Dynamic parameter adaptation for M-LWDF/M-LWWF scheduling

Article *in* IEEE Transactions on Wireless Communications · March 2012

DOI: 10.1109/TWC.2012.020712.101398

3 authors:

Ju Yong Lee
Korea Advanced Institute of Science and Technology

Sameh Sorour
Queen's University

Shahrokh Valaee
University of Toronto

Dynamic Parameter Adaptation for M-LWDF/M-LWWF Scheduling

Ju Yong Lee, *Member, IEEE*, Sameh Sorour, *Student Member, IEEE,*
Shahrokh Valaee, *Senior Member, IEEE*, and Wonyoung Park

***Abstract*—M-LWWF/M-LWDF scheduling schemes have attracted much interest due to their ability to both stabilize queues whenever possible and control delay through parameter selection. However, a good implementation of these schedulers would require a mechanism to minimize the required fraction of the bandwidth while satisfying its stability and delay requirements. To the best of our knowledge, previous works on these scheduling policies did not address the problem of minimizing the bandwidth utilization while satisfying delay constraints.**

In this paper, we explore the solution of this problem using a joint bandwidth and weight adaptation approach. We characterize the problem solution space for M-LWWF and M-LWDF scheduling, assuming time-varying traffic. We also show that, starting from any point in the solution space, simple dynamic bandwidth and weight updates can surely lead to the convergence to the optimal operation point in this space. Based on these characteristics, we propose a dynamic parameter adaptation algorithm that is able to track the time-varying optimal operation points for dynamic traffic and channel conditions. Simulation results show the efficiency of our proposed algorithm in tracking the optimal operation points in dynamic traffic and channel settings.

***Index Terms*—M-LWWF/M-LWDF scheduling, parameter optimization, bandwidth minimization, Lyapunov stability.**

I. INTRODUCTION

OPPORTUNISTIC scheduling has been a target of intensive studies in academia and industry in the past decade due to its ability to exploit user mobility and wireless channel variations to achieve higher system performance. An attractive feature of opportunistic scheduling is its ability to achieve these performance improvements using simple and easily implementable index policies [1].

Different opportunistic scheduling approaches have been proposed in the literature. Some of these approaches [2–5] consider the infinite backlog model that assumes permanent availability of data in user queues. The goal of these approaches is to maximize the achievable wireless capacity under either temporal bandwidth or proportional fairness constraints. Liu *et al.* [2] proposed an opportunistic scheduling algorithm with weighted temporal ratio fairness while Borst *et al.* [3] propose a scheduling algorithm with weighted bandwidth ratio fairness. These algorithms propose the adjustment of user weight values to satisfy bandwidth or temporal ratio fairness. Park *et al.* [4] proposed a scheduler using cumulative distribution function of air channel capacity, where temporal fairness can be guaranteed and average throughput can be derived from channel distribution and temporal ratios. However, it needs a large amount of statistics to measure exact channel distribution, and thus, it is difficult to adapt to the dynamically changing environment.

One problem with the aforementioned scheduling approaches is the infinite backlog model assumption, which is not generally a practical assumption. It is more practical to consider finite backlog model, where traffic is not permanently present in user queues but rather fluctuates according to some stochastic models. In such works, the main focus of the designed scheduler is to minimize packet delays in different queues and/or stabilize user queues whenever any other scheduler can. These schedulers are known as throughput optimal schedulers. One of the leading works in this direction is [6], where it is proved that for symmetric input traffic and packet loss probabilities, the longest connected queue (LCQ) scheduling minimizes the average delay. In [7, 8], Ganti *et al.* generalized [6] to the case for symmetric Bernoulli traffic and ON/OFF channels. Neely *et al.* [9] computed upper bounds on the delay of stabilizing largest-queue type strategies for heterogeneous downlinks. Later, Neely proposed in [10] a dynamic queue-length aware algorithm that maximizes throughput and achieves average delay over ON/OFF channels that is independent of the number of queues. They apply this policy to both symmetric and asymmetric systems.

Along the same line, another set of throughput optimal algorithms were proposed in [11–13], namely the modified large weighted delay first (M-LWDF), modified largest weighted work first (M-LWWF) and the exponential (EXP) rule schedulers. These schedulers select the user whose delay (or queue length) weighted sum (or a function) of user rates is maximum. An additional weight is also employed in the selection policy of these schedulers to control the delay violation probability of the different queues [14]. These weights are generally fixed to a certain value related to the delay bound and an upper-

Manuscript received August 4, 2010; revised January 1 and July 6, 2011; accepted January 4, 2012. The associate editor coordinating the review of this paper and approving it for publication was G. Song.

J. Y. Lee was with the Edward S. Rogers Sr. Department of Electrical and Computer Engineering, University of Toronto, Canada. He is now with KI for IT Convergence, KAIST, Daejeon, Korea (e-mail: jylee@itc.kaist.ac.kr).

S. Sorour and S. Valaee are with the Edward S. Rogers Sr. Department of Electrical and Computer Engineering, University of Toronto, 10 King's College Road, Toronto, ON, M5S 3G4, Canada (e-mail: {samehsorour, valaee}@comm.utoronto.ca).

W. Park is with the Telecommunication Network Business, Samsung Electronics Co. Ltd., Suwon, Korea (e-mail: wony.park@gmail.com).

This research was supported by the KCC (Korea Communications Commission), Korea, under the R&D program supervised by the KCA (Korea Communications Agency) (KCA-2011-11913-04001).

Digital Object Identifier 10.1109/TWC.2012.020712.101398

bound on the delay violation probability. M-LWWF and M-LWDF can be regarded as the special case of MaxWeight scheduling [15], whose stability has been discussed in [16]. Recently, Sadiq *et. al* [17] proposed the Log rule, where for linearly growing queues, the scheduler allocates service in a manner that de-emphasizes the balancing of unequal queues in favor of maximizing current system throughput (being opportunistic). They provide the important insight that queue length information is critical in developing throughput optimal scheduling schemes.

The opportunistic scheduling literature, for both infinite and finite backlog models, assumed that the whole system bandwidth is available to serve the incoming traffic flows. This assumption is clearly valid for the infinite backlog model as all the system bandwidth is required to serve the infinite amount of traffic. However, the full bandwidth utilization is not always needed in the finite backlog model for traffic that is required to satisfy average delays. Moreover, user weights are assumed to be fixed in some of these scheduling techniques. However, in practical systems, both incoming traffic flows from different users and channel qualities are time varying. In [18], it is pointed out that, although the flexibility in assigning these weights allows one to handle heterogeneity in channel capacity distribution, it is very difficult to determine these weights because they are dependent on the channel capacity distributions and traffic characteristics of all users. Moreover, Andrews *et al.* [14] argue that the feasibility of queue stability and delay requirements in throughput optimal schedulers (M-LWDF in this article) requires a good binding of the flows' characteristics with respect to the available bandwidth and an efficient call admission control.

In addition, wireless networks generally serve both delay-constrained and best-effort traffic. Consequently, higher service can be given to the best-effort traffic if the delay-constrained traffic is satisfied with the minimum possible bandwidth. All of these aspects require a mechanism for joint dynamic bandwidth and user weight adaptation such that the bandwidth is minimized while satisfying the delay constraints of different traffic flows. This parameter adaptation process is also important for designing call admission control algorithms when these scheduling techniques are practically implemented in real networks. For example, if a new call requires the resource less than the remaining bandwidth, the call can be admitted. Otherwise, it is rejected.

In this paper, we first formulate the joint bandwidth and weight adaptation problem. Then, we characterize the solution space of the adaptation problem in terms of bandwidth and user weights, for both M-LWDF and M-LWWF schedulers, and show that the optimal operation point can be reached from any point in this solution space using a simple parameter update approach. We then find an efficient dynamic parameter adaptation algorithm for both schedulers without any assumptions on the traffic and channel fluctuations. The proposed algorithm tends to track the time-varying optimal operation point of the system. We then evaluate the performance of our proposed algorithm in an OFDMA based WiMAX wireless environment.

The rest of this paper is organized as follows: In Section II, we describe the optimization problem in M-LWWF/M-LWDF schedulers. Section III proposes a parameter optimization algorithm and validates its stability. We develop a dynamic parameter adaptation algorithm in mobile WiMAX system in Section IV. Section V presents the simulation results in a downlink OFDMA-based WiMAX system. Finally, Section VI concludes the paper.

II. Problem Statement in M-LWWF/M-LWDF Scheduling

We assume there are K users denoted with the set $\mathcal{K} = \{1, 2, \ldots, K\}$. Let $A_k(t)$ be the traffic arrived to the k-th user queue during $[0, t)$, where input traffic packets are assumed to arrive in discrete time. We define $A_k(t), k \in \mathcal{K}$ as:

$$A_k(t) = A_{k,i} \quad \text{for } t_{k,i-1} \le t < t_{k,i}, i = 1, 2, \ldots, \tag{1}$$

where $A_{k,i-1} < A_{k,i}$. Then, $A_k(\cdot)$ is a piecewise constant and nondecreasing function. Let $c_k(t)$ be the channel capacity if the k-th user is selected for transmission at time t, where $c_k(\cdot)$ is assumed to be uniformly bounded and piecewise constant function. Let $S_k(t)$ be the traffic served in the k-th queue during $[0, t)$. Note that $S_k(\cdot)$ are nondecreasing functions. We denote by $Q_k(t)$, the occupancy of the k-th queue at time t expressed as:

$$Q_k(t) = A_k(t) - S_k(t). \tag{2}$$

We also define $D_k(t)$ as the delay that the head-of-line traffic in the k-th queue suffers at time t. Then, $D_k(\cdot)$ can be expressed as:

$$D_k(t) = t - H_k(S_k(t)). \tag{3}$$

where $H_k(x) = t_{k,i}$ for $A_{k,i} \le x < A_{k,i+1}$, $i = 1, 2, \ldots$. Since $A_k(t) \ge S_k(t)$, $Q_k(t) \ge 0$ and $D_k(t) \ge 0$ $k \in \mathcal{K}, \forall t \ge t_0$.

