Q(\mathbf{x}(t+1), \delta)\right] \tag{4}$$

End loop

where E_π means taking expectation under policy π.

III. Proposed Network Selection Algorithms

In practical implementations, the prior user demand transition probability matrix $\mathbf{P}(l)$ and the network dynamics are unknown, making the direct optimization of (3) intractable. To get around this limitation, we resort to reinforcement learning. In the following, an online network selection algorithm based on Q-learning is proposed only relying on the interaction with the environment.

The main idea of the online network selection algorithm is learning the Q value $Q(\mathbf{x}, \delta)$ for each state-action pair, which is the estimated expected long term reward for action δ in state $\mathbf{x}$. The action for each state is selected according the updated Q values. The algorithm works as shown in Algorithm 1: At initiation, $Q(\mathbf{x}, \delta)$ is set with zero for each state-action pair. Then, for each state, the network δ maximizing the corresponding Q value is selected. Finally, the immediate reward r and the new state are used to update the Q value, where α_t is the learning rate satisfying $\sum_{t=0}^{\infty} \alpha_t = \infty, \sum_{t=0}^{\infty} {\alpha_t}^2 < \infty$.

Although the online network selection algorithm can converge to the optimal policy, it may be complex and slow to convergence. In order to achieve fair performance, each state-action pair $Q(\mathbf{x}, \delta)$ must be sampled sufficiently. In our problem, there are $|\mathcal{S}||\mathcal{N}|$ states and $|\mathcal{N}|$ actions, resulting in $|\mathcal{S}||\mathcal{N}|^2$ state-action pairs. The relatively large state-action pair space may lead to poor convergence performance of algorithm 1. The slow convergence speed may be one limitation in practical implementations. Nevertheless, we found that there exists rooms to speed up the above algorithm. The dependence among utilities in different states enables us to update a group of state-action pairs simultaneously. There are two possible cases.

Case 1

Condition: The network handoff cost $c_{m,n}, m, n \in \mathcal{N}$ is constant and known. *Result*: Given some state $\mathbf{x} = [s, n]$ and the action δ, we can update $|\mathcal{N}|$ state-action pairs $r_i = u(\mathbf{x}_i, \delta), \forall \mathbf{x}_i \in \mathbf{X}(s_{\mathbf{x}})$, where $\mathbf{X}(s_{\mathbf{x}}) = \{\mathbf{x}' | \forall \mathbf{x}' \in \mathbf{X}, s_{\mathbf{x}'} = s_{\mathbf{x}}\}$ is the set of states whose user demand type is the same with $\mathbf{x}$. Note that the differences

of $u(\mathbf{x}_i, \delta)$ for $\forall \mathbf{x}_i \in \mathbf{X}(s_{\mathbf{x}})$ only lie in the handoff cost according to (1). Thus, it makes sense to update the state-action pairs within group $\mathbf{X}(s_{\mathbf{x}})$. We define a new matrix G the same size with Q and derive online network selection 2 by replacing (4) in Algorithm 1 with the following (5) and (6),

$$G(\mathbf{x}, \delta(t)) = (1-\alpha_t) Q(\mathbf{x}, \delta(t)) + \alpha_t \left[r(t) + \beta \max_{\delta \in \mathcal{N}} Q(\mathbf{x}(t+1), \delta) \right], \forall \mathbf{x} \in \mathbf{X}\left(s_{\mathbf{x}(t)}\right) \quad (5)$$

$$Q(\mathbf{x}, \delta(t)) = G(\mathbf{x}, \delta(t)), \forall \mathbf{x} \in \mathbf{X}\left(s_{\mathbf{x}(t)}\right) \quad (6)$$

