Combined Packet Scheduling and Call Admission Control with Minimum Throughput Guarantee in Wireless Networks

Hyang-Won Lee, *Student Member, IEEE,* and Song Chong, *Member, IEEE*

***Abstract*— In this paper, a scheduling problem is considered in the cellular network where there exist CBR (constant bit rate) users requiring exact minimum average throughput guarantee, and EMG (elastic with minimum guarantee) users requiring minimum average throughput guarantee and more if possible. We propose a combined scheduling and call admission control algorithm that exactly guarantees the minimum requirements of CBR and EMG users, and then allocates the leftover capacity to EMG users. The proposed algorithm is developed using utility maximization problem without minimum throughput constraints and newly defined utility functions. In the algorithm, it is easy to give priority to particular users so that their requirements are guaranteed prior to any other user. Moreover, the priority structure enables the proposed measurement-based call admission control algorithm to perform admission trial without affecting the minimum required throughput of ongoing users. We verify the performance of our algorithm through mathematical analysis and simulations.**

***Index Terms*— Scheduling algorithm, call admission control, minimum average throughput guarantee, cellular downlink.**

I. Introduction

QUALITY of service (QoS) in wireless networks has become a very important issue as the wireless systems have been required to support high data rates and various applications. As a consequence, many schedulers have been proposed for QoS guarantee in wireless networks. In such QoS schedulers, it is important not only to guarantee the minimum requirements of users but also to cope with the case where not all the requirements can be satisfied. If the requirements of ongoing users cannot be fulfilled completely, it would be desirable that the scheduler guarantees their requirements in the order of predetermined priorities. Moreover, if the admission of a new arrival deteriorates the minimum required performance of ongoing users, then it should be blocked to maintain feasibility.

In [1] and [2], the authors present two scheduling algorithms including M-LWDF (modified largest weighted delay first) and EXP. For each time slot, the M-LWDF algorithm selects the user having the maximum decision metric[1] which is the product of the achievable data rate and HOL (head-of-line) delay of packet queue. If the scheduler is combined with token counter, it can provide minimum throughput guarantee. The EXP algorithm uses exponential function of HOL delays instead of just HOL delay, and it achieves better performance than M-LWDF. In fact, these schedulers belong to the class of MaxWeight policy which is known to be throughput optimal, meaning that it stabilizes a system whenever stability is achievable. Similar to M-LWDF and EXP, one could consider modifying the MaxWeight scheduling rule proposed by Stolyar [3] in order to develop a QoS scheduling algorithm.

The scheduling algorithm in [4] guarantees minimum performance by maximizing the expected utility subject to explicit minimum performance constraints. Those constraints however can cause the feasibility problem. If the constraints are feasible, then the problem can be solved and thus the scheduler developed through the problem provides minimum performance guarantee as designed. But otherwise, the scheduler will produce unpredictably fluctuating results because it will keep trying to satisfy the constraints which cannot be satisfied at all. Unfortunately, it is extremely hard to know the feasibility of such constraints due to randomly time-varying wireless channels. In [5], Andrews et al. consider asymptotic utility maximization problem with minimum and maximum rate constraints. They propose a solution to the problem (i.e., scheduling algorithm) by modifying the token counter suggested in [1]. In the paper, two specific forms of the scheduling algorithms are shown to guarantee the minimum and maximum rates.

In this paper, we consider a single-cell downlink where a single carrier is used and only one user is served at a time. There are two classes of users (or applications) in the system. One is CBR (constant bit rate) application which generates the data at some fixed rate, e.g., voice. The performance of such application severely degrades if the minimum throughput (usually encoding rate) is not guaranteed, so they would demand minimum throughput guarantee. However, allocating more throughput than the requirement is nothing but the waste of resource because they cannot utilize the excessive throughput. The other is EMG (elastic with minimum guarantee) application which requires minimum throughput guarantee and more if possible. For example, MPEG-4 FGS (fine granularity scalability) enables to freely adjust the video rate to an

Manuscript received January 24, 2006; revised August 25, 2006; accepted November 9, 2006. The associate editor coordinating the review of this paper and approving it for publication was W. Liao. This work was supported by the center for Broadband OFDM Mobile Access (BrOMA) at POSTECH through the ITRC program of the Korean MIC, supervised by IITA (IITA-2006-C1090-0603-0037).

The authors are with the School of Electrical Engineering and Computer Science, Korea Advanced Institute of Science and Technology (e-mail: mslhw@netsys.kaist.ac.kr; song@ee.kaist.ac.kr).

Digital Object Identifier 10.1109/TWC.2007.06004.

[1]The value that a scheduler compares to decide which user to serve.

arbitrary value in real time, as long as the target rate is greater than or equal to that of the base layer[2] [6]. Consequently, such application would require minimum throughput guarantee for minimum acceptable video quality and require more for enhanced quality. Of course, any premium data user could require such QoS guarantee. Note that the elastic traffic, one of the important application types [7], belongs to EMG class with zero minimum requirement. The users in CBR and EMG class excluding elastic users will be called QoS users throughout the paper.

The previous works [1], [2], [4], [5] would fail to support the system of our interest in several aspects. First, they do not describe how to deal with the capacity which remains after guaranteeing the requirements of QoS users. Although it is shown in [1] that the leftover capacity is allocated to non-real time users, the authors do not completely describe how the allocation is accomplished. Second, it seems to be difficult for them to handle the case where the capacity is not enough and thus not all the requirements can be satisfied. Note that even if they adopt call admission control (CAC), the feasibility of the requirements cannot always be maintained due to the time-varying capacity. In this case, it is desirable that the capacity is distributed according to predetermined priorities, i.e., the requirements of the users with high priorities are fulfilled prior to the users of lower priorities. But, to do this, two conditions should be met. One is that the decision metrics of the users with high priorities should be greater than those of other users before the requirements of the high-priority users are fulfilled. The other is that if the requirement of a QoS user is fulfilled, the decision metric of the user should drop to zero so that the remaining capacity can be distributed to other QoS users whose requirements are not fulfilled yet. However, it is difficult to meet the two conditions with the previous scheduling algorithms. Third, QoS scheduling problem has been widely studied in the previous works, but the joint consideration of CAC and QoS scheduling has been hardly discussed in TDMA (time division multiple access) systems. Note however that there are several researches [8], [9] addressing the CAC/QoS problem in wireless communication networks with power control where the performance metric is SINR (signal to interference and noise ratio).

To address the above issues, we define new utility functions and use utility maximization problem, without minimum throughput constraint. We propose a combined scheduling and CAC algorithm that guarantees the minimum requirements of QoS users and then allocates the leftover capacity to EMG users. The proposed scheduling algorithm can easily give *priority* to particular QoS users so that their requirements are guaranteed prior to any other user. Thus, if the requirements of all the ongoing QoS users cannot be fully met, only the high-priority users will be provided minimum throughput guarantee. This feature is important because even if we have CAC, the feasibility cannot always be maintained due to time-varying wireless channels. Moreover, our scheduler *distributes the leftover capacity to EMG users* in such a way that a user with better channel condition yields higher throughput from the leftover capacity with some degree of fairness, and some users could achieve zero throughput from the leftover capacity if their channel conditions are extremely bad relative to other users. Our CAC algorithm is carried out based on measurement, i.e., it accepts and serves a new arrival and decide to admit or block the arrival after certain trial period. In such a *measurement-based call admission control*, it is very important that the admission trial does not deteriorate the minimum performance of ongoing users. We achieve this by taking advantage of the priority structure.