M-LWWF and M-LWDF schedulers stabilize the system whenever any other feasible scheduler can stabilize the system [11]. We consider these schedulers in continuous time domain. Let $I(t)$ be the index of the queue scheduled at time t. Then, $I(t)$ for the M-LWWF and M-LWDF schedulers can be expressed as [11]:

$$I(t) \in \arg\max_{k \in \mathcal{K}} \left(\alpha_k Q_k(t) c_k(t)\right), \tag{4}$$

$$I(t) \in \arg\max_{k \in \mathcal{K}} \left(\alpha_k D_k(t) c_k(t)\right), \tag{5}$$

respectively, where $\alpha_k > 0$ is the k-th user weight. We consider cases in which more than one index satisfy the above maximizations. In such situations, we break ties by serving queues with the largest capacity (better utilization). If the i-th queue and the j-th queue with ties have the same capacity and $i < j$, the i-th queue is served. Then, (4) and (5) give unique solution for $Q_k(\cdot)$ and $D_k(\cdot), k \in \mathcal{K}$.

We assume that the total available resources are used to serve both QoS guaranteed and best effort traffic. The objective is to minimize the total consumed capacity for guaranteed traffic, while maintaining the required QoS. The minimization of the allocated resources to the QoS traffic provides a better service for the best effort traffic.

The ratio of the allocated resources for the QoS guaranteed traffic to the total available resources is denoted by Ψ, where $0 < \Psi \le 1$. This representation aligns well with OFDMA and TDMA schemes, which are the modulations of choice for future wireless technologies. For instance, in OFDMA one can assume that a subset of subcarriers in each symbol is allocated to QoS guaranteed sources, and the rest left for best effort service. With this definition, the total available bandwidth for a scheduled queue k at time t is $\Psi c_k(t)$.

First, we consider the M-LWWF scheduling. Suppose that $\alpha_i Q_i(t_1)c_i(t_1) = \alpha_j Q_j(t_1)c_j(t_1)$, and $A_k(\cdot)$ and $c_k(\cdot)$, $\forall k \in \mathcal{K}$ do not change during $[t_1, t_2)$. If there exists a $\delta(< t_2 - t_1)$ such that the i-th queue is served and the j-th queue is not served during $[t_1, t_1+\delta)$, the i-th queue size decreases and the j-th queue size does not decrease, and thus, $\alpha_i Q_i(t_1+\delta/2)c_i(t_1+\delta/2) < \alpha_j Q_j(t_1+\delta/2)c_j(t_1+\delta/2)$, which contradicts the scheduling policy (4). Therefore, for some δ, $\alpha_i Q_i(t)c_i(t) = \alpha_j Q_j(t)c_j(t), t_1 \le t < t_1+\delta$, where both queues are served and overall resource Ψ is utilized by these queues simultaneously during $(t_1, t_1+\delta)$. Let $\Omega(t)$ be the set of served queues at time t. Let $J_k(t)$ be the allocated bandwidth ratio for the k-th queue such that the service given to the queue k at time t is

$$S_k(t) = \Psi \int_0^t J_k(\tau)c_k(\tau)d\tau, \tag{6}$$

where $0 < J_k(t) \le 1$ for $k \in \Omega(t)$.

Suppose that no packet arrives in queue i and j during $[t_1, t_2)$, $c_i(t)$ and $c_j(t)$ are constant during $[t_1, t_2)$, and $i, j \in \Omega(t)$ for $t \in [t_1, t_2)$. Then,

$$\alpha_i Q_i(t)c_i(t) = \alpha_j Q_j(t)c_j(t). \tag{7}$$

Thus,

$$\begin{aligned} &\alpha_i \left(A_i(t_1) - \Psi \int_0^t J_i(\tau)c_i(\tau)d\tau \right) c_i(t_1) \\ &= \alpha_j \left(A_j(t_1) - \Psi \int_0^t J_j(\tau)c_j(\tau)d\tau \right) c_j(t_1). \end{aligned} \tag{8}$$

Differentiating both sides of (8),

$$\frac{J_i(t)}{J_j(t)} = \frac{\alpha_j (c_j(t))^2}{\alpha_i (c_i(t))^2}. \tag{9}$$

Since $\sum_{k\in\Omega(t)} J_k(t) = 1$,

$$J_k(t) = \frac{1}{\alpha_k (c_k(t))^2} \left(\sum_{i\in\Omega(t)} \frac{1}{\alpha_i (c_i(t))^2} \right)^{-1}. \tag{10}$$

Therefore, $J_k(\cdot)$ is a piecewise constant function, and thus, $S_k(\cdot)$ in (6) is a continuous and piecewise linear function. Moreover, $Q_k(\cdot)$ is a piecewise linear function.

Now we consider the M-LWDF scheduling. Since $c_k(\cdot)$ is uniformly bounded, $S_k(\cdot)$ is a continuous function. Thus, $H_k(S_k(t))$ is a piecewise constant function for t. If $H_k(S_k(t))$ is constant for $t_1 \le t < t_2$,

$$\frac{d}{dt} D_k(t) = 1, \quad t_1 < t < t_2. \tag{11}$$

Therefore, from the scheduling policy in (5), only one user can be served during (t_1, t_2), that is, $J_k(t) = 1$ or $J_k(t) = 0$. The discontinuity of $D_k(\cdot)$ occurs when $S_k(t) = A_{k,i}$, $i = 1, 2, \ldots$. Thus, $D_k(\cdot)$ is a piecewise linear function.

Users may have different quality-of-service (QoS) requirements. In this paper, we deal with satisfying the average delay requirements. A similar approach can be suggested for average queue occupancy requirements. Let d_k and $\hat{d}_k$ be the average delay and the target average delay for the k-th user, respectively. In M-LWDF/M-LWWF schedulers, d_k is a function of Ψ and $\boldsymbol{\alpha} = (\alpha_1, \alpha_2, \ldots, \alpha_K)$. The objective is to minimize the total consumed resources while satisfying the QoS requirements (identified as the average delay) for all users. Accordingly, we consider the following problem:

$$\begin{aligned} \min_{\Psi, \boldsymbol{\alpha}} \quad & \Psi \\ \text{such that} \quad & d_k(\Psi, \boldsymbol{\alpha}) \le \hat{d}_k, \quad k \in \mathcal{K}. \end{aligned} \tag{12}$$

We denote the solution of (12) by $(\Psi^*, \boldsymbol{\alpha}^*)$.

Our objective in minimization (12) is to reduce the total resources consumed by QoS guaranteed sources and hence improve the performance of the best-effort flows. As a result, one might introduce a more general problem to optimize for a control policy $\Psi(Q_1, \ldots, Q_K; c_1, \ldots, c_K)$ such that $E\Psi(\cdot)$ is minimized [19]. However, in this paper we consider the minimization (12) for two reasons. First, as we will show in Section IV, this case aligns well with the current and upcoming standards such as WiFi Point Coordination Function, WiMax and LTE. Second, (12) can be easily solved recursively with decomposing into two subproblems that can be iterated to arrive at the optimal solution.

In general, (12) is a nonlinear nonconvex optimization problem. We propose an algorithm that can converge to the optimal value iteratively. The proposed algorithm operates by modifying Ψ or $\boldsymbol{\alpha}$ while keeping the other parameters constant until the optimum point is reached. In the following section we study the solution space and show that the proposed iterative algorithm converges to $(\Psi^*, \boldsymbol{\alpha}^*)$.

Next, we study the behavior of the system when Ψ or $\boldsymbol{\alpha}$ varies. We show that increasing Ψ, while keeping $\boldsymbol{\alpha}$ constant, has a constructive effect on all sample paths of delay and queue size. Consider two systems X and Y with different parameters $(\Psi^X, \boldsymbol{\alpha}^X)$ and $(\Psi^Y, \boldsymbol{\alpha}^Y)$, respectively, where $\boldsymbol{\alpha}^X = (\alpha_1^X, \alpha_2^X, \ldots, \alpha_K^X)$ and $\boldsymbol{\alpha}^Y = (\alpha_1^Y, \alpha_2^Y, \ldots, \alpha_K^Y)$. Let $S_k^X(t)$ and $S_k^Y(t)$ be the served traffic of the k-th queue in X and Y, respectively. We assume that input traffic patterns, $A_k(t), k \in \mathcal{K}$ and the channel capacity $c_k(t), k \in \mathcal{K}$ are the same for both systems for $t \ge t_0$, where t_0 is the time origin at which $S_k^X(t_0) = S_k^Y(t_0)$, $k \in \mathcal{K}$. Then, $Q_k^X(t_0) = Q_k^Y(t_0)$, $k \in \mathcal{K}$. Let $I^X(t)$ and $I^Y(t)$ be the queue indexes which are selected for transmission through scheduling at time t in the systems X and Y, respectively.

Lemma 1: *Suppose that $\Psi^X \le \Psi^Y$ and $\boldsymbol{\alpha}^X = \boldsymbol{\alpha}^Y$. Then, in the M-LWWF scheduler, for $k \in \mathcal{K}$ and $t \ge t_0$,*

$$Q_k^X(t) \ge Q_k^Y(t). \tag{13}$$

In the M-LWDF scheduler, for $k \in \mathcal{K}$ and $t \ge t_0$,

$$D_k^X(t) \ge D_k^Y(t). \tag{14}$$

Proof: The proof is in A. ■

Since (13) and (14) hold for all sample paths, we can extend them to expected values. Let the expected values of $Q_k^X(t), Q_k^Y(t), D_k^X(t)$ and $D_k^Y(t)$ be represented by

$q_k(\Psi^X, \alpha^X), q_k(\Psi^Y, \alpha^Y), d_k(\Psi^X, \alpha^X)$ and $d_k(\Psi^Y, \alpha^Y)$, respectively. We assume that q_k and d_k are differentiable with Ψ and $\boldsymbol{\alpha}$, which is common for practical setting. Then, we have the following corollary.

