This full text paper was peer reviewed at the direction of IEEE Communications Society subject matter experts for publication in the ICC 2007 proceedings.

Scheduling and Source Control with Average Queue-Length Control in Cellular Networks

Hyang-Won Lee†, Cheoljung Kim‡, and Song Chong†
† Department of Electrical Engineering and Computer Science, KAIST, Korea
(e-mail: {mslhw@netsys, song@ee}.kaist.ac.kr)
‡ Department of Next Generation WLAN Research, ETRI, Korea
(e-mail: kimcj@etri.re.kr)

***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 and delay guarantee, and EMG (elastic with minimum guarantee) users requiring minimum average throughput and delay guarantee and more throughput if possible. We first define a new utility function of average queue length. Based on the new queue utility function and the throughput utility function proposed before, we design a 2-dimensional weight function that will be used in our scheduling algorithm, and suggest a source control algorithm that can work with the scheduling algorithm. We show through simulations that the proposed algorithm guarantees the QoS (quality of service) requirements of CBR and EMG users.**

I. Introduction

With the rapid growth of the use of portable computing/networking devices, the wireless access system is required to support various applications and high data rates. As a consequence, QoS (quality of service) guarantee in wireless networks has become a very attractive research issue. Especially, the reliable delivery of multimedia applications over wireless networks provide researchers a lot of challenges as the multimedia applications require both throughput and delay guarantee.

In [1] and [2], the authors present two scheduling algorithms including M-LWDF (modified largest weighted delay first) and EXP (exponential). The schedulers give priority to the user with high HOL (head of line) delay and good channel condition. By properly weighting the decision metric[1], they can improve delay performance compared to other schedulers such as Max. C/I and PF (proportional fair) [3]. Song et al. define a utility maximization problem with the utility functions of average delay [4]. The scheduler developed via the problem is shown to improve the delay performance of users. The basic idea behind these schedulers is that if the queue length (delay) is bounded, then the throughput is also guaranteed in average sense. Liu et al. address QoS issues by considering stochastic optimization problem with several types of fairness constraints or minimum performance constraints [5]. In [6], the authors consider an asymptotic utility maximization problem with minimum and maximum rate constraints, and propose a scheduling algorithm solving the problem by modifying the token counter suggested in [1].

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) and delay are not guaranteed, so they would demand minimum throughput and delay guarantee. However, allocating more throughput or delay than the requirement is nothing but the waste of resource because they cannot utilize the excessive. The other is EMG (elastic with minimum guarantee) application which requires minimum average throughput and delay guarantee and more if possible. For example, MPEG-4 FGS (fine granularity scalability) enables to freely adjust the video rate to an arbitrary value in real time, as long as the target rate is greater than or equal to that of the base layer[2] [7]. Consequently, such application would require minimum throughput and delay 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 [8], belongs to EMG class with zero minimum requirement.

To this end, we define a new utility function of average queue length which is strictly concave and decreasing. Based on the new queue utility function and the throughput utility function we proposed before [9], we design a 2-dimensional weight function that will be used in our scheduling algorithm. A source control algorithm working with the scheduler is also suggested using average queue length information. We show through simulations that the proposed algorithm guarantees the minimum throughput requirements of CBR and EMG users, and distributes the leftover throughput to EMG users. Moreover, the average queue length of each user is explicitly controlled in that it converges close to the threshold determined by average delay requirement, which implies that the

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).

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

[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.

This full text paper was peer reviewed at the direction of IEEE Communications Society subject matter experts for publication in the ICC 2007 proceedings.

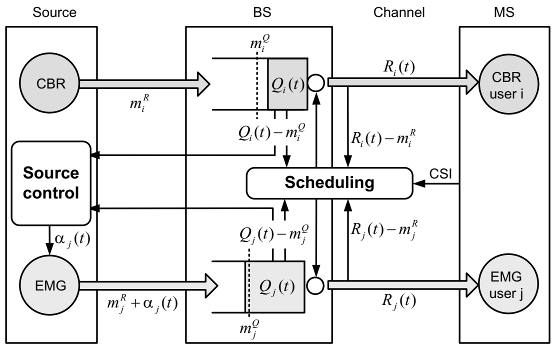


Fig. 1. System model in time-averaged view

delay requirement is satisfied.