The rest of the paper is organized as follows. In Section II, we discuss the motivation for this work and the background needed to understand this paper. In Section III, we propose a scheduling and CAC algorithm that accomplishes our goal, and mathematically analyze the properties of the algorithm. Our algorithm is examined in Section IV, and finally, we conclude the paper in Section V.

II. Motivation and Background

A. Utility Maximization Problem

Since the seminal work by Kelly [10] which adopted utility maximization problem for network flow control, the utility maximization framework has been frequently used in wired and wireless networks. Recently, Kushner and Whiting have analyzed the optimality and convergence of PF (proportional fairness) scheduler based on the utility maximization problem and some standard results in stochastic approximation [11]. We summarize the result of the paper here.

Consider a time-slotted cellular downlink, and let $R_i(t)$ be the average throughput of user i up to time t. Then, $R_i(t)$ is given by

$$R_i(t) = \frac{\sum_{\tau=1}^{t} r_{i,\tau} I_{i,\tau}}{t} \tag{1}$$

where $r_{i,\tau+1}$ is the achievable data rate of user i during $[\tau, \tau+1)$, i.e., $(\tau+1)$-th time slot, and $I_{i,\tau+1}$ is the indicator function such that $I_{i,\tau+1} = 1$ if user i is chosen at time τ to be served in slot $\tau + 1$ and $I_{i,\tau+1} = 0$ otherwise. (1) can be rewritten in recursive form as

$$R_i(t+1) = R_i(t) + \epsilon_t \left[r_{i,t+1} I_{i,t+1} - R_i(t) \right] \tag{2}$$

where $\epsilon_t = \frac{1}{t+1}$. Define $U(R(t)) = \sum_i \log(R_i(t))$, then by first order Taylor expansion in the neighborhood of $\epsilon_t = 0$, we have

$$U(R(t+1)) - U(R(t)) = \epsilon_t \sum_i \frac{r_{i,t+1} I_{i,t+1} - R_i(t)}{R_i(t)} + O(\epsilon_t^2), \tag{3}$$

of which the derivation is shown for general utility functions in Appendix II. Obviously, selecting user i^* such that

$$i^* = \arg\max_i \frac{r_{i,t+1}}{R_i(t)} \tag{4}$$

maximizes the first order term and as $\epsilon_t \to 0$, the scheduler results in maximizing $\lim_{t\to\infty} U(R(t+1)) - U(R(t))$. $R_i(t)$ dynamically evolves according to (2) and (4) where $r_{i,t+1}$ is defined only stochastically. Thus $R_i(t)$ is a dynamically defined stochastic process, and we need stochastic approximation theory [12] to prove its asymptotic convergence and optimality.

[2]The decoding of the encoded video is impossible without the base layer, and hence, at least the bit rate corresponding to the base layer should be guaranteed for MPEG-4 FGS.

For the proof of convergence, they first show that the limit point of the iteration (2) corresponding to (4) weakly converges to the set of limit points of the solution of an ODE (ordinary differential equation). After that, the existence, uniqueness and global asymptotic stability of the limit points of the ODE are proved. For the optimality, they use the strict concavity of logarithmic utility function and show that the scheduler (4) maximizes $\lim_{t\to\infty} U(R(t))$. As well-known, (4) is nothing but PF scheduler [13] and therefore PF scheduler achieves proportional fairness. All the above results can be extended to the algorithms based on any strictly concave utility function [14]. Thus, one could develop various types of schedulers by applying this result to particularly designed strictly concave utility functions. We will apply this result to newly defined strictly concave utility functions and develop a scheduling and CAC algorithm to accomplish our goal.

B. Minimum Throughput Guarantee

For minimum throughput guarantee, one might want to explicitly add minimum throughput constraints to the utility maximization problem. Obviously, the solution (i.e., scheduler) to the problem will guarantee minimum throughput if feasible, but it is difficult not only to find the solution in time-slotted systems but also to investigate the feasibility of the problem. In practice, CAC is used to maintain feasibility, and QoS scheduling is used to provide minimum throughput guarantee. Namely, CAC is necessary in order for QoS scheduler to work as desired. Motivated by this, we will propose a joint scheduling and CAC algorithm using utility maximization problem with new utility functions, but without minimum throughput constraint.

III. Proposed Algorithm

In this section, we present our algorithm for the time-slotted single-cell downlink where only one user can be served at a time. We first define new utility functions which are strictly concave and increasing. Using the utility functions and Kushner's result [11], we derive a scheduling algorithm which enables to set priority in providing minimum throughput guarantee and distributes the leftover capacity to EMG users. The scheduler is then used for the development of measurement-based call admission control.

A. New Utility Function

Traditionally, the elastic traffic such as FTP is modeled as a strictly concave utility function and the hard real-time traffic (or minimum-guarantee application) is modeled as a step utility function [7], which is usually approximated by a sigmoidal function for mathematical tractability. Because Kushner's analysis holds only for strictly concave utility functions, the traditional utility functions cannot be used for our objective. We would like to redefine the utility functions of QoS users as strictly concave ones. Let $S = C \cup E$ where C and E are the set of CBR users and EMG users, respectively. Denote by R_i the average throughput of user i. We define the utility functions as follows: for $i \in C$,

$$U_i(R_i) = c_i \left\{ 1 - \frac{\log\left(1 + e^{-b_i(R_i - m_i)}\right)}{\log\left(1 + e^{b_i m_i}\right)} \right\} \tag{5}$$

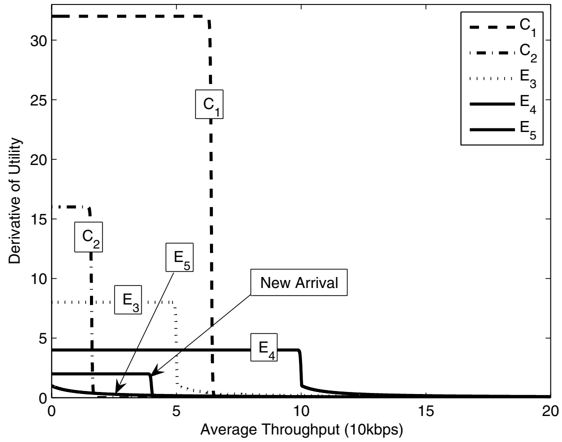


Fig. 1. Derivative of utility functions

and for $i \in E$,

$$U_i(R_i) = \begin{cases} c_i \left\{ 1 - \frac{\log\left(1+e^{-b_i(R_i-m_i)}\right)}{\log\left(1+e^{b_i m_i}\right)} \right\}, & R_i < m_i^\delta \\ a_i \log\left(1 + R_i - m_i^\delta\right) + \Delta_i, & R_i \geq m_i^\delta \end{cases} \tag{6}$$

where a_i, b_i, c_i and Δ_i are positive constants, m_i is the minimum demand rate of user i, and $m_i^\delta = m_i + \delta_i$ where δ_i is a small nonnegative constant. a_i and c_i are determined according to user i's priority. In particular, a_i is set to be equal for all $i \in E$ and we will explain the detail below. b_i is set to be equal for all $i \in S$ and see Appendix I-C for the role of b_i. Given a_i, b_i and c_i, the values of Δ_i and δ_i are selected such that $U_i(R_i)$ and $U_i'(R_i)$ are continuous. See Appendix I-A and I-B for the continuity and strict concavity of (5) and (6). It should be noted that (6) becomes the traditional utility function for elastic user if $m_i^\delta = 0$, and thus it can be viewed as a generalized version of the traditional utility function. We will let $m_i^\delta = 0$ for elastic user. From now on, the region $R_i \geq m_i^\delta$ for EMG users will be called *elastic part*. Note that elastic user has only elastic part and unless otherwise specified, "elastic part" or "elastic part of EMG user" will contain elastic user.