Case 2

Condition: (i) The network handoff cost $c_{m,n}, m, n \in \mathcal{N}$ is constant and known. (ii) All the QoS parameters in $\mathbf{v}$ are observable in each epoch. *Result*: In this case, we can obtain $|\mathcal{S}||\mathcal{N}|$ state-action pairs $r_i = u(\mathbf{x}_i, \delta), \forall \mathbf{x}_i \in \mathbf{X}$. Note that the above samples comprise one real sample $\mathbf{x}_i = \mathbf{x}$ and $|\mathcal{S}||\mathcal{N}| - 1$ virtual samples. The virtual samples are derived based on the fact that the difference among $u(\mathbf{x}_i, \delta)$ for $\forall \mathbf{x}_i \in \mathbf{X}$ lies in the demand utility $D_s(\mathbf{v})$ and the handoff cost. If all the QoS parameters in $\mathbf{v}$ are observable in each epoch, the group update manner is feasible. Therefore, we propose online network selection algorithm 3, by replacing (4) in Algorithm 1 with (7) and (8),

$$G(\mathbf{x}, \delta(t)) = (1-\alpha_t) Q(\mathbf{x}, \delta(t)) + \alpha_t \left[r(t) + \beta \max_{\delta \in \mathcal{N}} Q(\mathbf{x}(t+1), \delta) \right], \forall \mathbf{x} \in \mathbf{X} \quad (7)$$

$$Q(\mathbf{x}, \delta(t)) = G(\mathbf{x}, \delta(t)), \forall \mathbf{x} \in \mathbf{X} \quad (8)$$

IV. Performance Evaluation

We differentiate three types of user demand based on the running application/traffic: video traffic demand, audio traffic demand and elastic traffic demand and denote the user demand set $\mathcal{S} = \{video, audio, elastic\}$. The corresponding three demand utility functions are defined according to QoE models in [10][11]. These demand utility functions map the experienced QoS into the mean opinion score (MOS) that indicates user's satisfaction degree. The MOS is in the range of [1, 5], where a larger value means a higher satisfaction degree. The video traffic demand utility function $D_{video}(P_{snr})$ reflects the peak SNR's dominating role in the MOS of video traffic,

$$D_{video}(P_{snr}) = 4.5 - \frac{3.5}{1 + \exp[b_1(P_{snr} - b_2)]}$$

where P_{snr} is the peak SNR, b_1 and b_2 are relevant parameters. The audio traffic demand utility function is defined based on the R-factor R_f,

$$D_{audio}(R_f) = 1 + 0.035R_f + 7 \cdot 10^{-6} R_f (R_f - 60)(100 - R_f)$$

where R_f reflects the overall effect of loss e and delay d on the MOS. Note that $R_f = 94.2 - I_e - I_d$, where $I_e = \gamma_1 + \gamma_2 \ln(1 + \gamma_3 e)$ is the impairment caused by the loss, $I_d = 0.024d + 0.11(d - 177.3)\mathbf{I}_{\{d-177.3\}}$ is the impairment caused by delay, $\mathbf{I}_{\{\cdot\}}$ is the indicator function. The loss can be further divided by $e = e_{network} + (1 - e_{network}) e_{playout}$, where $e_{network}$ and $e_{playout}$ are the losses caused by the network and the playout, respectively. Similarly, the delay $d = d_{codec} + d_{playout} + d_{network}$, where d_{codec}, $d_{playout}$ and $d_{network}$ are the delay in codec, playout and network, respectively. Since the elastic traffic is mainly affected by the throughput, the demand utility is defined by

$$D_{elastic}(\theta) = b_3 \log(b_4 \theta),$$

where θ is the throughput, b_3 and b_4 are parameters determined by the maximal and minimal expected throughput thresholds of the user. In our simulation, $e_{playout} = 0.005$, $d_{playout} = 60$ms and $d_{codec} = 25$ms.

TABLE I
PARAMETER SET

	e_m	e_u	K_e	d_m	d_u	K_d	θ_m	θ_u	K_θ
LTE	0.02	0.02	3	10	10	5	250	50	6
WLAN1	0.02	0.02	5	50	10	4	450	60	4
WLAN2	0.04	0.02	5	60	10	5	250	50	4