II. System Model and Background

A. Notation and Model Description

Consider a single cellular downlink with single carrier where only one user can be served at a time. Let C and E be the set of CBR users and EMG users in the system, respectively. We denote by M the set of all users in the system, i.e., $M = C \cup E$. Let $r_{i,t+1}, t = 0, 1, 2, \cdots$ be the maximum achievable data rate of user i during $[t, t+1)$, i.e., $(t+1)$-th time slot when there is enough data to transmit. If user $i \in M$ is selected to be served during $(t+1)$-th time slot, then the actually achievable data rate $z_{i,t+1}$ will be $\min\{r_{i,t+1}, q_i(t)/T_s\}$ where $q_i(t)$ and T_s are the instantaneous queue length of user i at time t and the length of a time slot, respectively. Define $R_i(t), t = 1, 2, \cdots$ to be the average throughput of user i up to time t. Then, $R_i(t)$ is written as

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

where $I_{i,\tau+1}$ is the indicator function such that $I_{i,\tau+1} = 1$ if user i is selected at time τ to be served during $(\tau+1)$-th time slot, and $I_{i,\tau+1} = 0$ otherwise. Similarly, define $Q_i(t), t = 1, 2, \cdots$ to be the average queue length of user i up to time t. Then, $Q_i(t)$ is written as

$$Q_i(t) = \frac{\sum_{\tau=1}^{t} q_i(\tau)}{t}. \tag{2}$$

Fig. 1 depicts the system model in time-averaged view. CBR source i sends its data at a constant rate m_i^R while as EMG source j sends its data at $m_j^R + \alpha_j(t)$ where $\alpha_j(t) \geq 0$ is the estimate of leftover throughput computed by *source control* using $Q_j(t)$. *Scheduling* is carried out based on the actually achievable data rate $z_{i,t+1}$, average throughput $R_i(t)$ and queue length $Q_i(t)$. Let D_i be the average queueing delay of CBR user i at steady state, and suppose that $Q_i(t)$ converges to $Q_i^* > 0$. Then, it is obvious that the minimum average throughput m_i^R is exactly guaranteed. Moreover, D_i can be written as $\frac{Q_i^*}{m_i^R}$ by Little's Law. So, if Q_i^* is not larger than a certain value m_i^Q, the average delay is guaranteed as $D_i \leq \frac{m_i^Q}{m_i^R}$. For EMG user j, suppose that $Q_j(t)$ and $\alpha_j(t)$ converge to Q_j^* and α_j^*, respectively. In this case, the user will achieve the average throughput $m_j^R + \alpha_j^*$, which is minimum plus leftover throughput. Similarly, it holds $D_j = \frac{Q_j^*}{m_j^R + \alpha_j^*}$ and hence the average delay will be guaranteed as $D_j \leq \frac{m_j^Q}{m_j^R}$ for $Q_j^* \leq m_i^Q$. Based on this argument, the values of m_i^R and m_i^Q for every user $i \in M$ will be determined according to average throughput and delay requirements, and we will propose a scheduling and source control algorithm that guarantees the minimum average throughput requirements (m_i^R) of all users, distributes the leftover capacity (α_i) to EMG users, and explicitly controls the average queue length close to given threshold (m_i^Q).

B. Throughput Utility Function

Each user $i \in M$ is associated with throughput utility function $G_i(R_i(t))$ and queue length utility function $H_i(Q_i(t))$. We adopt the throughput utility function defined in [9] as: for $i \in C$,

$$G_i(R_i(t)) = c_i^R \left\{ 1 - \frac{\log\left(1 + e^{-b_i(R_i(t) - m_i^R)}\right)}{\log\left(1 + e^{b_i m_i^R}\right)} \right\} \tag{3}$$

and for $i \in E$,

$$G_i(R_i(t)) = \begin{cases} c_i^R \left\{ 1 - \frac{\log\left(1+e^{-b_i(R_i(t) - m_i^R)}\right)}{\log\left(1+e^{b_i m_i^R}\right)} \right\}, & R_i(t) < m_i^R \\ a_i \log\left(1 + R_i(t) - m_i^R\right) + \Delta_i, & R_i(t) \geq m_i^R \end{cases} \tag{4}$$