We will informally describe the characteristics of our utility functions when they are used in utility maximization problem. For simplicity of exposition, the QoS class and its utility function are denoted together by $C_i(m_i, c_i, b_i)$ or $E_i(m_i, c_i, b_i, a_i)$. For example, $C_1(10, 5, 50)$ stands for CBR class 1 of which the utility function is given by (5) with parameters $m_1 = 10$, $c_1 = 5$ and $b_1 = 50$. Fig. 1 shows the derivative of the new utility functions corresponding to: $C_1(6.4, 204.8, 50)$, $C_2(1.6, 25.6, 50)$, $E_3(5, 40, 50, 1)$, $E_4(10, 40, 50, 1)$ and $E_5(0, 0, 0, 1)$. Notice that the users in E_5 are elastic users. The parameters are selected using the properties stated in Appendix I-C. Namely, b_i is set to a sufficiently large value for the sharpness of $U_i'(R_i)$, and c_i is set to $h \cdot m_i$ if the height of $U_i'(R_i)$ is wanted to be h. For example, the height of $U_1'(R_1)$ is 32 as c_1 has been set to 6.4×32. We can see that for $R_1 \leq m_1$, the slope of CBR class C_1 is higher than that of any other class. Obviously, this will give the highest priority to C_1 in maximizing the total utility because allocating the throughput to other classes results in less increase of the objective function than allocating to C_1. Consequently, the minimum requirement of C_1 will be satisfied prior to any other class. Moreover, the slope sharply drops down to zero above the minimum requirement so that the other classes can get the chance to be served. Clearly, we can expect the priority relationship $C_1 > C_2 > E_3 > E_4 > E_5$.

The leftover capacity will be shared by EMG classes because the slope of EMG class slowly decreases above its minimum requirement while as that of CBR class sharply drops to zero. All the elastic parts will have the same priority in the sharing of leftover capacity because their corresponding utilities are equivalently given as $a_i \log(1 + R_i)$. Here, a_i is determined such that the elastic part has the lowest priority, and Fig. 1 is actually an example of derivatives with such a_i. As a consequence, the sharing will be similar to the one by PF scheduler where the utility function is $a_i \log(R_i)$. The major difference between the two allocation strategies is that R_i can be zero with $a_i \log(1+R_i)$, but R_i cannot be zero with $a_i \log(R_i)$. Since the throughput allocated to elastic parts is desired to be zero when there is not enough capacity, we must use $a_i \log(1 + R_i)$. We will discuss this issue more formally in the following subsection.

B. Scheduling Algorithm and Its Analysis

Using the new utility functions and stochastic approximation analysis [11], we can easily derive the scheduling policy that maximizes the total utility. The decision metric for general utility functions is given by $r_{j,t+1}U_j'(R_j(t)), \forall j \in S$, and using the metric, the scheduler will select user j^* at time t such that

$$j^* = \arg\max_j r_{j,t+1}U_j'(R_j(t)) \quad (7)$$

where $U_j'(R_j(t))$ is given as

$$U_j'(R_j(t)) = \frac{b_i c_i}{\log\left(1 + e^{b_i m_i}\right)} \cdot \frac{e^{-b_i(R_j(t)-m_i)}}{1 + e^{-b_i(R_j(t)-m_i)}} \quad (8)$$

for $j \in C_i(m_i, c_i, b_i)$ and

$$U_j'(R_j(t)) = \begin{cases} \frac{b_i c_i}{\log\left(1+e^{b_i m_i}\right)} \cdot \frac{e^{-b_i(R_j(t)-m_i)}}{1+e^{-b_i(R_j(t)-m_i)}}, & R_j(t) < m_i^\delta \\ \frac{a_i}{1+R_j(t)-m_i^\delta}, & R_j(t) \geq m_i^\delta. \end{cases} \quad (9)$$

for $j \in E_i(m_i, c_i, b_i, a_i)$. See Appendix II for the derivation of (7). The optimality and convergence of the scheduler can be readily proved following the results in [11] because the utility functions are strictly concave.

In this paper, we analyze the properties of the limit point of algorithm (7) regarding our objectives mentioned in Section I. Let $R_i = \lim_{t\to\infty} R_i(t), \forall i$, and β_i be the mean rate of user i when transmitting, i.e., $\beta_i = \lim_{t\to\infty} \sum_{\tau=1}^{t} r_{i,\tau}/t$ [3]. Then, $\frac{R_i}{\beta_i}$ is the average fraction of time slots allocated to user i, and hence the sum of all the fractions should not exceed 1, i.e., $\sum_i \frac{R_i}{\beta_i} \leq 1$. As a capacity region, we use the intersection of the time fraction constraint and $R \geq 0$ for analytical tractability where $R = [R_i, \forall i \in S]$. Note that the capacity region we will use is actually the convex hull of $R = 0$ and corner points $[R_i = \beta_i \text{ and } R_j = 0, \forall j \neq i], \forall i$, which is the approximated capacity region.

Theorem 3.1: Assume that the capacity region is given by $\sum_i \frac{R_i}{\beta_i} \leq 1$ and $R \geq 0$, and that $\beta_i U_i'(R_i) > \beta_j U_j'(R_j)$ for $R_i < m_i$. Then, user j cannot achieve positive throughput unless the requirement of user i is satisfied.

The proof is given in Appendix III. For homogeneous channels ($\beta_i = \beta_j$), the above theorem shows that we can give priority to user i by setting the derivative as $U_i'(R_i) > U_j'(R_j)$ for $R_i < m_i$. Thus, the utility functions in Fig. 1 set the priority relationship as $C_1 > C_2 > E_1 > E_2 > E_3$. For the case of heterogeneous channels, if the height of $U_i'(R_i)$ is set to a sufficiently large value, then the inequality can be satisfied so that the priority relationship still holds. As discussed in Appendix I-C, the height of $U_i'(R_i)$ can be arbitrarily set by only adjusting the value of c_i without changing the drop-down property. Thus, the priority relationship between users can be arbitrarily established irrespective of channel conditions, even for the users with equal m_i.

We need the following definition in analyzing the share of the leftover capacity.

Definition 3.1: Let $x_i \geq 0$ be the throughput allocated to user $i \in E$, and assume that its feasible region is given by $\sum_i \frac{x_i}{\beta_i} \leq \alpha$ for some positive $\alpha \leq 1$. The throughput vector $x = [x_i, \forall i \in E]$ is said to be PF^z if it holds

$$x_i = \left[\frac{\beta_i}{\lambda} - 1\right]^+, \forall i \quad (10)$$

where $[\cdot]^+ = \max\{0, \cdot\}$ and λ is a positive constant such that $\sum_i \frac{x_i}{\beta_i} = \alpha$.

It is clear that under PF^z throughput allocation, a user with good channel condition (large β_i) will achieve high throughput, and some users with bad channel conditions (small β_i) will get zero throughput. In fact, (10) is the optimality condition to the problem of maximizing the sum of $\log(1+x_i)$ subject to $\sum_i \frac{x_i}{\beta_i} \leq \alpha$. This implies that PF^z allocation is a slightly distorted version of PF where the sum of $\log(x_i)$ is maximized.