Corollary 2: *For $k \in \mathcal{K}$,*

$$\frac{\partial q_k}{\partial \Psi}(\Psi, \boldsymbol{\alpha}) \leq 0 \quad \forall(\Psi, \boldsymbol{\alpha}), \tag{15}$$

$$\frac{\partial d_k}{\partial \Psi}(\Psi, \boldsymbol{\alpha}) \leq 0 \quad \forall(\Psi, \boldsymbol{\alpha}). \tag{16}$$

Now we consider the case when total resource, Ψ does not change and the weight of one user changes while weights of other users do not change. Suppose that, in the scheduling in (4), $\Psi^X = \Psi^Y = \Psi$ and

$$\alpha_l^X = \alpha_l, \quad \alpha_l^Y = \beta\alpha_l, \tag{17}$$

$$\alpha_j^X = \alpha_j^Y = \alpha_j \quad j \neq l, j \in \mathcal{K}. \tag{18}$$

where $\beta > 1$.

Let $J_l^X(t)$ and $J_l^Y(t)$ be the indication function for the l-th user at time t in system X and Y, respectively. Let us also define the cumulative scheduling function $\Phi_l^X(t)$ as

$$\Phi_l^X(t) = \int_0^t J_l^X(\tau) d\tau. \tag{19}$$

The effect of an increased weight for the l-th user in regime Y is that the user will be given service earlier and hence the expected value of $\Phi_l^Y(t)$ will be larger than the expected value of $\Phi_l^X(t)$. We show this formally in the sequel.

We first note that our analysis is for average case. Therefore, we can assume that the channel is stationary and the total expected capacity available for any source i is given by $c_i = Ec_i(t)$. Note that this assumption is for convenience and does not imply any limitation on the analysis. An alternative would be to first analyze the system conditioned on the channel capacity, and then average over the channel pdf.

Consider a typical busy period for the whole system in regime X, and assume without the loss of generality that the busy period starts at time 0. We first notice that since the capacity of the channel is the same for both regimes, the system busy periods are identical in both systems. Now consider any source i will be served in the periods marked by the identifier function $J_i^X(t)$. For tractability of the problem we make the following assumptions. The first time that the i-th source is served in regime X is identified by s_i^X. A similar parameter can be defined for regime Y. The consecutive service times for each source constitute a set of random epochs that exhibit a stationary independent increment property. This assumption can be justified by the fact that in between each two consecutive service of a source other contending traffics from independent sources are served. A similar assumption is usually made in queueing theory to enable tractability of the solution.

With this assumption $\Phi_i^X(t)$ and $\Phi_i^Y(t)$ are in general Levy processes, but with our setting of the problem can be modelled as compound Poisson processes. Therefore, $E\Phi_i^X(t) = \lambda(t - s_i^X)E\Delta$, where λ is the density of the Poisson points and $E\Delta$ is the expected value of the jumps in the Poisson process. We now make the following observations. First, we argue that $s_l^Y \leq s_l^X$. This is due to the fact that the larger scheduling coefficient in Y would give a priority for the l-th source to be served earlier. With this observation we get that

$$q_l^X \geq q_l^Y, \tag{20}$$

Since the l-th source is served earlier, some queues might have a service time in regime Y that is later than their service time in regime X. Therefore we have

$$q_j^X \leq q_j^Y, \quad j \in \mathcal{K}, \ j \neq l. \tag{21}$$

Since the arrival process for the two systems is the same, from Little's formula [20] we have

$$d_l^X \geq d_l^Y, \tag{22}$$

$$d_j^X \leq d_j^Y \quad j \in \mathcal{K}, \ j \neq l. \tag{23}$$

Based on the above discussion we have proven the following lemma.

Lemma 3: Assuming that the cumulative scheduling function is a compound Poisson process, we have for $i \in \mathcal{K}$,

$$\frac{\partial q_i}{\partial \alpha_i}(\Psi, \boldsymbol{\alpha}) \leq 0, \quad \frac{\partial d_i}{\partial \alpha_i}(\Psi, \boldsymbol{\alpha}) \leq 0, \quad \forall(\Psi, \boldsymbol{\alpha}), \tag{24}$$

Moreover, if $i \neq j$,

$$\frac{\partial q_j}{\partial \alpha_i}(\Psi, \boldsymbol{\alpha}) \geq 0, \quad \frac{\partial d_j}{\partial \alpha_i}(\Psi, \boldsymbol{\alpha}) \geq 0, \quad \forall(\Psi, \boldsymbol{\alpha}). \tag{25}$$

III. Parameter Optimization in M-LWWF/M-LWDF Scheduling

Based on the properties in the previous section, we investigate the characteristics of the solution space $(\Psi, \boldsymbol{\alpha})$ for (12). We define the sets, $\mathcal{S}$ and $\mathcal{V}$ as follows:

$$\mathcal{S}(\Psi, \boldsymbol{\alpha}) = \{k | d_k(\Psi, \boldsymbol{\alpha}) \leq \hat{d}_k, \quad k \in \mathcal{K}\}, \tag{26}$$

$$\mathcal{V}(\Psi, \boldsymbol{\alpha}) = \{k | d_k(\Psi, \boldsymbol{\alpha}) > \hat{d}_k, \quad k \in \mathcal{K}\}. \tag{27}$$

$\mathcal{S}(\Psi, \boldsymbol{\alpha})$ is the set of users whose delay requirements are *satisfied* at a given $(\Psi, \boldsymbol{\alpha})$ point and $\mathcal{V}(\Psi, \boldsymbol{\alpha})$ is the set of users whose delay requirements are *violated* at point $(\Psi, \boldsymbol{\alpha})$. For the brevity at presentation, we drop the arguments and denote the set of satisfied users by $\mathcal{S}$, and the set of unsatisfied users by $\mathcal{V}$. If all users are in $\mathcal{S}$, then the point $(\Psi, \boldsymbol{\alpha})$ is called a *feasible* point. Otherwise, it is called an *infeasible* point. Moreover, if all users are in $\mathcal{V}$, the point $(\Psi, \boldsymbol{\alpha})$ is called a *fully infeasible* point. If $(\Psi, \boldsymbol{\alpha})$ is an infeasible point but is not a fully infeasible point, then it is called a *partially infeasible* point.

We note that the above definition of feasibility is based on the mean delay and does not necessarily hold for all sample paths. In other words, for a feasible point $(\Psi, \boldsymbol{\alpha})$, the randomness of the input traffic and the capacity curve $c_k(t)$ can create sample paths of delay $D_k(t)$, for which $D_k(t) > \hat{d}_k$ for some t and k. However, if $(\Psi, \boldsymbol{\alpha})$ is feasible then $E[D_k(t)] = d_k(\Psi, \boldsymbol{\alpha}) \leq \hat{d}_k$.

We define the hyperplane $\Omega(\Psi), \Psi \in [0, 1]$ in the solution space as:

$$\Omega(\Psi) = \{(\Psi, \boldsymbol{\alpha}) | \boldsymbol{\alpha} \in \mathbb{R}_+^K\}. \tag{28}$$

Also, we define $\mathbb{F}(\Psi)$ and $\mathbb{I}(\Psi)$ as:

$$\mathbb{F}(\Psi) = \{\boldsymbol{\alpha} | (\Psi, \boldsymbol{\alpha}) \text{ is feasible}, \boldsymbol{\alpha} \in \mathbb{R}_+^K\}, \tag{29}$$

$$\mathbb{I}(\Psi) = \{\boldsymbol{\alpha} | (\Psi, \boldsymbol{\alpha}) \text{ is fully infeasible}, \boldsymbol{\alpha} \in \mathbb{R}_+^K\}. \tag{30}$$

Theorem 4: *If* $\Psi_1 \leq \Psi_2$, *then*

$$\mathbb{F}(\Psi_1) \subseteq \mathbb{F}(\Psi_2), \tag{31}$$

$$\mathbb{I}(\Psi_1) \supseteq \mathbb{I}(\Psi_2). \tag{32}$$

Proof: Suppose that $(\Psi_1, \boldsymbol{\alpha})$ is a feasible point. Then, $d_k(\Psi_1, \boldsymbol{\alpha}) \leq \hat{d}_k \; k \in \mathcal{K}$. On the other hand, from Corollary 2, $d_k(\Psi_2, \boldsymbol{\alpha}) \leq d_k(\Psi_1, \boldsymbol{\alpha}) \; k \in \mathcal{K}$ since $\Psi_2 \geq \Psi_1$. Therefore, $(\Psi_2, \boldsymbol{\alpha})$ is a feasible point, which proves (31).

In a similar way, we can show that if $(\Psi_2, \boldsymbol{\alpha})$ is a fully infeasible point, then $(\Psi_1, \boldsymbol{\alpha})$ is a fully infeasible point, which proves (32). ■

The above Theorem shows that the set of feasible points does not become smaller when Ψ increases. Therefore, $\boldsymbol{\alpha}^*$ should belong to the intersection of all feasible sets for $\Psi \geq \Psi^*$,

$$\boldsymbol{\alpha}^* \in \bigcap_{\Psi \geq \Psi^*} \mathbb{F}(\Psi). \tag{33}$$

Similarly, the set of fully infeasible points does not become smaller when Ψ decreases.

Each point $(\Psi, \boldsymbol{\alpha})$ may be a feasible point, a partially infeasible point or a fully infeasible point. The following Theorem indicates that there is not a hyperplane $\Omega(\Psi)$ that includes both feasible and fully infeasible points.