where a_i, b_i, c_i^R and Δ_i are positive constants, m_i^R is the minimum demand rate of user i. In [9], the priority relationship between uses are established by differentiating the values of $\frac{c_i^R}{m_i^R}$ (So, the value of c_i^R determines the priority), but in this paper c_i is selected such that $\frac{c_i^R}{m_i^R}$'s are equal for all $i \in M$ and thus there is no priority relationship between users. $G_i(\cdot)$ is strictly concave and increasing, and see [9] for other details. $H_i(\cdot)$ will be defined in the next section.

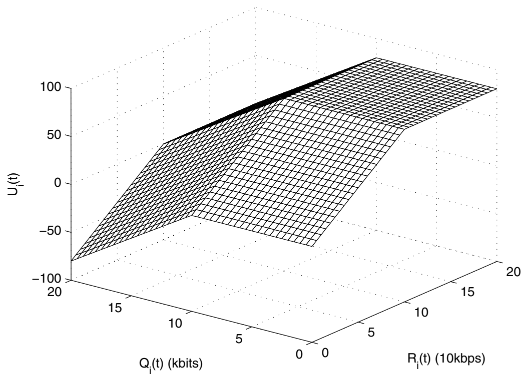
III. Proposed Algorithm

A. New Queue-Length Utility Function

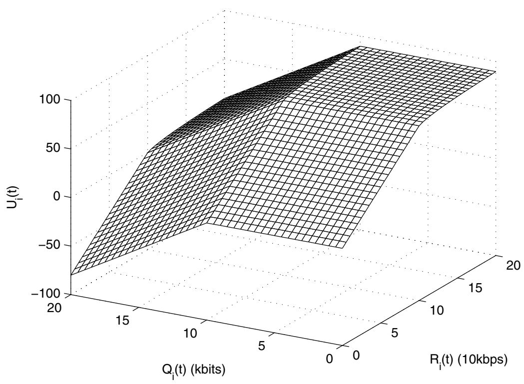
For each $i \in M$, we define a new utility function with respect to average queue length as follows:

$$H_i(Q_i(t)) = -c_i^Q \left\{ Q_i(t) + \frac{1}{b_i} \log\left(1 + e^{-b_i\left(Q_i(t) - m_i^Q\right)}\right) \right\} \tag{5}$$

where b_i and c_i^Q are positive constants, and m_i^Q is the threshold mentioned in Subsection II-A, i.e., the target average queue length. $-H_i'(\cdot)$ is a sigmoid function and b_i is set to a sufficiently large value for the sharpness of $-H_i'(\cdot)$. c_i^Q determines the height of the sigmoid function. If the height of $-H_i'(\cdot)$ is wanted to be h, then we set $c_i^Q = h$. $H_i(\cdot)$ is strictly concave and decreasing function, which is reasonable because as the queue length (delay) increases, the service quality perceived by user is deteriorated. Using