We consider a classical cellular WLAN integrated heterogeneous wireless network, where one LTE and two WLANs are available simultaneously. Although the instant SNR changes with time, the dynamics of peak SNRs can be slow and bounded. Thus, we assume that peak SNRs of networks are fixed throughout for simplicity. In realistic wireless communications, the dynamics of QoS parameters such as the delay, loss and throughput can be very complex. According to the trace data from android application "speedtest", we found that the Markov chain model in [7] can approximately model the dynamics of QoS parameters. Specifically, the joint delay-loss-throughput state is fixed in one epoch (10 seconds duration), while state transition at successive two epochs follows fixed probability. The parameters are set as shown in Table I and the transition probabilities between any two state are assumed equal. In the table, the delay d in ms, loss e and throughput θ in kbps are all characterized by three parameters: subscript "m" denotes the minimal values, subscript "u" denotes the basic unit, K denotes the number of levels. The nominal handoff cost matrix among LTE, WLAN1 and WLAN2, the user demand transition matrix among $video$, $audio$ and $elastic$ are

$$\mathbf{C} = \begin{matrix} 0 & 2 & 2 \\ 2 & 0 & 1 \\ 2 & 1 & 0 \end{matrix}, \mathbf{P} = \begin{matrix} 0.4 & 0.3 & 0.3 \\ 0.3 & 0.4 & 0.3 \\ 0.3 & 0.3 & 0.4 \end{matrix}$$

First, we set $\lambda = 0.5$, $\beta = 0.3$ and compare the convergence performance of the proposed three algorithms. As can be observed in Fig 1, algorithm 2 and algorithm 3 can achieve much larger average rewards than the original Q-learning based algorithm 1. This indicates that the group based update can effectively promote the convergence speed. Meanwhile, algorithms 3 outperforms algorithm 2 due to its larger number of samples in the group update. Note that although these three algorithms have different reward growth rates, they will converge to the same average reward given sufficient iterations.

Second, we compare the performance of the proposed algorithms with another five algorithms. The considered five algorithms include a random selection algorithm, a matching selection algorithms and three fixed selection algorithms. The

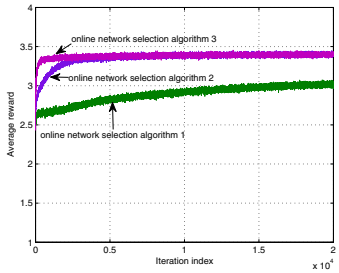

Fig. 1. Convergence performance comparison of the proposed algorithms.

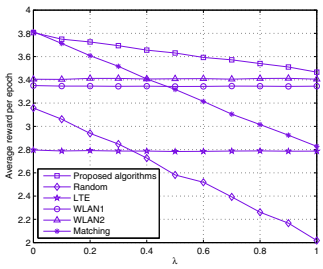

Fig. 2. Performance comparison of different algorithms with different handoff cost weights.

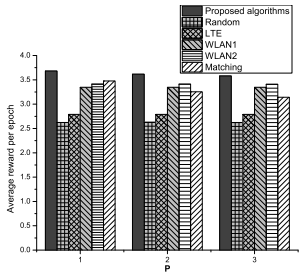

Fig. 3. Performance comparison of different algorithms with different user demand transition matrixes.

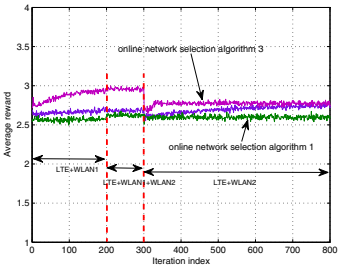

Fig. 4. Convergence performance comparison of the proposed algorithms in the mobility scenario.