Theorem 5: *Given* Ψ, $\mathbb{F}(\Psi) = \emptyset$ *if* $\mathbb{I}(\Psi) \neq \emptyset$ *and vice versa.*

Proof: The proof is in B. ■

Suppose that feasible points exist in $\Omega(\Psi)$ and $\Psi' \geq \Psi$. From Theorem 4, there exist feasible points in $\Omega(\Psi')$. Moreover, from Theorem 5, there cannot exist fully infeasible points in $\Omega(\Psi')$. In a similar argument, if fully infeasible points exist in $\Omega(\Psi)$ and $\Psi' \leq \Psi$, there exist fully infeasible points in $\Omega(\Psi')$ and there cannot exist a feasible point in $\Omega(\Psi')$. In summary, the optimal point is located in $\Omega(\Psi)$ with minimum Ψ such that $\Omega(\Psi)$ has a feasible set.

We now show that we can reach feasible region or fully infeasible region with a proper parameter adaptation when $\mathbb{F}(\Psi) \neq \emptyset$ or $\mathbb{I}(\Psi) \neq \emptyset$ in $\Omega(\Psi)$. For this, we consider the following continuous nonlinear control dynamics:

$$\frac{d}{dt}\alpha_k(t) = \begin{cases} -\alpha_k(t) & \text{if } k \in \mathcal{S}(t), \\ \alpha_k(t) & \text{if } k \in \mathcal{V}(t), \end{cases} \tag{34}$$

where $\boldsymbol{\alpha}(t)$ is assumed to be continuous and differentiable for all t, and $\mathcal{S}(t)$ and $\mathcal{V}(t)$ are the set of satisfied and unsatisfied users at time t. We now prove that this control system is asymptotically stable at the optimal solution, which means that we can reach the optimal solution (Ψ^*, α^*) by iteration regardless of the initial point (Ψ, α).

Lemma 6: The dynamics in (34) have the following properties:

$$\frac{d}{dt}d_i(t) \geq 0 \quad \text{if } i \in \mathcal{S}(t), \tag{35}$$

$$\frac{d}{dt}d_i(t) \leq 0 \quad \text{if } i \in \mathcal{V}(t). \tag{36}$$

Proof: The dynamics in (34) can be expressed as follows. As $\delta t \to 0$,

$$\alpha_k(t+\delta t) = \begin{cases} (1-\delta t)\alpha_k(t) & \text{if } k \in \mathcal{S}(t) \\ (1+\delta t)\alpha_k(t) & \text{if } k \in \mathcal{V}(t). \end{cases} \tag{37}$$

Note that we can obtain the same performance when all $\alpha_k(t)$'s are scaled by the same value.

$$\alpha_k(t+\delta t) = \begin{cases} \alpha_k(t)(1-\delta t)(1-\delta t) & \text{if } k \in \mathcal{S}(t), \\ \alpha_k(t)(1+\delta t)(1-\delta t) & \text{if } k \in \mathcal{V}(t). \end{cases} \tag{38}$$

If δt goes to zero in (38), then

$$\frac{d}{dt}\alpha_k(t) = \begin{cases} -2\alpha_k(t) & \text{if } k \in \mathcal{S}(t), \\ 0 & \text{if } k \in \mathcal{V}(t). \end{cases} \tag{39}$$

Therefore, (34) and (39) give the same performance.

We consider the case that $i \in \mathcal{V}(t)$. It follows from (39) and Lemma 3 that

$$\frac{d}{dt}d_i(t) = \sum_{k \in \mathcal{K}} \frac{\partial d_i}{\partial \alpha_k}\frac{d\alpha_k}{dt} = \sum_{k \in \mathcal{S}(t)} \frac{\partial d_i}{\partial \alpha_k}\left(-2\alpha_k(t)\right) \leq 0. \tag{40}$$

In a similar way, we can show that if $i \in \mathcal{S}(t)$,

$$\frac{d}{dt}d_i(t) \geq 0. \tag{41}$$

■

Theorem 7: Suppose that there exists a feasible region (or a fully infeasible region) in $\Omega(\Psi)$. If $\boldsymbol{\alpha}(t)$ changes according to the continuous dynamics (34), the feasible region (or the fully infeasible region) is globally asymptotically stable.

Proof: Suppose that there exists a feasible region. Then, from Theorem 5 there is no fully infeasible point. We adopt the following Lyapunov function:

$$V(\boldsymbol{\alpha}) = \sum_{k \in \mathcal{K}} \left([d_k(\boldsymbol{\alpha}) - \hat{d}_k]^+\right)^2, \tag{42}$$

where $[x]^+ = x$ if $x \geq 0$, and $[x]^+ = 0$ if $x < 0$. Note that $V(\boldsymbol{\alpha}) = 0$ if $\boldsymbol{\alpha}$ is in feasible region, and $V(\boldsymbol{\alpha}) > 0$ if $\boldsymbol{\alpha}$ is in partially infeasible region. From (40),

$$\frac{dV}{dt} = \sum_{k \in \mathcal{V}} 2[d_k(\boldsymbol{\alpha}) - \hat{d}_k]^+ \frac{d}{dt}d_k(t) \leq 0. \tag{43}$$

If $\mathcal{V} \neq \emptyset$, all users in $\mathcal{V}$ increase α_k. Therefore, the partially infeasible region with $\frac{dV}{dt} = 0$ is not included in the invariant set. Hence, feasible region is globally asymptotically stable from LaSalle's Theorem [21].

Suppose that there exists a fully infeasible region. In a similar argument, we can show that the fully infeasible region is globally asymptotically stable using the following Lyapunov function:

$$V(\boldsymbol{\alpha}) = \sum_{k \in \mathcal{K}} \left([\hat{d}_k - d_k(\boldsymbol{\alpha})]^+\right)^2. \tag{44}$$

This completes the proof. ■

IV. Parameter Adaptation in OFDMA Systems

Consider a cellular communication system employing OFDMA as the multiple access technique [22, 23]. The temporal structure of the system is divided in a time division duplex (TDD) mode into a downlink frame and an uplink frame. A frame is divided into several subcarriers in the frequency domain and subcarriers are grouped into M subchannels. A frame is also divided into symbols in temporal domain. A slot is a rectangular resource consisting of one subchannel and one

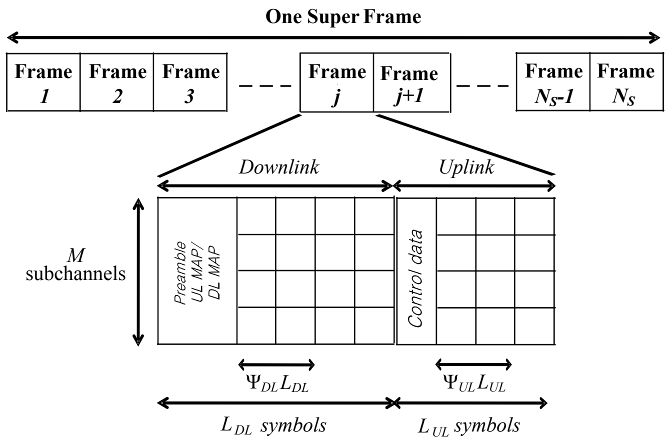


Fig. 1. Frame and super-frame structure.

symbol. One frame for downlink (uplink) has M subchannels and L_{DL} (L_{UL}) symbols, and thus, each frame has $M * L_{DL}$ ($M * L_{UL}$) slots. We introduce a superframe which has N_S frames. The OFDMA frame structure is depicted in Figure 1.

The cell base station (BS) serves K users with targetted delay requirement. The BS tries to use the minimum resources (slots) needed to satisfy the targetted QoS requirements. We define the frame occupancy ratio $\Psi \in [0,1]$ as the ratio of symbols occupied by the QoS traffic per frame to the overall frame size. Then, $\Psi_{DL} * L_{DL}$ and $\Psi_{UL} * L_{UL}$ are the needed resources for the downlink frame and the uplink frame, respectively.

Let $c_k(j,s)$ be the number of bits that user k, $k \in \mathcal{K}$, would be able to receive if it is scheduled to use slot s in frame j. These channel qualities can change on frame basis for all users, and the values $c_k(j,s)\ \forall\ k,s$, are reported to the base station before each frame j. Note that the slots in the same subchannel in each frame have the same channel quality. Let $I(j,s)$ be the user index which is scheduled at slot s in the j-th frame. Let $D_k(j,s)$ and $Q_k(j,s)$ be the head-of-line delay and the queue occupancy in the k-th queue at slot s in frame j, respectively. Then, M-LWDF and M-LWWF can be expressed as:

$$I(j,s) = \arg\max_{k\in\mathcal{K}} \left(\alpha_k D_k(j,s) c_k(j,s)\right), \tag{45}$$

$$I(j,s) = \arg\max_{k\in\mathcal{K}} \left(\alpha_k Q_k(j,s) c_k(j,s)\right). \tag{46}$$

We fix the parameters $(\Psi, \boldsymbol{\alpha})$ during one super-frame with length T_s. Suppose that the system starts with parameters $(\Psi(0), \boldsymbol{\alpha}(0))$. During each super-frame, the average delay values are estimated. By the end of the super-frame, the estimated performance metrics are compared to their required levels and then parameter adaptation is executed. Let $d_k(n)$ be the estimated average delay of user k during the n-th superframe, $[(n-1)T_s, nT_s)$, where the parameters $(\Psi(n-1), \boldsymbol{\alpha}(n-1))$ are used. We define $\mathcal{S}(n)$ and $\mathcal{V}(n)$ as follows:

$$\mathcal{S}(n) = \{k | d_k(n) \le \hat{d}_k, \quad k \in \mathcal{K}\}, \tag{47}$$

$$\mathcal{V}(n) = \{k | d_k(n) > \hat{d}_k, \quad k \in \mathcal{K}\}. \tag{48}$$

Given Ψ, starting in a partially infeasible region in $\Omega(\Psi)$, we propose an algorithm so that a feasible region or a fully infeasible region can be reached if it exists. The proposed