Let $f_j(R_j)$ denote the utility function corresponding to the elastic part of EMG user j, i.e., $f_j(R_j) = a_j \log(1 + R_j - m_j^\delta)$ for $R_j \geq m_j^\delta$. Using Theorem 3.1, we will make any elastic part have lower priority than any QoS user by setting $\beta_i U_i'(R_i) > \beta_j f_j'(R_j)$ for any QoS user i and EMG user j when $R_i < m_i$. Let $s \in S$ be a QoS user with the lowest priority, i.e., $\beta_s U_s'(R_s) < \beta_i U_i'(R_i)$ for every QoS user $i \neq s$ when $R_i < m_i$. For each $i \in S$, define $\tilde{R}_i = [\hat{R}_i]^+$ where $\hat{R}_i$ is the solution of $\beta_i U_i'(R_i) = \beta_s U_s'(m_s)$. Note that $\tilde{R}_j$ for elastic user j will be zero because $\hat{R}_j$ is non-positive due to $\beta_s U_s'(R_s) > \beta_j f_j'(R_j)$ for $R_s < m_s$. The drop-down property of the utility function will result in $\tilde{R}_i = m_i + \xi_i$ for each QoS user i where ξ_i is a small positive constant. Note also that even if there are multiple s's, the solution for each of such s will be almost the same because we have approximately equal $\beta_s U_s'(m_s)$ for such s's. So let us assume that $\tilde{R} = [\tilde{R}_i, \forall i \in S]$ is unique.

Theorem 3.2: Assume $\tilde{R} \in \mathcal{F}$, and $\beta_i U_i'(R_i) > \beta_j f_j'(R_j)$ for any QoS user i and EMG user j when $R_i < m_i$. Then, all the minimum requirements of QoS users are fulfilled and

[3]For the equivalence, we actually need the ergodicity of the channels, which is necessary in Kushner's analysis [11].

Algorithm 1 Call Admission Control Algorithm

1: Upon arrival of QoS user k with parameters m_k,c_k,b_k, and a_k(if $k \in E$):
- associate user k with $C_k(m_k, \tilde{c}_k, b_k)$ where $\tilde{c}_k$ satisfies

$$\beta_k U_k'(R_k) < \beta_s U_s'(R_s) \text{ for } R_s < m_s,$$
$$\beta_k U_k'(R_k) > \beta_j f_j'(R_j), \forall j \in E \text{ for } R_k < m_k \quad (11)$$

- serve user k according to (7)

2: After admission trial period,
- if $R_k \geq m_k$,
  - admit user k and associate it with $C_k(m_k, c_k, b_k)$(if $k \in C$) or $E_k(m_k, c_k, b_k, a_k)$(if $k \in E$)
- otherwise, it is blocked.

after that, the leftover capacity is shared by EMG users in PF^z manner.

Theorem 3.2 implies that if there is enough capacity (i.e., $\tilde{R} \in \mathcal{F}$), the proposed scheduler guarantees all the minimum requirements QoS users. Moreover, if the capacity remains after fulfilling all the requirements, EMG users will share the leftover capacity in PF^z manner.

C. Call Admission Control Algorithm

We propose a measurement-based call admission control algorithm jointly operating with the proposed scheduling algorithm. The following theorem shows that with our scheduler, it is possible to perform CAC without affecting the minimum performance of ongoing QoS users. Let user s be the ongoing QoS user of lowest priority defined as above.

Theorem 3.3: Suppose that user k arrives at the system. If $U_k(R_k)$ is selected such that $\beta_k U_k'(R_k) < \beta_s U_s'(R_s)$ for $R_s < m_s$, then serving user k according to (7) does not violate the minimum throughput guarantee of ongoing QoS users.

Proof: The proof is straightforward following Theorem 3.1. ■

Based on the above theorem, we suggest a CAC algorithm in Algorithm 1. When a new call arrives, it is admitted and served by using predefined utility function for admission trial. If its minimum requirement is satisfied after certain trial period, then it is admitted, and its parameters and utility function are set back to the originally intended ones. Otherwise, it is blocked. Note that the purpose of CAC is to test the feasibility of minimum requirements and thus EMG users are served like CBR users during admission trial.

According to Theorem 3.3, the minimum requirements of ongoing QoS users can be protected from admission trial if $\tilde{c}_k$ in Algorithm 1 satisfies the upper condition in (11). In addition, the lower condition in (11) allows user k take the leftover capacity formerly allocated to elastic parts. Let us explain more about the selection of such $\tilde{c}_k$. Namely, the condition (11) can be met if the following inequality holds.

$$2\beta_j a_j \leq \beta_k \frac{\tilde{c}_k}{m_k} \leq \beta_s \frac{c_s}{2m_s}, \forall j \in E. \quad (12)$$

It follows from (12) that

$$\beta_j f_j'(R_j) \leq \beta_j a_j \leq \beta_k \tfrac{\tilde{c}_k}{2m_k} \approx \beta_k U_k'(m_k) < \beta_k U_k'(R_k)$$
$$\beta_k U_k'(R_k) \leq \beta_k \tfrac{\tilde{c}_k}{m_k} \leq \beta_s \tfrac{c_s}{2m_s} \approx \beta_s U_s'(m_s) < \beta_s U_s'(R_s) \quad (13)$$

for $R_k < m_k$ and $R_s < m_s$, which leads to (11). $\tilde{c}_k$ satisfying (12) can be easily obtained if a_j and c_s have been set to a sufficiently small and large value respectively. As an example for homogeneous channels, (12) is rewritten as $2a_j \leq \frac{\tilde{c}_k}{m_k} \leq \frac{c_s}{2m_s}, \forall j \in E$. If $a_j = 1, \forall j \in E$ and $c_s = 4m_s$, then setting $\tilde{c}_k = 2m_k$ satisfies the inequality. Fig. 1 shows such an example. Note that the throughput of elastic parts of EMG users will probably decrease by admission trial, which is not unacceptable according to the properties of EMG users. Note also that β_i in Algorithm 1 is a known parameter if channels are unknown, but otherwise it is computed from running average.

IV. Simulation Results

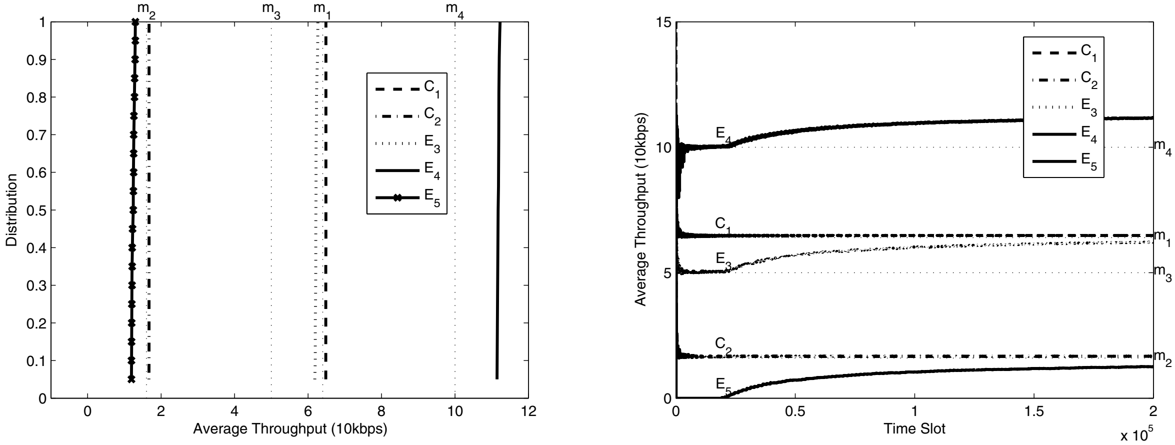
In this section, we demonstrate that our algorithm works as designed through simulations. The results are broken into two parts including, the case of homogeneous channels and the case of heterogeneous channels, and we will examine the characteristics of minimum guarantee, priority, share of leftover capacity and CAC. There are 5 classes including $C_1(64\text{kbps}, 204.8, 50)$, $C_2(16\text{kbps}, 25.6, 50)$, $E_3(50\text{kbps}, 40, 50, 1)$ and $E_4(100\text{kbps}, 40, 50, 1)$ and $E_5(0, 0, 0, 1)$, and each class has 20 users. The derivatives are equivalent to those shown in Fig. 1. We assume that the achievable data rate $r_{i,\tau}$ is given as the Shannon bound, i.e., $r_{i,\tau} = W \log_2(1 + S_i/N_i)$ where W, S_i and N_i are the bandwidth of the channel, received signal power and noise power, respectively.