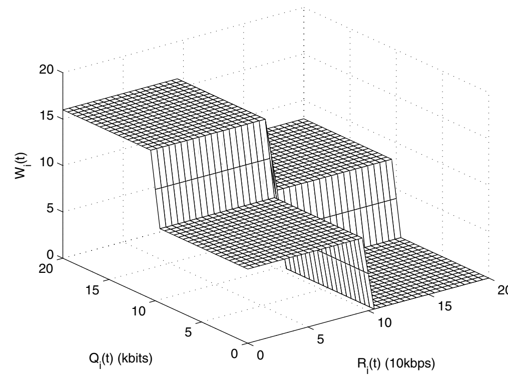


(a) Utility function $U_i(t)$ of CBR user

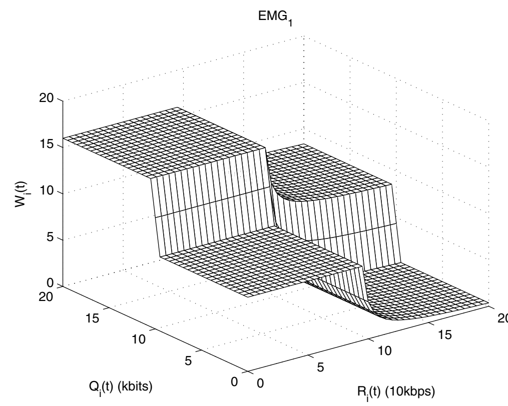


(b) Utility function $U_i(t)$ of EMG user

Fig. 2. Example of utility function



(a) Weight function $W_i(t)$ of CBR user



(b) Weight function $W_i(t)$ of EMG user

Fig. 3. Example of weight function

(3), (4) and (5), user i's utility function $U_i(t)$ is defined as $U_i(t) = G_i(R_i(t)) + H_i(Q_i(t))$.

For simplicity of exposition, QoS class and its utility function are denoted together by $C_i(m_i^R, c_i^R, m_i^Q, c_i^Q, b_i)$ or $E_i(m_i^R, c_i^R, m_i^Q, c_i^Q, a_i, b_i)$. In addition to that, we will let m_i^R and m_i^Q be scaled by 10kbps and 1kbits, respectively. So, for example, C_1(10,80,10,80,50) stands for CBR class 1, requiring average 100kbps and 100ms guarantee, of which the utility function is given by $U_i(t)$ with parameters $m_1^R = 10$, $c_1^R = 80$, $m_1^Q = 10$, $c_1^Q = 80$ and $b_1 = 50$. Fig. 2(a) is the plot of such $U_i(t)$, from which we can see that as $R_i(t)$ increases, $U_i(t)$ also increases for fixed $Q_i(t)$, but $U_i(t)$ is almost constant above m_i^R. This property is an immediate consequence of $G_i(R_i(t))$ which is increasing but bounded. In contrast, $U_i(t)$ is almost constant for $Q_i(t) < m_i^Q$ and fixed $R_i(t)$, and decreases when $Q_i(t)$ goes beyond m_i^Q. Fig. 2(b) plots the utility function corresponding to E_1(10,80,10,80,4,50). We can see that the utility function of EMG user is equivalent to that of CBR user for $R_i(t) < m_i^R$ except that it is not bounded from above for $R_i(t) \geq m_i^R$. The effect of this difference on our scheduler will be discussed below.

B. Scheduling Algorithm

Let $U(t) = \sum_{i \in M} U_i(t)$. Then, it is easy to see from Fig. 2 that the scheduler maximizing $\lim_{t\to\infty} U(t)$ will achieve our objective because $U(t)$ is maximized with $Q_i(t) \leq m_i^Q$ and $R_i(t) \geq m_i^R$ for all $i \in M$. Moreover, if there remains any throughput after satisfying the two conditions, the leftover throughput will be allocated to EMG users since allocating the throughput to CBR users results in less increase of $U(t)$ than allocating to EMG users. Then, the question is how to derive the scheduler that maximizes $\lim_{t\to\infty} U(t)$. We propose a scheduler that maximizes the drift $\Delta U(t)$ of $U(t)$, i.e., $\Delta U(t) = U(t+1) - U(t)$ for each time t as follows. At every time t, select user i^* such that

$$i^* = \arg\max_{i \in M} z_{i,t+1} W_i(t) \tag{6}$$

where $W_i(t)$ is defined as

$$W_i(t) = G_i'(R_i(t)) - T_s H_i'(Q_i(t)). \tag{7}$$

See the appendix for the derivation.

Fig. 3 illustrates an example of $W_i(t)$ with C_1(10,80,10, 8000,50) and E_1(10,80,10,8000,4,50) and T_s = 1ms. As mentioned above, the heights of $G_i'(R_i(t))$ and $H_i'(Q_i(t))$ are respectively given by c_i^R/m_i^R and c_i^Q. The parameters are selected such that the first and second terms in (7) have the equal height of 8. As a consequence, $W_i(t)$ in Fig. 3 has the height of 16.