Input: $(\Psi(n-1), \boldsymbol{\alpha}(n-1))$, $\mathcal{S}(n)$, $\mathcal{V}(n)$
Output: $(\Psi(n), \boldsymbol{\alpha}(n))$
begin
  if $\mathcal{V}(n) = \emptyset$ **then**
    $\Psi(n) = \Psi(n-1)(1-\eta)$;
    $\boldsymbol{\alpha}(n) = \boldsymbol{\alpha}(n-1)$;
  else if $\mathcal{S}(n) = \emptyset$ **then**
    $\Psi(n) = \Psi(n-1)(1+\eta)$;
    $\boldsymbol{\alpha}(n) = \boldsymbol{\alpha}(n-1)$;
  else if $i = N_p$ **then**
    $\Psi(n) = \Psi(n-1)(1+\eta)$;
    $\boldsymbol{\alpha}(n) = \boldsymbol{\alpha}(n-1)$;
    $i = 0$;
  else
    $\Psi(n) = \Psi(n-1)$;
    for *each* k *in* $\mathcal{S}(n)$ **do**
      $\alpha_k(n) = \alpha_k(n-1)(1-\epsilon)$;
    end
    for *each* k *in* $\mathcal{V}(n)$ **do**
      $\alpha_k(n) = \alpha_k(n-1)(1+\epsilon)$;
    end
    $i = i + 1$;
end

Algorithm 1: Parameter adaptation algorithm

algorithm has two main steps. Starting from a partially infeasible point in the solution space, for fixed Ψ, we change $\boldsymbol{\alpha}$ so that either a fully infeasible or a feasible point is reached. Then, for fixed $\boldsymbol{\alpha}$, we decrease Ψ if a feasible point was found in the first stage. On the other hand, we increase Ψ if a fully infeasible point was found. If we cannot find feasible region or fully infeasible region in a certain number of adaptation steps, we increase Ψ. We iterate this procedure until a point arbitrarily close to $(\Psi^*, \boldsymbol{\alpha}^*)$ is reached.

If both $\mathcal{S}(n)$ and $\mathcal{V}(n)$ are not empty, then the system is in a partially infeasible point. In this case, $\boldsymbol{\alpha}(n)$ is updated as follows:

$$\alpha_k(n) = \begin{cases} \alpha_k(n-1)(1-\epsilon) & \text{if } k \in \mathcal{S}(n) \\ \alpha_k(n-1)(1+\epsilon) & \text{if } k \in \mathcal{V}(n) \end{cases} \tag{49}$$

where ϵ is a sufficiently small positive constant. If we arrive at a feasible region, we reduce the frame occupancy ratio $\Psi(n)$ as follows:

$$\Psi(n) = \Psi(n-1)(1-\eta) \tag{50}$$

where η is a positive constant. If we arrive at a fully infeasible region or we cannot arrive at a feasible region or a fully infeasible region in N_p updates of $\boldsymbol{\alpha}$, we increase Ψ as follows:

$$\Psi(n) = \Psi(n-1)(1+\eta) \tag{51}$$

The proposed procedure is shown in Algorithm 1.

The dynamics in (49) is the discrete time version of (34). In practice, we should measure $d_k(\boldsymbol{\alpha}), k \in \mathcal{K}$, and thus, there is a measurement error. However, using the stochastic approximation theory [24], it can be shown that the feasible region or fully infeasible region can be reached in a globally asymptotically stable manner with the dynamics in (49) if $\epsilon \to 0$.

TABLE I
SYSTEM PARAMETERS

cell radius	$1km$
total power	$20W$
system bandwidth	$10MHz$
number of sub-carriers per symbol	1024
BS antenna gain	$15dB$
noise level (N_0)	$-174dBm/Hz$

TABLE II
SIMULATION PARAMETERS FOR SCENARIO 1

Class	1	2	3	4
Number of users	2	4	5	10
Rate (kbps)	1000	200	100	50
Velocity (km/h)	0	0	0	0
$\hat{d}_k(sec)$	1	0.2	0.1	0.01

If the input traffic patterns and the air channel capacity do not vary, we can arrive at a neighborhood of the optimal solution $(\Psi^*, \boldsymbol{\alpha}^*)$ with an arbitrarily small radius if ϵ and η are properly selected. However, if the input traffic patterns and the channel capacities change, the proposed algorithm traces the optimal solution with adequate η and ϵ. In the following section, we will show that our algorithm works well in a dynamically varying environment.

V. SIMULATION RESULTS

A. Simulation Environment

We consider downlink transmission in a mobile WiMAX base station. The number of OFDMA subcarriers is 1024 where only 768 subcarriers are used for data transmission and the rest of the subcarriers are used for guard bandwidth. Each 48 data subcarriers are grouped into one subchannel leading to a total of 16 subchannels. The frame consists of 42 OFDMA symbols and 28 symbols (=L_{DL}) are used for downlink transmission. Then, there are 448 slots in each downlink frame. We construct the wireless environment with the system parameters in Table I. Although we consider only downlink, a similar approach is possible for uplink.

We adopt the modified Okumura-Hata model as the propagation model [25] expressed in dB as follows:

$$PL(r) = 128.1 + 37.6 \log_{10}(r)(dB), \tag{52}$$

where r is in km. Shadowing has a log-normal distribution with mean 0 and standard deviation 10, which has correlation 0.5 among different cells. The multi-path fading parameter varies every $5ms$ (frame length) and is exponentially distributed with mean 1. We ignore inter-frame correlation of fading. We consider only cochannel interference of six neighbor cells in the first tier. We assume that all the neighbor cells are fully loaded, that is, they are transmitting packets at all times.

Users are divided into four classes with different input traffic parameters and different average delay requirements. We consider two scenarios. In Scenario 1, each user generates a Poisson traffic with fixed packet size $1kbyte$. All mobiles are located in random positions inside the coverage area of the cell and all users are stationary during the simulation.

TABLE III
SIMULATION PARAMETERS FOR SCENARIO 2

Class	1	2	3	4
Number of users	4	8	10	20
Rate during ON interval (kbps)	10000	2000	1000	500
$E(T_{ON})(sec)$	0.1	0.1	0.1	0.1
$E(T_{OFF})(sec)$	0.9	0.9	0.9	0.9
Shape parameter of T_{ON}	1.5	1.5	1.5	1.5
Shape parameter of T_{OFF}	1.5	1.5	1.5	1.5
Velocity (km/h)	0	4	30	120
$\hat{d}_k(sec)$	1	0.2	0.1	0.01

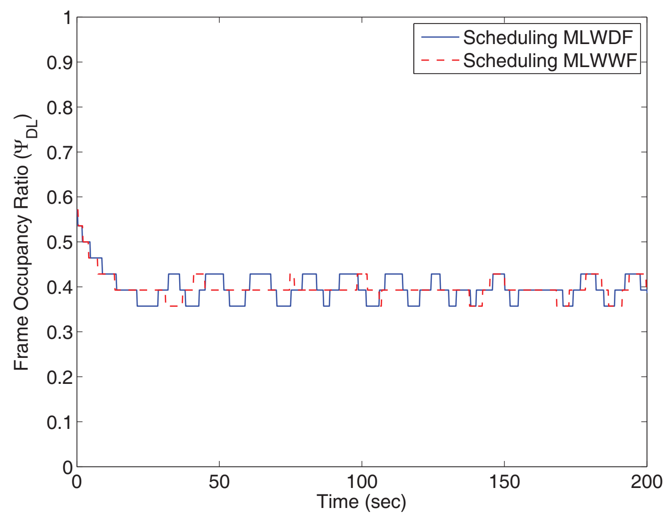


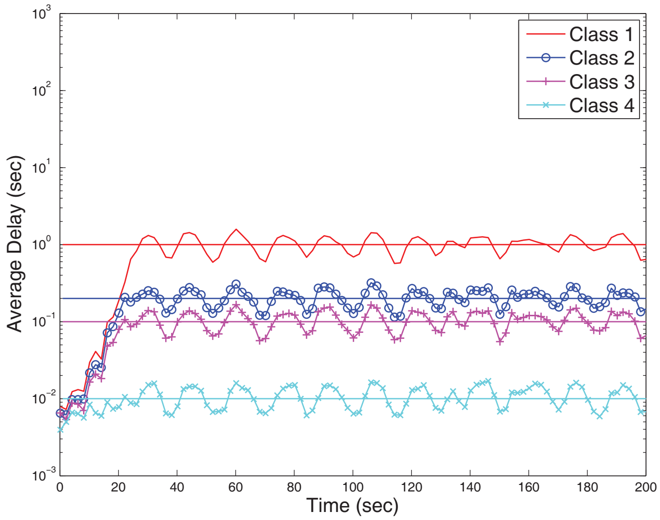
Fig. 2. Comparison of required Ψ (Scenario 1).

In Scenario 2, traffics are generated by ON/OFF sources, where ON and OFF intervals, T_{ON} and T_{OFF}, are Pareto distributed. During T_{ON}, packets are generated uniformly with the fixed packet size, $1kbyte$. Each user moves with its velocity in random directions every $10sec$. If the user moves out of the cell, the user enters the cell again in the opposite direction of the cell. Thus, the number of users in the cell does not change. Tables II and III show different simulation parameters in two scenarios. If the integration of autocorrelation function of traffic amounts during unit time is finite, then the process is short range dependent. Otherwise, if it is infinite, the process is long range dependent [26]. Note that traffic streams of Scenario 1 have short range dependence, and traffic streams of Scenario 2 have long range dependence.