A. Homogeneous Channel

We examine the performance of our scheduler by varying W in the Shannon bound under homogeneous Rayleigh fading channels where the SNR values S_i/N_i's are independently and identically distributed over all i's for each time slot. The simulation was run over 200000 time slots. Fig. 2 shows the distribution of the average throughput allocated to all users and the average throughput over time when $W =$ 1MHz. As seen in Fig. 2(a), CBR users are exactly guaranteed their minimum average throughput requirements while as EMG users are guaranteed minimum plus certain share of leftover capacity. Observe that $R_i(t) - m_i$'s for $i \in E$ are almost equal and this is the desired result because PF^z share is equal for all users under homogeneous channels. We present the average throughput of five users (one user from each class) over time in Fig. 2(b), from which we can see that those of CBR users converge to their minimum requirements and those of EMG users converge to minimum + PF^z share.

To verify the priority structure, we decrease the value of W[4], and Fig. 3 shows the maximum ($R_i^{\max}$) and minimum ($R_i^{\min}$) of users' achieved throughputs for each class i. $R_i^{\min}$ obviously indicates whether or not the minimum requirement

[4]Note that in practice, W does not change, but we change the value just to see what happens when the system capacity decreases.



(a) Distribution of average throughput

(b) Average throughput over time

Fig. 2. Simulation results with $W = 1$MHz: minimum guarantee and convergence

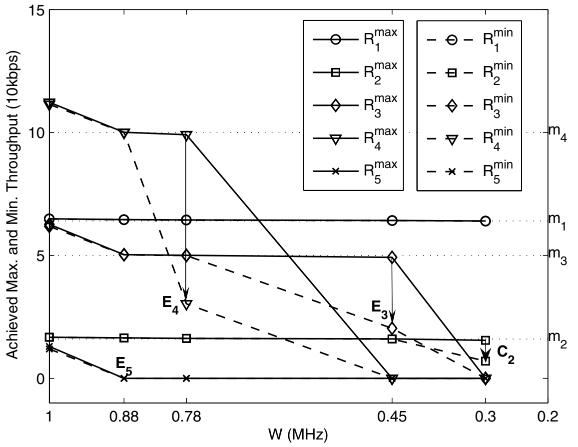


Fig. 3. $R_i^{\max}$ and $R_i^{\min}$ for each class i with different W's: priority

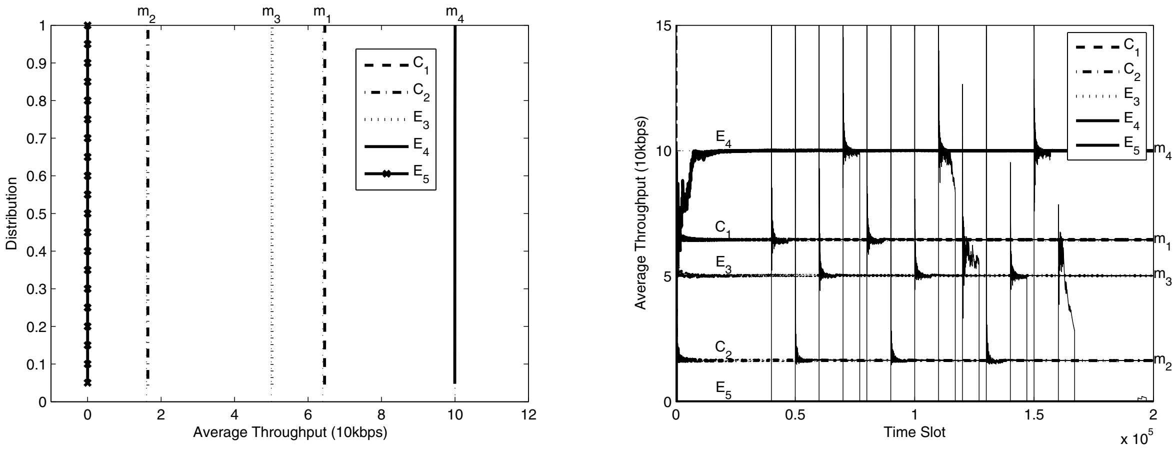
of corresponding class is satisfied. For example, the minimum requirement of E_4 is satisfied when $W = 1$MHz, but not when $W = 0.78$, because some user in E_4 has achieved lower throughput than its requirement. As seen in the figure, all the users in a class achieve zero throughput if the requirement of higher priority class is not satisfied. Therefore, we can say that the priority is given as $C_1 > C_2 > E_3 > E_4 > E_5$, and this is exactly what we expected.

We test our combined scheduling and call admission control algorithm under the scenario where ongoing users exist as above and new users arrive at the system. The admission trial period is set to 7000 slots, and a new call arrives 3000 slots after the decision on the previous arrival is made. The first new call arrives at 40000-th time slot. New users arrive in the order of C_1, C_2, E_3 and E_4, i.e., the sequence of the minimum requirements of new arrivals is 64kbps(CBR), 16kbps(CBR), 50kbps(EMG), 100kbps(EMG), 64kbps(CBR), and so on. As discussed in Subsection III-C, we temporarily set $\tilde{c}_i = 2m_i$ for trial to protect the minimum performance of ongoing users. See Fig. 1 for an example of such $U_i'(R_i)$. We assume that no other new calls arrive during admission trial, which is not impractical because the system can delay the admission trial on new calls for correct decision. W is set to 0.95MHz.

Fig. 4 is the results when there are new arrivals. First, Fig. 4(a) depicts the distribution of the average throughput of ongoing users and newly admitted users. As seen in the figure, the minimum requirements are satisfied, which implies that our CAC maintains the feasibility of requirements. The throughput trace in Fig. 4(b) shows that the arrival 1$\sim$3, 5$\sim$7, 10 and 12 are admitted and the other arrivals are blocked. There can be two reasons for the block of the arrivals. One is insufficient capacity and the other is insufficient time for the convergence of throughput to m_i. For example, the 4-th arrival was blocked while as the 12-th arrival belonging to the same class was admitted. Since the channels are homogeneous, the reason for the block of the 4-th arrival is the latter. We remark that if the admission trial period is set to a larger value, our CAC will make more correct decision, but it will incur larger delay for admission trial. So, there is a tradeoff between the correctness of decision and the latency for decision, and consequently, the admission trial period should be selected according to system requirements. Lastly, observe from Fig. 4(b) that the minimum performance of ongoing QoS users is not affected by new arrivals, which is very important property for measurement-based CAC.