The scheduler will give high priority to the users with $R_i(t) < m_i^R$ and $Q_i(t) > m_i^Q$, i.e., the users of which both throughput and queue length requirements are unfulfilled. Consequently, such users will be selected by the scheduler (6). For any user $i \in M$, it is not possible that $R_i(t) < m_i^R$ and $Q_i(t) \leq m_i^Q$ in long-term sense because $Q_i(t)$ will keep increasing by excessive arrival. So, if either one of QoS requirements is satisfied, it means that $R_i(t) \geq m_i^R$ and $Q_i(t) > m_i^Q$. For CBR user, $W_i(t)$ sharply drops to medium level (8 in Fig. 3(a)) when such a case happens. $W_i(t)$ of EMG user in Fig. 3(b) also drops to medium level in that case although it does slowly. The scheduler will therefore serve the users of which any of m_i^R and m_i^Q is not satisfied, and then those of which only m_i^R is satisfied. Thus, both m_i^R and m_i^Q will be eventually satisfied for all $i \in M$.

Let us now see how the leftover throughput is distributed. Suppose that $W_i(t)$ of EMG user sharply drops for $R_i(t) > m_i^R$ as that of CBR user, then the scheduler will serve CBR and EMG users without any discrimination after all users' QoS requirements are satisfied. But, $W_i(t)$ we use for EMG user slowly drops so that EMG users are selected when all users' QoS requirements are satisfied. Thus, the leftover throughput will be allocated to EMG users. We remark that if one wants the scheduler to react more strongly on the violation of queue length (respectively, throughput) requirement than throughput (respectively, queue length) requirement, then higher c_i^Q (respectively, c_i^R) would be selected, in which case the dropdown shape of $W_i(t)$ remains the same and only its height changes.

The scheduler is *short-term optimal* in that it finds the steepest ascent direction for every time t, but the long-term optimality is not guarantee. Although we cannot prove its optimality and convergence, such a derivation of scheduler is a common technique for long-term maximization of utility [4], [10].

C. Source Control Algorithm

The purpose of source control is to estimate the leftover throughput and to adapt the source transmission rate of EMG user according to the estimation. We want the leftover throughput to be equally allocated to EMG users, and propose the following update of $\alpha_i(t)$ for $i \in E$.

$$\alpha_i(t+1) = \alpha_i(t) + \gamma \sum_{j \in M} (m_j^Q - Q_j(t)) \tag{8}$$

where γ is a positive step size. We assume that $\alpha_i(t)$ is immediately known to the source, i.e., no communication delay. The basic idea behind the control is that: a) if $Q_i(t) < m_i^Q$, the capacity is being excessively used for delay guarantee and the excessive capacity needs to be transferred to throughput m_i^Q, b) if $Q_i(t) > m_i^Q$, the delay requirement is violated and the leftover throughput $\alpha_i(t)$ should be sacrificed for delay guarantee. All EMG users will receive equal $\alpha_i(t)$ under the source control (8). Notice that the average queue lengths can converge to different values for different users even if $\alpha_i(t)$ converges. Consider the case of EMG user i and arbitrary user j, and suppose $m_i^Q - Q_i^* = -\delta$ and $m_j^Q - Q_j^* = \delta$ for some $\delta > 0$. In this case, $\alpha_i(t)$ converges, but $Q_i^* > m_i^Q$ while $Q_j^* < m_j^Q$. Observe that even slight increase of $Q_i(t)$ from m_i^Q incurs sharp increase of $W_i(t)$ and thus user i is likely to be selected, which leads to the decrease of $Q_i(t)$. Similarly, even slight decrease $Q_j(t)$ from m_j^Q incurs sharp decrease of $W_j(t)$ and thus user j will be hardly selected, which in turn will increase $Q_j(t)$. Therefore, we expect that such δ will be very small, and we will verify this argument in the following section.