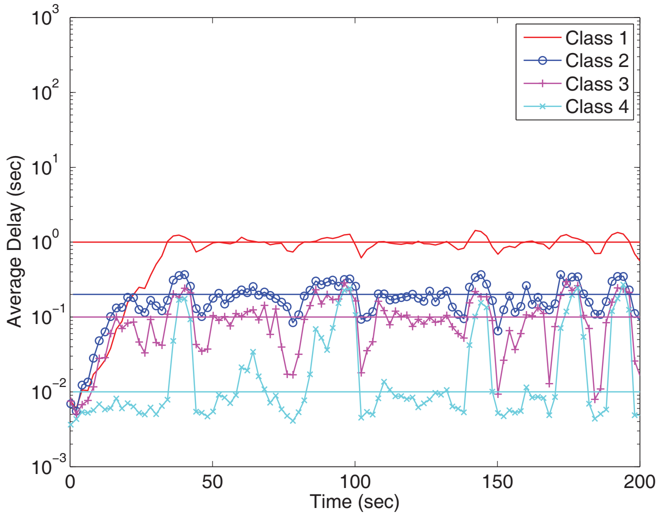
B. Dynamic Parameter Adaptation

We consider a dynamic environment where parameters $(\Psi, \boldsymbol{\alpha})$ are adapted based on the measured average delay. The super-frame length is set to $200msec$. The step size for $\boldsymbol{\alpha}$ and Ψ are set as $\epsilon = 0.1$ and $\eta = 0.01$. If no packets are served for a flow during a super-frame, we do not change the weight of that flow.

First, using the simulation parameters in Scenario 1, we compare the performance of our parameter adaptation algorithm for M-LWDF and M-LWWF. For each of the schedulers, Fig. 2 depicts the variation of the required resource. It takes

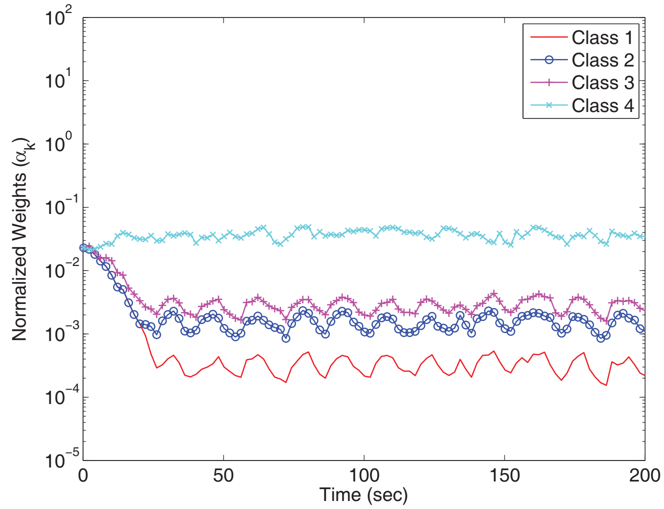


(a) M-LWDF

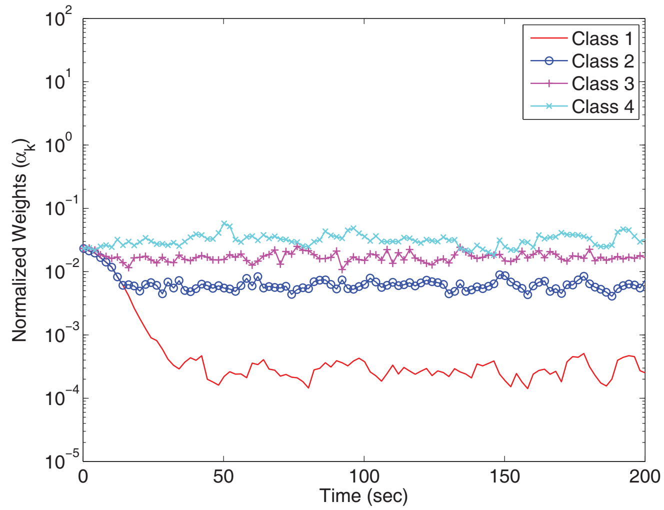


(b) M-LWWF

Fig. 3. Average delay variation (Scenario 1).



(a) M-LWDF



(b) M-LWWF

Fig. 4. Weight variation (Scenario 1).

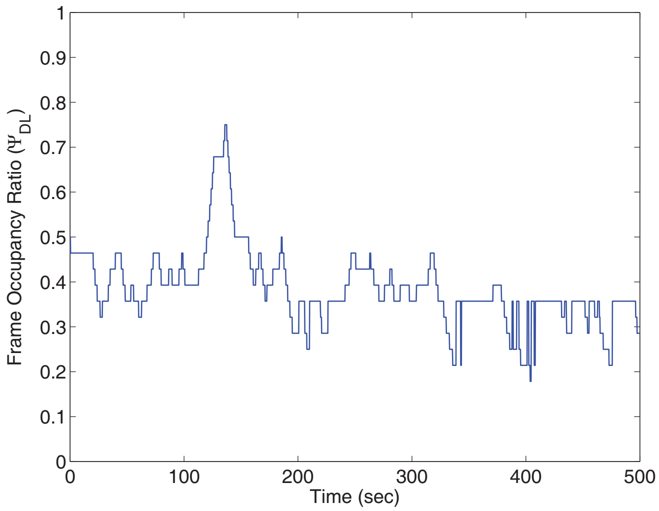
about $20sec$ to converge. After convergence, the mean and the standard deviation of required Ψ for M-LWDF and M-LWWF are (0.3953, 0.0280) and (0.3944, 0.0211), respectively, which shows that the required Ψ for M-LWDF and M-LWWF are similar.

Fig. 3 shows the average delay for all users of each class. The average delay fluctuates about the target average delay, which means that the system is in the optimal state. The average delay fluctuation for stringent delay requirement in the M-LWWF scheduler is higher than that for the M-LWDF scheduler. The reason is that M-LWWF uses queue size rather than delay for scheduling. Fig. 4 shows the weight variation for each scheduler, where weights are normalized so that the sum of weights is equal to 1.

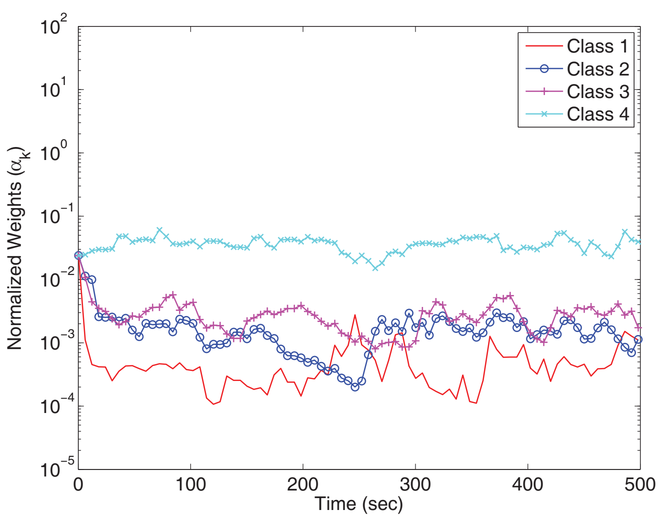
Here, we report the results for Scenario 2 with M-LWDF scheduler. Although the mean arrival rate is the same as Scenario 1, the input traffic with self-similarity is bursty. Moreover, channel capacity of each user changes dynamically since users move according to the velocity in Table III. Fig. 5 (a) shows the adaptation of Ψ where the mean and standard deviation of required Ψ are 0.3862 and 0.0852, respectively. Compared with the results in Fig. 2, the variation of the necessary resource Ψ for Scenario 2 is higher than that for Scenario 1, which is due to the burstiness of traffic and user's mobility. Fig. 5 (b) shows the adjustment of weight values for one of the users in each class. The weight values change dynamically according to the input traffic variation and channel capacity variation. Fig. 5 (c) shows the average delay fluctuation according to the adaptation of Ψ and weight values. The fluctuation is higher than that in Fig. 3 due to user mobility and traffic burstiness.

VI. Conclusion

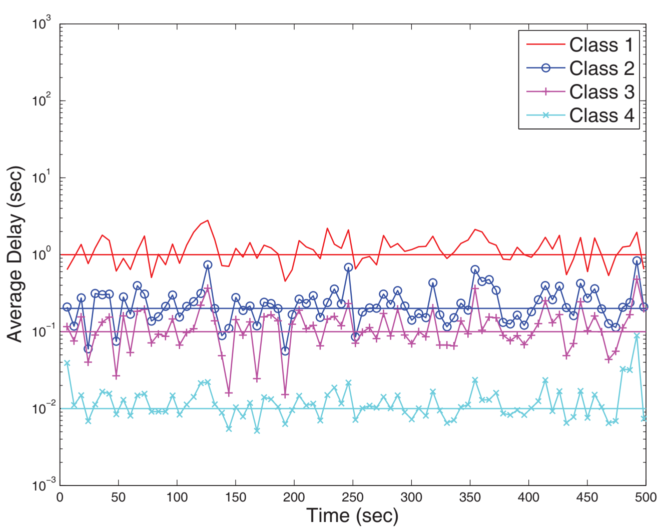
In this paper, we have investigated the problem of dynamic parameter adaptation for throughput optimal opportunistic schedulers. We have first modeled parameter adaptation as an optimization problem and showed a detailed description of its solution space structure and characteristics for opportunistic scheduling algorithms. Based on the solution space structure, we have proposed a practical dynamic parameter adaptation



(a) Variation of Ψ



(b) Weight variation



(c) Average delay variation

Fig. 5. Parameter and Performance Variation with M-LWDF Scheduling in Scenario 2.

algorithm that tracks the system dynamics with the necessary parameters. We have also proved the convergence of the algorithm to the optimal solution for a given channel and queue state.