B. Heterogeneous Channel

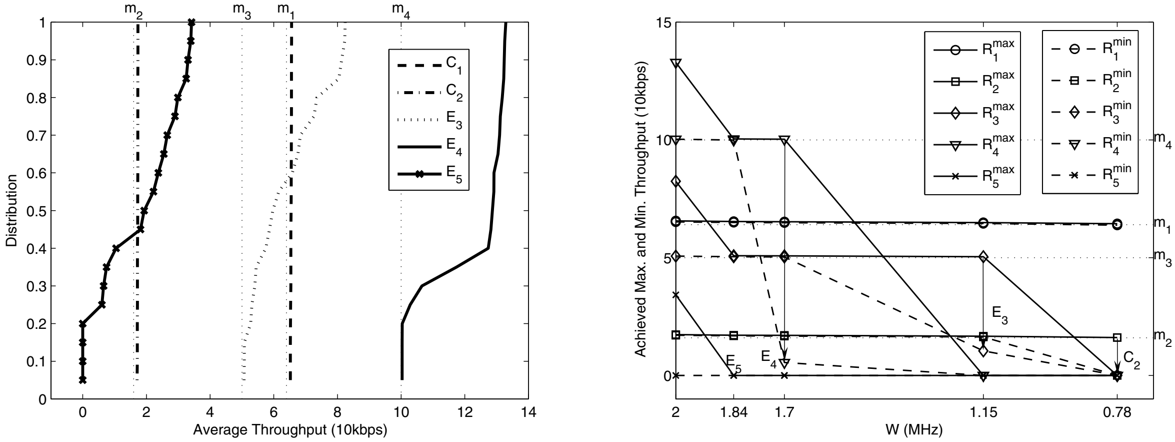
The heterogeneous channels are generated by reflecting path loss and Rayleigh fading. The path loss model is $PL(d) = 16.62 + 37.6\log_{10}(d)$[dB] where d is the distance between a user and the base station in meters. The distances are assumed to be uniformly distributed in $[100, 1000]$m. The Rayleigh fading of each user is generated with different maximum Doppler shifts uniformly distributed in $[6, 200]$Hz. So the means of the SNR values S_i/N_i's are different for different users, but the variances are identical although each user has different changing speed of S_i/N_i over time due to having different Doppler shift. The users and all the parameters are the same as the pervious case except that $b_i = 70, \forall i$, $c_1 = 12800m_1$, $c_2 = 1920m_2$, $c_3 = 320m_3$ and $c_4 =$



(a) Distribution of average throughput with new arrivals

(b) Average throughput over time with new arrivals

Fig. 4. Simulation results with $W = 0.95$MHz: call admission control



(a) Distribution of average throughput with $W = 2$MHz

(b) $R_i^{\max}$ and $R_i^{\min}$ for each class i with different W's: priority

Fig. 5. Simulation results under heterogeneous channels

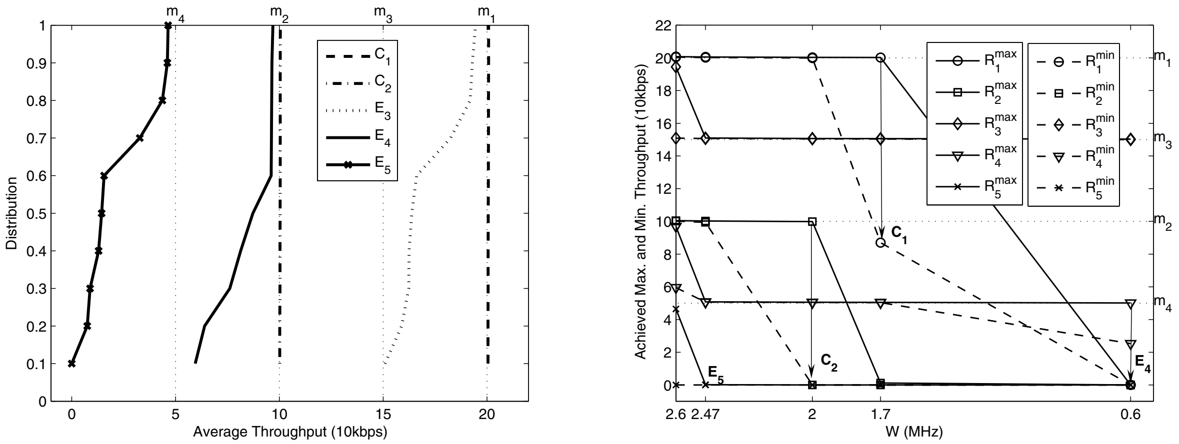
$16m_1$ are used to overcome the heterogeneity of channels in imposing priority. Note that any values of c_i's sufficiently widening the difference in heights of $U_i'(R_i)$'s will work as well. The distribution of throughput with $W = 2$MHz is shown in Fig. 5(a), from which we can see that CBR users are exactly guaranteed their minimum requirements and EMG users are also guaranteed their requirements. In contrast to the previous case, the elastic parts of EMG users attain different shares from leftover capacity. Some elastic parts yield zero throughput from leftover capacity while others yield nonzero throughput up to about 40kbps. We note that the users achieving zero throughput from leftover capacity are far from the base station, thereby reflecting PF^z share. To examine the priority relationship, we show the results with different W's in Fig. 5(b). The priority can be observed as expected, and thus it is possible to impose priority even for heterogeneous channels, by choosing sufficiently large c_i. We could see that the proposed call admission control algorithm also performs as designed.

We simulate a new scenario where the classes are given as $C_1(200\text{kbps}, 6400, 100)$, $C_2(100\text{kbps}, 160, 100)$, $E_3(150\text{kbps}, 192000, 100, 1)$ and $E_4(50\text{kbps}, 9600, 100, 1)$ and $E_5(0, 0, 0, 1)$, and each class has 10 users. The priority is thus given as $E_3 > E_4 > C_1 > C_2 > E_5$. Fig. 6 plots the results and we can observe that the expected results are achieved in this scenario as well.

V. Conclusions

In this paper, we proposed a combined scheduling and call admission control algorithm that guarantees the minimum average throughput requirements according to users' priorities, allocates the leftover capacity to EMG users in PF^z manner, and performs admission trial without deteriorating the minimum performance of ongoing users. We verified the performance of the proposed scheme through mathematical analysis and simulations. Moreover, the proposed algorithm performs as designed in heterogeneous channels as well as homogeneous channels.

Although our work targets multimedia traffic, it is not yet sufficient to be applied in practice. We have considered only minimum throughput guarantee assuming persistent traffic, but the multimedia traffic usually has more complex characteris-



(a) Distribution of average throughput with $W = 2.6\text{MHz}$

(b) $R_i^{\max}$ and $R_i^{\min}$ for each class i with different W's: priority

Fig. 6. Simulation results under heterogeneous channels with new setting of classes

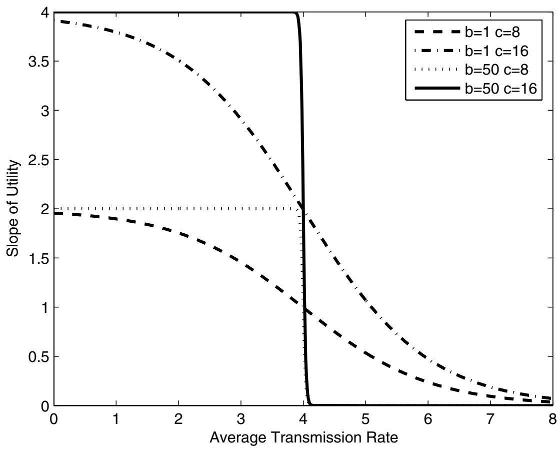


Fig. 7. $U_i'(R_i)$ with different values of b_i and c_i

tics: a) not only minimum rate demand but also peak rate constraints, b) bursty packet arrival rather than infinite backlog, and c) multiple QoS parameters like throughput, delay and delay jitter. So it would be interesting to extend our work by taking into account these characteristics of multimedia traffic. Furthermore, our CAC assumes that at most one user can be under admission trial. More challenging study is to consider CAC problem while allowing multiple users to take admission trial. In fact, this problem was considered for power-controlled wireless networks in [8], but not for TDMA systems. For analytical tractability, we used approximated capacity region represented as a polyhedron. But, the capacity region might be depicted as a nonlinear shape which is hard to describe explicitly. We leave the analysis over the real capacity region as a future study.