random selection algorithm randomly selects a network in each epoch. The matching selection algorithm selects the optimal network according to the user demand. In our setting, matching selection means the optimal networks for video traffic demand, audio traffic demand and elastic traffic demand are WLAN2, LTE and WLAN1, respectively. The fixed selection algorithm indicates sticking to any one of networks without handoff. In Fig 2, the average rewards per epoch in 1000 runs of different algorithms are shown. As can be seen, the performance of the random selection, the matching and the proposed algorithm generally decrease as the handoff cost weight increases. In particular, since the random selection algorithm considers neither the user demand nor the handoff cost, its performance decays quickly and becomes the worst of all algorithms when $\lambda > 0.5$. On the other hand, since the matching algorithm selects the optimal networks for each type of user demand, its performance is the same with the proposed algorithms when $\lambda = 0$. However, when λ increases, its performance becomes worse than that of the proposed algorithms, due to the negative influence of handoff cost. Even though the handoff cost decays the performance gain, the proposed algorithms outperform the other five algorithms with the maximal gains over random, matching and fixed selections up to 250%, 25%, 12%. Moreover, we compare the performance of the mentioned algorithms in three different user demand transition matrixes where $\mathbf{P}_1$ represents that the user demand prefers to maintain the current type with higher probability, $\mathbf{P}_2$ represents that the user demand can be any one of three types with equal probability in each epoch, $\mathbf{P}_3$ represents that the user demand changes with higher probability. In Fig 3, we can find that although the performance of the proposed algorithms may vary in different cases, they are better than the other algorithms with gains in the range of 5%-40%. Finally, we consider the mobility scenario that the user travels through different overlapping areas of networks. As shown in Fig 4, the simulation process can be divided into three stages where the available networks vary but the network environment remains the same with Table I. The results indicates that the proposed algorithms, especially algorithm 2 and algorithm 3, can adapt to the changes in available networks to some extent.

V. Conclusion

In this letter, we studied the network selection problem considering time-varying user demand and non-uniform network handoff costs. In order to balance the user QoE and the network handoff cost in a dynamic environment, we proposed three online network selection algorithms. The proposed algorithms can attain the optimal network selection policy on the fly. Simulations reveal that the proposed algorithms significantly outperform existing methods.

References

[1] S. P. Yeh, S. Talwar, G. Wu, *et al.*, "Capacity and coverage enhancement in heterogeneous networks," *IEEE Wireless Commun.*, vol. 18, no. 3, pp. 32–38, 2011.

[2] Q. T. Nguyen-Vuong, Y. Ghamri-Doudane, and N. Agoulmine, "On utility models for access network selection in wireless heterogeneous networks," in *2008 IEEE Network Operations and Management Symposium.*

[3] J. Hou and D. C. O'Brien, "Vertical handover-decision-making algorithm using fuzzy logic for the integrated radio-and-OW system," *IEEE Trans. Wireless Commun.,* vol. 5 no. 1, pp. 176–185, 2006.

[4] K. Pahlavan, P. Krishnamurthy, A. Hatami, M. Ylianttila, J. P. Makela, R. Pichna, and J. Vallstron, "Handoff in hybrid mobile data networks," *IEEE Personal Commun.,* vol. 7, no. 2, pp. 34–47, 2000.

[5] B. Chang and J. Chen, "Cross-layer-based adaptive vertical handoff with predictive RSS in heterogeneous wireless tetworks," *IEEE Trans. Veh. Technol.*, vol. 57, no. 6, pp. 3679–3692, 2008.

[6] D. Ma and M. Ma, "A QoS oriented vertical handoff scheme for WiMAX/WLAN overlay networks," *IEEE Trans. Parallel Distrib. Syst.*, vol. 23, no. 4, pp. 598–606, 2012.

[7] E. Stevens-Navarro, Y. Lin, and V. W. S. Wong, "An MDP-based vertical handoff decision algorithm for heterogeneous wireless networks," *IEEE Trans. Veh. Technol.*, vol. 57, no. 2, pp. 1243–1254, 2008.

[8] S. Busanelli, M. Martalo, G. Ferrari, *et al.*, "Vertical handover between WiFi and UMTS networks: experimental performance analysis," *International J. Energy, Inf. and Commun.,* vol. 2, no. 1, pp. 75–96, 2011.

[9] K. Piamrat, A. Ksentini, C. Viho, *et al.*, "QoE-based network selection for multimedia users in IEEE 802.11 wireless networks," in *Proc. 2008 IEEE Local Computer Networks*, pp. 388–394.

[10] A. B. Reis., J. Chakareski, A. Kassler, *et al.*, "Distortion optimized multi-service scheduling for next-generation wireless mesh networks," in *Proc. 2010 IEEE INFOCOM*, pp. 4473–4477.

[11] S. Shamik, C. Mainak, and G. Samrat, "Improving quality of VoIP streams over WiMax," *IEEE Trans. Comput.*, vol. 57, no. 8, pp. 145–156, 2008.