Appendix A
Proof of Lemma 1

First, we consider the M-LWWF scheduling. We define the following functions for $t \geq t_0$ and $k \in \mathcal{K}$:

$$f_k(t) = Q_k^X(t) - Q_k^Y(t). \tag{53}$$

Since $Q_k^X(\cdot), Q_k^Y(\cdot), k \in \mathcal{K}$ are piecewise linear functions, $f_k(\cdot)$ is a piecewise linear function, and thus, $f_k'(\cdot)$ is a piecewise constant function. Since the input traffic patterns for X and Y, $A_k(\cdot), k \in \mathcal{K}$, are the same, it follows from (2) that

$$f_k(t) = S_k^Y(t) - S_k^X(t), \tag{54}$$

and $f_k(t_0) = 0,\ k \in \mathcal{K}$. Since $S_k^X(\cdot)$ and $S_k^Y(\cdot)$ are continuous functions, $f_k(\cdot)$ is continuous. The condition in (13) can be expressed as:

$$f_k(t) \geq 0, \quad t \geq t_0,\ k \in \mathcal{K}. \tag{55}$$

From (6) and (54),

$$\begin{aligned} f_k'(t) &= \lim_{h \to 0+} \frac{f_k(t+h) - f_k(t)}{h} \\ &= \Psi^Y J_k^Y(t) c_k(t) - \Psi^X J_k^X(t) c_k(t). \end{aligned} \tag{56}$$

Let us assume that the above claim (55) is not true and let τ_{min} be the first time instant at which $f_k(t)$ becomes negative for at least one k, that is,

$$\tau_{min} = \inf_{t \geq t_0} \{t | f_k(t) < 0, \exists k \in \mathcal{K}\}. \tag{57}$$

Since $f_k(\cdot)$ is continuous and piecewise linear, and $f_k'(\cdot)$ is piecewise constant, there exist $\delta(> 0)$ and a set $\Theta(\neq \emptyset)$ such that if $k \in \Theta$,

$$f_k(\tau_{min}) = 0, \tag{58}$$

$$f_k(t) < 0 \quad \text{for } \tau_{min} < t < \tau_{min} + \delta \tag{59}$$

$$f_k'(t) < 0 \quad \text{for } \tau_{min} < t < \tau_{min} + \delta, \tag{60}$$

and if $k \notin \Theta$,

$$f_k(t) \geq 0 \quad \text{for } t < \tau_{min} + \delta. \tag{61}$$

Suppose that $k \in \Theta$ and $\tau_{min} \leq t < \tau_{min} + \delta$ for which $\Omega^X(t)$ and $\Omega^Y(t)$ do not change. From (56) and (60), $J_k^X(t) \neq 0$. Thus, $k \in \Omega^X(t)$. Therefore,

$$\Theta \subset \Omega^X(t). \tag{62}$$

From (59),

$$Q_k^X(t) < Q_k^Y(t). \tag{63}$$

Moreover, since $k \in \Omega^X(t)$,

$$\alpha_k Q_k^X(t) c_k(t) \geq \max_{i \in \mathcal{K}} \alpha_i Q_i^X(t) c_i(t). \tag{64}$$

If $j \in \Omega^Y(t)$,

$$\alpha_k Q_k^Y(t) c_k(t) \leq \alpha_j Q_j^Y(t) c_j(t). \tag{65}$$

Combining (63), (64), (65), for $j \in \Omega^Y(t)$,

$$\alpha_j Q_j^X(t)c_j(t) < \alpha_j Q_j^Y(t)c_j(t). \tag{66}$$

Thus, from (53),

$$f_j(t) < 0 \quad \text{for } j \in \Omega^Y(t). \tag{67}$$

Since $f_k(\cdot), k \in \mathcal{K}$ are continuous and $f_k(t) \geq 0$ for $t \leq \tau_{min}$, $f_j(\tau_{min}) = 0$ and $f_j'(t) < 0$, $\tau_{min} < t < \tau_{min} + \delta$ if $j \in \Omega^Y(t)$. Thus, $\Omega^Y(t) \subset \Theta$. Therefore, from (62), $\Omega^Y(t) \subset \Omega^X(t)$. For $j \in \Omega^Y(t)$,

$$\begin{aligned} f_j'(t) = & \frac{\Psi^Y}{\alpha_j(c_j(t))} \left(\sum_{i \in \Omega^Y(t)} \frac{1}{\alpha_i (c_i(t))^2} \right)^{-1} \\ & - \frac{\Psi^X}{\alpha_j(c_j(t))} \left(\sum_{i \in \Omega^X(t)} \frac{1}{\alpha_i (c_i(t))^2} \right)^{-1} \geq 0. \end{aligned} \tag{68}$$

Since $\Omega^Y(t) \subset \Theta$, there exists j such that $f_j'(t) \geq 0$, $j \in \Theta$, which contradicts (60) and proves (55).

Now, we consider the M-LWDF scheduling. We define the following function for $t \geq t_0$ and $k \in \mathcal{K}$:

$$g_k(t) = D_k^X(t) - D_k^Y(t). \tag{69}$$

Then, the condition in (14) can be expressed as:

$$g_k(t) \geq 0, \quad t \geq t_0,\ k \in \mathcal{K}. \tag{70}$$

From (3),

$$g_k(t) = H_k(S_k^Y(t)) - H_k(S_k^X(t)), \tag{71}$$

and $g_k(t_0) = 0,\ k \in \mathcal{K}$. Note that $g_k(\cdot)$ is a piecewise constant function. We define $t_{k,i}^X$ and $t_{k,i}^Y$ as:

$$t_{k,i}^X = \inf_{t \geq t_0} \{t | S_k^X(t) = A_{k,i}\}, \tag{72}$$

$$t_{k,i}^Y = \inf_{t \geq t_0} \{t | S_k^Y(t) = A_{k,i}\}. \tag{73}$$

Then, the claim in (70) can be restated as:

$$t_{k,i}^X \geq t_{k,i}^Y \quad \forall k, i. \tag{74}$$

Suppose that the claim (70) is not true. Then, there exist l and m such that

$$t_{l,i}^X \geq t_{l,i}^Y, i < m, \tag{75}$$

$$t_{l,m}^X < t_{l,m}^Y, \tag{76}$$

$$D_k^X(t) \leq D_k^Y(t), \quad t < t_{l,m}^X,\ k \in \mathcal{K}. \tag{77}$$

Note that

$$\begin{aligned} A_{l,m} - A_{l,m-1} &= S_l^Y(t_{l,m}^Y) - S_l^Y(t_{l,m-1}^Y) \\ &= S_l^X(t_{l,m}^X) - S_l^X(t_{l,m-1}^X). \end{aligned} \tag{78}$$

Since $S_l^X(\cdot)$ and $S_l^Y(\cdot)$ are non-decreasing functions, from (75), (76), (78),

$$\begin{aligned} & g_l(t_{l,m}^X) - g_l(t_{l,m-1}^X) \\ & = (S_l^Y(t_{l,m}^X) - S_l^Y(t_{l,m-1}^X)) - (S_l^X(t_{l,m}^X) - S_l^X(t_{l,m-1}^X)) \\ & < 0. \end{aligned} \tag{79}$$

Note that for $t \in [t_{l,m-1}^X, t_{l,m}^X)$,

$$D_l^X(t) = D_l^Y(t), \tag{80}$$

$$D_k^X(t) \geq D_k^Y(t) \quad k \in \mathcal{K}. \tag{81}$$

Therefore, if $\alpha_l D_l^X(t)c_l(t) \geq \max_{k \in \mathcal{K}} \alpha_k D_k^X(t)c_k(t)$, $\alpha_l D_l^Y(t)c_l(t) \geq \max_{k \in \mathcal{K}} \alpha_k D_k^Y(t)c_k(t)$. Suppose that $\alpha_j D_j^X(t)c_j(t) = \alpha_l D_l^X(t)c_l(t), j \neq l$, then we can break ties by comparing $c_l(t)$ and $c_j(t)$. If $c_l(t) = c_j(t)$, the indexing order is used to break ties. For any case, if $J_l^X(t) = 1$, then $J_l^Y(t) = 1$. Thus,

$$\begin{aligned} & g_l(t_{l,m}^X) - g_l(t_{l,m-1}^X) \\ & = \Psi^Y \int_{t_{l,m-1}^X}^{t_{l,m}^X} J_l^Y(t)c_l(t)dt - \Psi^X \int_{t_{l,m-1}^X}^{t_{l,m}^X} J_l^X(t)c_l(t)dt \\ & \geq 0. \end{aligned} \tag{82}$$

which contradicts (79) and proves (70).

Appendix B
Proof of Theorem 5

Let $\boldsymbol{\alpha}^X = (\alpha_1^X, \alpha_2^X, \ldots, \alpha_K^X)$ be a fully infeasible point in $\Omega(\Psi)$. Then,

$$d_k(\boldsymbol{\alpha}^X) > \hat{d}_k \quad k \in \mathcal{K}. \tag{83}$$

We choose a point $\boldsymbol{\alpha}^Y = (\alpha_1^Y, \alpha_2^Y, \ldots, \alpha_K^Y)$ in $\Omega(\Psi)$. We introduce a vector $\boldsymbol{\alpha}^Z$ as:

$$\boldsymbol{\alpha}^Z = \left(\frac{\alpha_m^Y}{\alpha_m^X} \right) \boldsymbol{\alpha}^X, \tag{84}$$

where

$$m = \arg\min_{k \in \mathcal{K}} \left(\frac{\alpha_k^Y}{\alpha_k^X} \right). \tag{85}$$

Then, for both schedulers M-LWWF and M-LWDF, we have,

$$d_k(\boldsymbol{\alpha}^Z) = d_k(\boldsymbol{\alpha}^X) \quad k \in \mathcal{K}. \tag{86}$$

On the other hand,

$$\alpha_k^Z = \frac{\alpha_m^Y}{\alpha_m^X} \alpha_k^X \leq \alpha_k^Y, \quad \forall k, \tag{87}$$

and

$$\alpha_m^Z = \alpha_m^Y. \tag{88}$$

Therefore,

$$\boldsymbol{\alpha}^Y = \boldsymbol{\alpha}^Z + \sum_{k \neq m} (\alpha_k^Y - \alpha_k^Z)\mathbf{e}^k. \tag{89}$$

Note that $\alpha_k^Y - \alpha_k^Z \geq 0$ for all k from (87). Therefore, Lemma 3 shows that

$$d_m(\boldsymbol{\alpha}^Y) \geq d_m(\boldsymbol{\alpha}^Z). \tag{90}$$

Combining (83), (86) and (90),

$$d_m(\boldsymbol{\alpha}^Y) > \hat{d}_m, \tag{91}$$

which shows that $\boldsymbol{\alpha}^Y$ is not a feasible point.