APPENDIX I
PROPERTIES OF NEW UTILITY FUNCTION

A. Continuity

For the continuity of $U_i'(R_i)$ in (9), the upper one and the lower one in (9) should be equal at $R_i = m_i^\delta$. By straightforward calculation, we can see that the continuity of $U_i'(R_i)$ holds when δ_i satisfies $a_i = \frac{b_i c_i}{\log\left(1+e^{b_i m_i}\right)} \frac{e^{-b_i \delta_i}}{1+e^{-b_i \delta_i}}$. By using δ_i computed from the equation, Δ_i is determined such that the continuity of $U_i(R_i)$ in (6) is satisfied, and it can be easily shown that $\Delta_i = c_i \left\{ 1 - \frac{\log\left(1+e^{-b_i \delta_i}\right)}{\log\left(1+e^{b_i m_i}\right)} \right\}$ leads to the continuity. Thus, $U_i(R_i)$ in (6) is a continuously differentiable function.

B. Strict Concavity of $U_i(R_i)$

The second derivative of $U_i(R_i)$ in (5) is given by

$$U_i''(R_i) = -\frac{b_i^2 c_i}{\log\left(1 + e^{b_i m_i}\right)} \cdot \frac{e^{-b_i(R_i - m_i)}}{\left(1 + e^{-b_i(R_i - m_i)}\right)^2} < 0$$

which implies that $U_i(R_i)$ in (5) is strictly concave. The strict concavity of $U_i(R_i)$ in (6) can also be proved easily. Therefore, $U_i(R_i)$ is strictly concave.

C. Shape of $U_i'(R_i)$ according to b_i and c_i

The derivative $U_i'(R_i)$ for $i \in C$ is given as (8), and its plot is shown in Fig. 7 for $m_i = 4$ and different values of b_i and c_i. As seen in the figure, the slope with $b_i = 50$ drops to zero above m_i much more sharply than that with $b_i = 1$. Precisely, for large b_i, we have $U_i'(0) \approx \frac{c_i}{m_i}$, $U_i'(m_i - \xi) \approx \frac{c_i}{m_i}$, $U_i'(m_i) \approx \frac{c_i}{2m_i}$ and $U_i'(m_i + \xi) \approx 0$ where ξ is a very small positive value. Thus, R_i will not increase over $m_i + \xi (\approx m_i)$ when $U_i(R_i)$ is plugged into utility maximization problem. For this reason, b_i is set to a large value. When $b_i = 50$ and $c_i = 8$, the height of $U_i'(R_i)$ is 2 and drops to zero above m_i. When $c_i = 16$, the shape is exactly the same as when $c_i = 8$ except that the height is 4, which is doubled. We will take advantage of these properties in developing scheduling and CAC algorithm. $U_i'(R_i)$ for $i \in E$ has the same property except for the part $R_i \geq m_i^\delta$.

APPENDIX II
TAYLOR EXPANSION OF $U(R(t+1)) - U(R(t))$

By the definition of $U(R(t))$ and (2), we can write

$$
\begin{aligned}
&U(R(t+1)) - U(R(t)) \\
&\quad = \sum_{i\in S} U_i(R_i(t+1)) - U_i(R_i(t)) \\
&\quad = \sum_{i\in S} U_i\left(R_i(t) + \epsilon_t\left[r_{i,t+1}I_{i,t+1} - R_i(t)\right]\right) - U_i(R_i(t)).
\end{aligned}
$$

It follows from first order Taylor expansion in the neighborhood of $\epsilon_t = 0$ that

$$
\begin{aligned}
&U(R(t+1)) - U(R(t)) \\
&\quad = \sum_{i\in S} U_i(R_i(t)) + \left[r_{i,t+1}I_{i,t+1} - R_i(t)\right] U_i'(R_i(t))\epsilon_t \\
&\qquad + O(\epsilon_t^2) - U_i(R_i(t)). \\
&\quad = \sum_{i\in S} \left[r_{i,t+1}I_{i,t+1} - R_i(t)\right] U_i'(R_i(t))\epsilon_t + O(\epsilon_t^2) \\
&\quad = \sum_{i\in S} r_{i,t+1}I_{i,t+1}U_i'(R_i(t))\epsilon_t - \sum_{i\in S} R_i(t)U_i'(R_i(t))\epsilon_t \\
&\qquad + O(\epsilon_t^2).
\end{aligned}
$$

Since $I_{i,t+1} \in \{0,1\}$ and $\sum_{i\in S} I_{i,t+1} = 1$, selecting the user having maximum $r_{i,t+1}U_i'(R_i(t))$ maximizes the first order term of $U(R(t+1)) - U(R(t))$.

Appendix III
Proof of Theorem 3.1

According to Kushner's analysis, the limit point of (7) can be viewed as a solution to the following utility maximization problem.

$$
\begin{aligned}
\max \quad & \sum_i U_i(R_i) \\
\text{subject to} \quad & R \in \mathcal{F}
\end{aligned} \tag{14}
$$

where $\mathcal{F}$ is the feasible region of R. Because $\mathcal{F}$ is a convex set (see [4] for the proof) and the objective function is strictly concave, 14) has a unique optimal solution. By assumption, $\mathcal{F}$ is given as $\sum_i \frac{R_i}{\beta_i} \leq 1$ and $R \geq 0$, which is also convex. Let R^* be an optimal solution to (14), and suppose $R_i^* < m_i$ and $R_j^* > 0$.

First, note that R^* exists at the boundary $\mathcal{F}^0$ of $\mathcal{F}$, i.e., $\sum_i \frac{R_i}{\beta_i} = 1$ because all the utility functions are strictly increasing. So, we will confine the feasible region of R to $\mathcal{F}^0$. Since we have $R_j^* > 0$, we can obviously shift a small positive amount $\eta\beta_j$ from R_j^* to R_i^*. In this case, R_i^* should increase by $\eta\beta_i$ so that the new point still remains at $\mathcal{F}^0$. The change in the objective function by this shift is $\eta\beta_i U_i'(R_i^*) - \eta\beta_j U_j'(R_j^*)$ and this change must be non-positive due to the optimality of R^*, i.e.,

$$
\beta_i U_i'(R_i^*) \leq \beta_j U_j'(R_j^*). \tag{15}
$$

This contradicts to the hypothesis $\beta_i U_i'(R_i) > \beta_j U_j'(R_j)$ for $R_i < m_i$. Therefore, $R_i^* < m_i$ and $R_j^* > 0$ are impossible to happen. For better understanding, we show an example of $\beta_i U_i'(R_i)$ and $\beta_j U_j'(R_j)$ satisfying the hypotheses of Theorem 3.1 in Fig. 8. It is easy to see that the inequality (15) cannot hold before the requirement of user i is fulfilled.