IV. Simulation Results

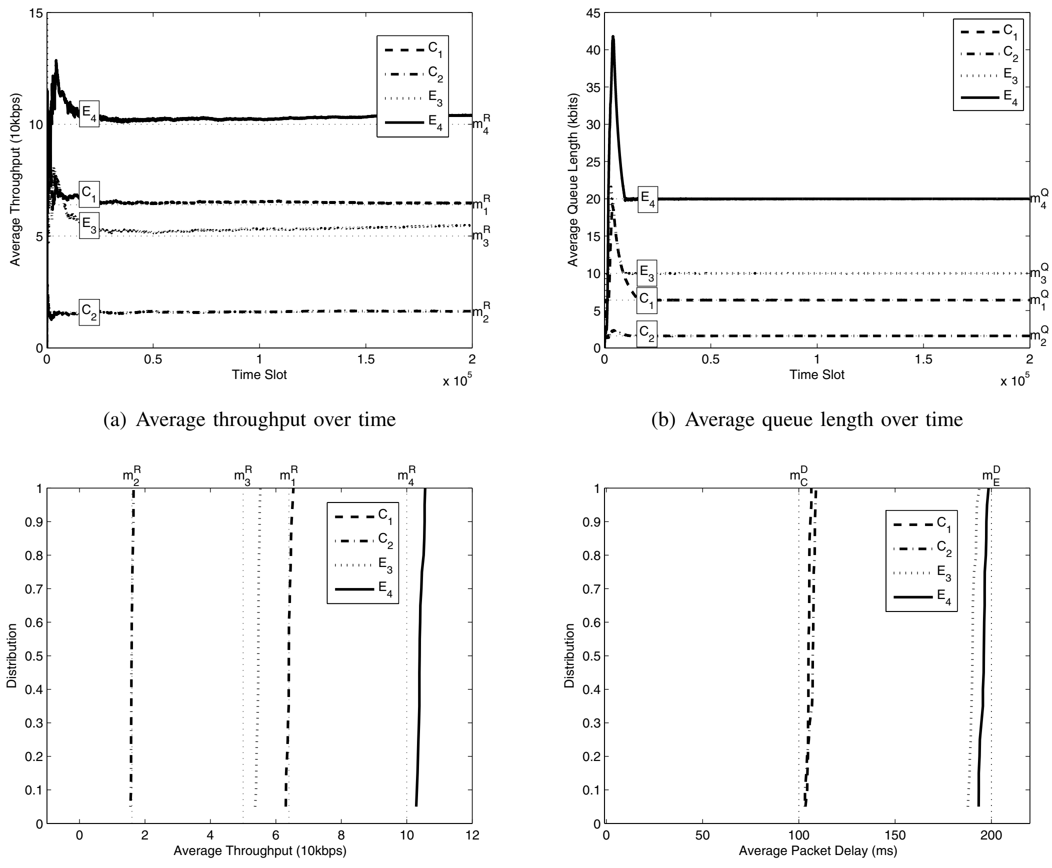
A. Simulation Setup

In this section, we demonstrate our algorithm through simulations. There are 4 classes including C_1(6.4,51.2,6.4,80000,100), $C_2(1.6, 12.8, 1.6, 80000, 100)$, E_3(5,40,10,80000,4,100) and $E_4(10, 80, 20, 80000, 4, 100)$, and each class has 20 users. So, CBR and EMG classes respectively require 100ms and 200ms delay guarantee in average sense. The length T_s of a time slot is set to 1ms and the step size γ in (8) is set to 5×10^{-6}. Since $T_s = 1$ms, the second term in (7) will have the height of 80 while as the first term have 8, which implies that we impose more weight on the violation of queue length requirement. We assume that the achievable data rate $r_{i,t+1}$ is given as [11], i.e., $r_{i,t+1} = B \log_2(1 + \beta\Gamma_{i,t+1})$ where B and $\Gamma_{i,t+1}$ are the frequency bandwidth of the channel and SNR (signal to noise ratio) during $[t, t+1)$, respectively. β is a constant depending on modulation, coding and BER (bit error rate) requirement.

We consider Poisson packet arrival with fixed packet size (125byte). The mean arrival rates of Poisson process for CBR and EMG are given by m_i^R and $m_i^R + \alpha_i(t)$, respectively. $\alpha_i(t)$ is assumed to be fed back to the source without feedback delay. If user i is selected for slot $[t, t+1)$, $z_{i,t+1}T_s$ bits are drained from its queue (so a packet can be partially served at a time and multiple packets can be served at a time). The instantaneous delay of a packet is defined to be the duration between its arrival time and the time when the packet is completely served. The average packet delay of a user is the average of instantaneous delays of all the packets during simulation time. The simulation was run over 200000 time slots under Rayleigh fading channels.