So far, we have proven that, if there exists a feasible region, there cannot exist a fully infeasible region. We now show the opposite direction. Suppose that there is a fully infeasible region. If a feasible region exists, then as per the first part of the proof, there cannot exist a fully infeasible region, which is a contradiction and the opposite direction is proved.

REFERENCES

[1] X. Lin, N. Shroff, and R. Srikant, "A tutorial on cross-layer optimization in wireless networks," *IEEE J. Sel. Areas Commun.*, vol. 24, no. 8, pp. 1452–1463, Aug. 2006.

[2] X. Liu, E. Chong, and N. Shroff, "Opportunistic transmission scheduling with resource-sharing constraints in wireless networks," *IEEE J. Sel. Areas Commun.*, vol. 19, no. 10, pp. 2053–2064, Oct. 2001.

[3] S. Borst and P. Whiting, "Dynamic rate control algorithms for HDR throughput optimization," in *Proc. 2001 IEEE Conference on Computer Communications*, vol. 2, pp. 976–985.

[4] D. Park, H. Seo, H. Kwon, and B. G. Lee, "Wireless packet scheduling based on the cumulative distribution function of user transmission rates," *IEEE Trans. Commun.*, vol. 53, no. 11, pp. 1919–1929, Nov. 2005.

[5] X. Liu, E. K. P. Chong, and N. B. Shroff, "A framework for opportunistic scheduling in wireless networks," *Computer Networks*, vol. 41, no. 4, pp. 451–474, 2003.

[6] L. Tassiulas and A. Ephremides, "Dynamic server allocation to parallel queues with randomly varying connectivity," *IEEE Trans. Inf. Theory,*, vol. 39, no. 2, pp. 466–478, Mar. 1993.

[7] A. Ganti, E. Modiano, and J. N. Tsitsiklis, "Transmission scheduling for multi-channel satellite and wireless networks," *2002 Allerton Conference for Communication, Control, and Computing*.

[8] ——, "Optimal transmission scheduling in symmetric communication models with intermittent connectivity," *IEEE Trans. Inf. Theory*, vol. 53, no. 3, pp. 998–1008, Mar. 2007.

[9] M. Neely, E. Modiano, and C. Rohrs, "Power allocation and routing in multibeam satellites with time-varying channels," *IEEE/ACM Trans. Networking*, vol. 11, no. 1, pp. 138–152, Feb. 2003.

[10] M. Neely, "Order optimal delay for opportunistic scheduling in multi-user wireless uplinks and downlinks," *IEEE/ACM Trans. Networking*, vol. 16, no. 5, pp. 1188–1199, Oct. 2008.

[11] M. Andrews, K. Krishnan, K. Ramanan, A. Stolyar, R. Vijayakumar, and P. Whiting, "CDMA data QoS scheduling on the forward link with variable channel conditions," *Bell Labs. Tech. Rep.*, Apr. 2000.

[12] S. Shakkottai and A. L. Stolyar, "Scheduling algorithms for a mixture of real-time and non-real-time data in HDR," in *Proc. 2001 Int. Teletraffic Congress*, pp. 793–804.

[13] ——, "Scheduling for multiple flows sharing a time-varying channel: the Exponential Rule," *Analytic Methods in Applied Probability*, vol. 207, pp. 185–202, 2002.

[14] M. Andrews, K. Kumaran, K. Ramanan, A. Stolyar, P. Whiting, and R. Vijayakumar, "Providing quality of service over a shared wireless link," *IEEE Commun. Mag.*, vol. 39, no. 2, pp. 150–154, Feb. 2001.

[15] A. L. Stolyar, "Maxweight scheduling in a generalized switch: state space collapse and workload minimization in heavy traffic," *The Annals of Applied Probability*, vol. 14, no. 1, pp. 1–53, 2004.

[16] S. Meyn, "Stability and asymptotic optimality of generalized maxweight policies," *SIAM J. Control Optim.*, vol. 47, pp. 3259–3294, Jan. 2009.

[17] B. Sadiq, S. J. Baek, and G. de Veciana, "Delay-optimal opportunistic scheduling and approximations: the Log Rule," in *Proc. 2009 IEEE Conference on Computer Communications*, pp. 1692–1700.

[18] S. Patil and G. de Veciana, "Managing resources and quality of service in heterogeneous wireless systems exploiting opportunism," *IEEE/ACM Trans. Networking*, vol. 15, no. 5, pp. 1046–1058, Oct. 2007.

[19] A. Wierman, L. Andrew, and A. Tang, "Power-aware speed scaling in processor sharing systems," in *Proc. 2009 IEEE Conference on Computer Communications*, pp. 2007–2015.

[20] L. Kleinrock, *Queueing Systems Volume I: Theory*. John Wiley & Sons, 1975.

[21] V. M. Haddad and V. Chellaboina, *Nonlinear Dynamical Systems and Control: A Lyapunov-Based Approach*. Princeton University Press, 2008.

[22] IEEE 802.16e-2005, Part 16:"Air interface for fixed and mobile broadband wireless access systems", Amendment 2 for physical and MAC layers for combined fixed and mobile operation in licensed bands Std., 2005.

[23] WiMAX Forum, "Mobile WiMAX - Part I: A Technical Overview and Performance Evaluation," Std., Aug. 2006.

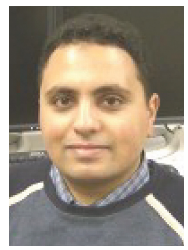
[24] H. F. Chen, *Stochastic Approximation and its Applications*. Kluwer Academic Publishers, 2002.

[25] "Physical layer aspect for evolved universal terrestrial radio access (UTRA)," 3GPP TR 25.814 Std., 2007.

[26] S. Kim, J. Y. Lee, and D. K. Sung, "Bandwidth allocation for fluid input generalized processor sharing servers," *IEEE Trans. Commun.*, pp. 1514–1523, May 2010.

Ju Yong Lee (M'10) received the B.S., M.S., and Ph.D. degrees in electrical engineering and computer science from the Korea Advanced Institute of Science and Technology (KAIST), Daejeon, Korea, in 1995, 1997, and 2003, respectively. He has won the bronze and silver prizes in Samsung Humantech Paper Contest in 1994 and 2003, respectively.

He worked as a senior researcher in the System R&D team of the Network Systems Division with Samsung Electronics from 2003 to 2006. He was also a Postdoctoral Fellow at University of Toronto in Canada from 2007 to 2008. He was a research professor at Sungkyunkwan University, Korea, in 2009. He is now a team leader and research professor of the communication and energy team at KAIST Institute for Information Technology Convergence in KAIST, Korea. His current research interests are in the fifth generation wireless communication network.



Sameh Sorour (S'98) received his B.Sc. and M.Sc. degrees in Electrical Engineering from Alexandria University, Egypt, in 2002 and 2006, respectively. He then received his Ph.D. from the Wireless and Internet Research Laboratory (WIRLab), Department of Electrical and Computer Engineering, University of Toronto, Canada. He is currently a Postdoctoral Fellow at University of Toronto. In 2002, he joined the Department of Electrical Engineering, Alexandria University, where he worked as a Teaching and Research Assistant for three years and was promoted to Assistant Lecturer in 2006. He also served as the chair of local arrangements for IEEE PIMRC 2011. His research interests include opportunistic, random and instantly decodable network coding applications in wireless, vehicular, cognitive and relay-assisted networks, indoor localization, adaptive resource allocation, OFDMA, and wireless scheduling.

Shahrokh Valaee (S'88-M'00-SM'02) received the B.Sc. and M.Sc. degrees from the University of Tehran, and the Ph.D. degree from McGill University, all in electrical engineering. From 1994 to 1995, he was a Research Associate at INRS Telecom, University of Quebec. From 1996 to 2001, he was with the Department of Electrical Engineering, Tarbiat Modares University, and the Department of Electrical Engineering, Sharif University of Technology. Since 2001 he has been a faculty member of the Edward S. Rogers Sr. Department of Electrical and Computer Engineering at the University of Toronto, Canada, where he is the Associate Chair for Undergraduate Studies, holds the Nortel Institute Junior Chair of Communication Networks and is the Director of the Wireless and Internet Research Laboratory (WIRLab).

Prof. Valaee is an Editor of IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS, and an Associate Editor of IEEE SIGNAL PROCESSING LETTERS. He is the TPC Co-Chair and the Chair of the Organizing Committee of IEEE PIMRC 2011. He was the Co-Chair for the Wireless Communications Symposium of IEEE GLOBECOM 2006, a Guest Editor for *IEEE Wireless Communications Magazine*, a Guest Editor for *Wiley Journal on Wireless Communications and Mobile Computing*, and a Guest Editor of *EURASIP Journal on Advances in Signal Processing*. His current research interests are in wireless vehicular and sensor networks, location estimation and cellular networks.

Wonyoung Park received a BS in Electronic Communication Engineering from Hanyang University, Seoul, Korea in 1998, an MS and a Ph.D. degrees in Electrical Engineering & Computer Science from KAIST, Daejeon, Korea in 2000 and 2007, respectively. He has been with Samsung Electronics, telecommunication systems business since 2006. He is currently working as a senior engineer in the field of LTE base station research and development.