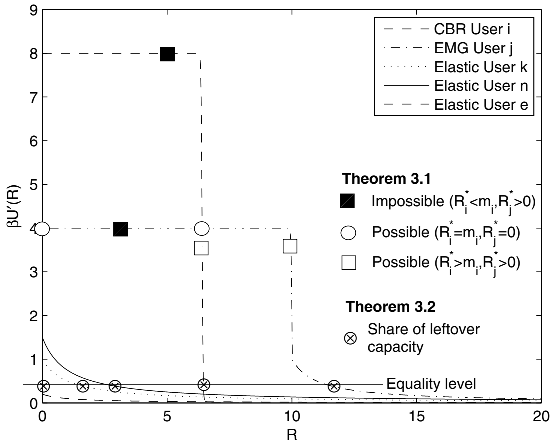


Fig. 8. Example for Theorem 3.1 and 3.2

Appendix IV
Proof of Theorem 3.2

Let R^* be an optimal solution to (14), and suppose $R_s^* < m_s$. It follows from Theorem 3.1 that $R_i^* = 0$ for every elastic user i. It holds $\sum_i U_i\left(R_i^*\right) \geq \sum_i U_i(\tilde{R}_i)$ due to $\tilde{R} \in \mathcal{F}$. Since all the utility functions are strictly increasing and $R_s^* < m_s = \tilde{R}_s$, there must exist QoS user i such that $R_i^* > \tilde{R}_i$. Then, it follows from the strict concavity of the utility function that

$$
\beta_i U_i'(R_i^*) < \beta_i U_i'(\tilde{R}_i) = \beta_s U_s'(m_s) < \beta_s U_s'(R_s^*). \tag{16}
$$

Similar to (15), we can obtain $\beta_s U_s'(R_s^*) \leq \beta_i U_i'(R_i^*)$ by shifting a small amount from R_i^* to R_s^*. This however contradicts to (16). Consequently, we will have $R_s^* \geq m_s$ if the assumption of Theorem 3.2 is satisfied. Since s has the lowest priority, all other minimum requirements will also be satisfied according to Theorem 3.1.

As mentioned in Appendix I-C, $U_i'(R_i)$ for $i \in C$ drops to zero more sharply as b_i increases. So, we can make $R_i^* \approx m_i$ for $i \in C$ by setting b_i to an arbitrarily large value. Similarly, we can also obtain $m_i^\delta \approx m_i$ for $i \in E$. Since all the minimum requirements are satisfied, we can let $R_i = m_i, i \in C$ and $R_i = m_i^\delta + x_i, i \in E$ in (14) where x_i is a nonnegative variable. Note that x_i is the leftover capacity allocated to EMG user i. Since we take equal a_i's in (6), (14) can be reduced as

$$
\begin{aligned}
\max_{x\geq 0} \quad & \sum_{i\in E} \log(1 + x_i) \\
\text{subject to} \quad & \sum_{i\in E} \frac{x_i}{\beta_i} \leq \alpha
\end{aligned} \tag{17}
$$

where $x = [x_i, \forall i \in E]$ and $\alpha = 1 - \sum_{i\in C} m_i/\beta_i - \sum_{i\in E} m_i^\delta/\beta_i$. It is easy to see that the KKT (Karush-Kuhn-Tucker) optimality condition to the problem (17) is given by (10). Therefore, the leftover capacity is shared by EMG users in PF^z manner[5].

For any $x_i, x_j > 0$, (10) can be rearranged as $\frac{\beta_i}{1+x_i} = \frac{\beta_j}{1+x_j}$, which is equivalent to $\beta_i f_i'(R_i) = \beta_j f_j'(R_j)$. Using this, we

[5]If we assume that all the requirements are satisfied for the real capacity region, we can similarly show that $\sum_{i\in E} \log(1 + R_i - m_i^\delta)$ is maximized over the region formed by the intersection of the real capacity region and $R_i \geq m_i, \forall i \in S$.

show an example of PF^z share in Fig. 8 with $a_j = a_k = a_n = a_e = 1$. We can infer from the figure that $\beta_n > \beta_k \approx \beta_j > \beta_e$, and can observe that user e whose channel is relatively bad gets zero throughput from the leftover capacity and other users achieve positive throughput proportional to their channel qualities. User e could achieve positive throughput if there remains more capacity and thus the equality level can go down further.

REFERENCES

[1] S. Shakkottai and A. L. Stolyar, "A study of scheduling algorithms for a mixture of real and non-real time data in hdr," Bell Laboratories, Lucent Technologies, Oct. 2000.

[2] M. Andrews, K. Kumaran, K. Ramanan, A. L. Stolyar, R. Vijayakumar, and P. Whiting, "CDMA data QOS scheduling on the forward link with variable channel conditions," Bell Laboratories, Lucent Technologies, Apr. 2000.

[3] 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, Feb. 2004.

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

[5] M. Andrews, L. Qian, and A. Stolyar, "Optimal utility based multi-user throughput allocation subject to throughput constraints," in *Proc. IEEE INFOCOM 2005*.

[6] W. Li, "Overview of fine granularity scalability in MPEG-4 video standard," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 11, no. 3, pp. 301–317, Mar. 2001.

[7] S. Shenker, "Fundamental design issues for the future Internet," *IEEE J. Sel. Areas Commun.*, vol. 13, no. 7, pp. 1176–1188, Sept. 1995.

[8] N. Bambos, S. C. Chen, and G. J. Pottie, "Channel access algorithms with active link protection for wireless communication networks with power control," *IEEE/ACM Trans. Networking*, vol. 8, no. 5, pp. 583–597, Oct. 2000.

[9] S. Singh, V. Krishnamurthy, and H. V. Poor, "Integrated voice/data call admission control for wireless DS-CDMA systems," *IEEE Trans. Signal Processing*, vol. 50, no. 6, pp. 1483–1495, June 2002.

[10] F. Kelly, A. K. Maulloo, and D. K. H. Tan, "Rate control for communication networks: shadow prices, proportional fairness and stability," *J. Oper. Res. Soc.*, vol. 49, no. 3, pp. 237–252, Sept. 1998.

[11] H. J. Kushner and P. A. Whiting, "Convergence of proportional-fair sharing algorithms under general conditions," *IEEE Trans. Wireless Commun.*, vol. 3, no. 4, pp. 1250–1259, July 2004.

[12] H. J. Kushner and G. G. Yin, *Stochastic Approximation and Recursive Algorithms and Applications*. New York: Springer, 2003.

[13] A. Jalali, R. Padovani, and R. Pankaj, "Data throughput of cdma-hdr: A high efficiency-high data rate personal communication wireless system," in *Proc. IEEE VTC2000-Spring*.

[14] H. J. Kushner and P. A. Whiting, "Convergence of proportional-fair sharing algorithms: extensions of the algorithm," Brown University, Applied Math. LCDS Report, 2003.

Hyang-Won Lee received his B.S. and M.S. degrees in Electrical Engineering and Computer Science from Korea Advanced Institute of Science and Technology (KAIST), Daejeon, Korea, in 2001 and 2003, respectively. He is currently a Ph.D. student in the Department of Electrical Engineering and Computer Science, KAIST. His research interests include congestion/flow control, radio resource management, QoS and optimization in wireless networks.

Song Chong received the B.S. and M.S. degrees in Control and Instrumentation Engineering from Seoul National University, Seoul, Korea, in 1988 and 1990, respectively, and the Ph.D. degree in Electrical and Computer Engineering from the University of Texas at Austin in 1995. In March 2000, he joined the Department of Electrical Engineering and Computer Science, Korea Advanced Institute of Science and Technology (KAIST), Daejeon, Korea. He is currently an Associate Professor, leading the Network Systems Laboratory. Prior to joining KAIST, he was with the Performance Analysis Department, AT&T Bell Laboratories, Holmdel, New Jersey, as a Member of Technical Staff. His research interests include high-speed networks, wireless networks and performance evaluation. He has published more than 50 papers in international journals and conferences and holds three U.S. patents in these areas. He is currently an Editor of the *Journal of Communications and Networks*. He has served as a Technical Program Committee member of IEEE INFOCOM '97, '99, '03, IEEE ICC '05, IEEE GLOBECOM '05, IEEE VTC '06, an Organizing Committee member of PV '01, a Program Committee member of PV '02-'06, a Technical Program Co-chair of ICBN '04.