B. Results

The results with $B = 1$MHz are shown in Fig. 4. Fig. 4(a) is the average throughput (over time) of each user randomly selected from each class. We can see that the throughput of CBR class converges close to m_i^R while as the throughput of EMG user converges to m_i^R plus some value. The average queue length plotted in Fig. 4(b) converges close to m_i^Q, which validates the above analysis on the queue length. Fig. 4(c) depicts the distribution of throughput achieved by all users. Observe from the figure that the minimum throughput



(a) Average throughput over time

(b) Average queue length over time

(c) Distribution of average throughput

(d) Distribution of average packet delay

Fig. 4. Simulation results: B =1MHz

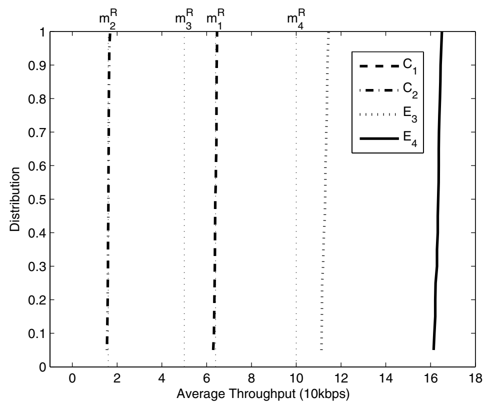
requirement of CBR class is exactly guaranteed and EMG class achieves minimum plus some positive throughput. Hence, the proposed scheduling and source control accomplish our goal regarding throughput. The distribution of average packet delays experienced by each class is shown in Fig. 4(d) where m_C^D and m_E^D denote the delay requirement of CBR and EMG classes respectively. The average delay of CBR class is kept closed to the requirement m_C^D, and the average delay of EMG class is controlled under the requirement m_E^D. This result coincides with the above analysis where the average delay experienced by EMG class was expected to be $Q_i^*/(m_i^R + \alpha_i^*) \leq Q_i^*/m_i^R \approx m_i^Q/m_i^R = m_E^D$. In brief, the average throughput and delay requirements of CBR and EMG classes are guaranteed by the proposed scheduling and source control.

To further examine our algorithm, B is set to 1.6MHz. Note that in practice, B is usually fixed, but we change the value just to see what happens when the system capacity increases. Fig. 5 shows the distribution of throughput and delay. Similar to the previous case, CBR users achieve their minimum throughput requirement m_i^R exactly and the average packet delays are controlled close to the requirement m_C^D. In contrast, EMG users achieve minimum plus about 80kbps throughput and their delays are far below m_E^D. The decreased delay is a consequence of much larger α_i^* than the previous case. This result obviously shows that EMG users share the leftover capacity which remains after fulfilling all the minimum requirements. We remark that the delay requirement of CBR class can be strictly guaranteed by conservatively selecting m_i^Q. In conclusion, our scheduling and source control algorithm works as designed.

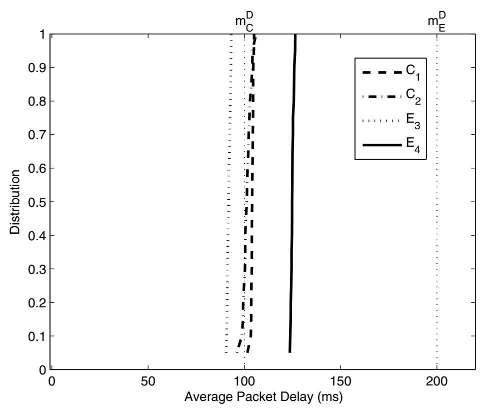
V. CONCLUSIONS

In this paper, we proposed a scheduling and source control algorithm under which CBR users are guaranteed their minimum throughput requirements exactly and maximum tolerable average delay requirements, and EMG user are guaranteed their minimum throughput requirements, maximum tolerable average delay requirements and the leftover throughput. The key idea is to explicitly control average queue length to a target value determined by delay requirement.

This work targets the service of multimedia applications in the cellular networks, but there are still many issues to be solved. First, the performance of our algorithm needs to be validated in the presence of communication delay between source and scheduling entity. Because the source of multimedia application is usually far away from the node where scheduling is implemented, $\alpha_i(t)$ inevitably experiences



(a) Distribution of average throughput



(b) Distribution average packet delay

Fig. 5. Simulation results: B =1.6MHz

delay to reach the source. The impact of this communication delay was not studied in this paper. Second, what is more important for the acceptable quality of multimedia application would be instantaneous delay guarantee rather than average delay guarantee. So, it will be interesting to investigate how our control affects the instantaneous delay and to develop an algorithm taking into account the instantaneous delay if necessary.

APPENDIX

(1) and (2) can be rewritten in recursive form as

$$R_i(t+1) = R_i(t) + \epsilon_t \left[z_{i,t+1} I_{i,t+1} - R_i(t)\right] \tag{9}$$

$$Q_i(t+1) = Q_i(t) + \epsilon_t \left[q_i(t+1) - Q_i(t)\right] \tag{10}$$

where $\epsilon_t = \frac{1}{t+1}$. We assume that for all $t = 0, 1, 2, \cdots$, the data arriving at the queue during $[t, t+1)$ cannot be served at that time slot, and express the evolution of instantaneous queue length as

$$q_i(t+1) = \left[q_i(t) - T_s z_{i,t+1} I_{i,t+1}\right]^+ + a_{i,t+1} \tag{11}$$

where $[\cdot]^+$ denotes $\max\{0, \cdot\}$ and $a_{i,t+1}$ is the amount of arrived data during $[t, t+1)$.

By the definition of $U(t)$, (9) and (10), we can write

$$\begin{aligned} &U(t+1) - U(t) \\ &= \sum_{i \in M} \{G_i\left(R_i(t) + \epsilon_t\left[r_{i,t+1} I_{i,t+1} - R_i(t)\right]\right) - G_i(R_i(t)) \\ &\qquad + H_i\left(Q_i(t) + \epsilon_t\left[q_i(t+1) - Q_i(t)\right]\right) - H_i(Q_i(t))\} \end{aligned}$$

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

$$\begin{aligned} &U(t+1) - U(t) = \epsilon_t \sum_{i \in M} \Big\{ \left[z_{i,t+1} I_{i,t+1} - R_i(t)\right] G_i'(R_i(t)) \\ &+ \Big[\left[q_i(t) - T_s z_{i,t+1} I_{i,t+1}\right]^+ + a_{i,t+1} - Q_i(t)\Big] H_i'(Q_i(t)) \Big\} \\ &+ O(\epsilon_t^2). \end{aligned}$$

Because $z_{i,t+1} = \min\left\{\frac{q_i(t)}{T_s}, r_{i,t+1}\right\}$, we can drop $[\cdot]^+$ and hence the above equation can be rearranged as

$$\begin{aligned} &U(t+1) - U(t) \\ &= \epsilon_t \sum_{i \in M} z_{i,t+1} I_{i,t+1} \{G_i'(R_i(t)) - T_s H_i'(Q_i(t))\} + X(\epsilon_t). \end{aligned}$$

where $X(\epsilon_t) = \epsilon_t \sum_{i \in M} \big\{[q_i(t) + a_{i,t+1} - Q_i(t)] H_i'(Q_i(t)) - R_i(t) G_i'(R_i(t))\big\} + O(\epsilon_t^2)$. Therefore, the scheduler (6) maximizes the first order term of $U(t+1) - U(t)$ and results in maximizing $U(t+1) - U(t)$ as $t \to \infty$.

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. 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*, Tokyo, May 2000.

[4] G. Song, Y. G. Li, L. J. Cimini, and H. Zheng, "Joint channel-aware and queue-aware data scheduling in multiple shared wireless channels," in *Proc. IEEE WCNC*, Atlanta, Mar. 2004.

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

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

[7] 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.

[8] S. Shenker, "Fundamental design issues for the future internet," *IEEE Journal on Selected Areas Communications*, vol. 13, no. 7, pp. 1176–1188, Sep. 1995.

[9] H.-W. Lee and S. Chong, "Combined qos scheduling and call admission control algorithm in cellular networks," in *Proc. WiOpt 2006*, Boston, Apr. 2006.

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

[11] X. Qiu and K. Chawla, "On the performance of adaptive modulation in cellular systems," *IEEE Transactions on Communications*, vol. 47, no. 6, pp. 884–895, Jun. 1999.