QoS Based Fair Resource Allocation in Multi-Cell TD/CDMA Communication Systems

Vincent Huang and Weihua Zhuang, *Senior Member, IEEE*

***Abstract*— In a wireless multimedia code division multiple access (CDMA) system, the resources in terms of transmission rate and power should be efficiently distributed to each user to guarantee its quality-of-service (QoS) requirements. In this paper, a resource allocation algorithm which combines packet scheduling and power assignment is proposed to achieve efficient resource utilization under QoS constraints. The packet scheduling is based on the fair packet loss sharing (FPLS) principle, and the power assignment is determined by the received power limited (RPL) scheme. The basic idea of FPLS is to schedule the transmission of multimedia packets in such a way that all the users have a fair share of packet loss according to their QoS requirements, which maximizes the number of the served users with QoS satisfaction. The RPL scheme minimizes the received power for each packet. Given the propagation path loss, it in turn minimizes the transmitted power as well. The intercell interference from the scheduled packets is also limited in order to increase the system capacity.**

***Index Terms*— Code-division multiple access (CDMA), fair packet loss sharing, packet scheduling, quality of service (QoS).**

I. INTRODUCTION

THE next generation wireless communication systems are anticipated to provide a broad range of multimedia services (including voice, data, and video) to mobile users, using code division multiple access (CDMA) technologies. Packetized transmission over wireless links makes it possible to achieve a high statistical multiplexing gain via proper medium access control (MAC). A flexible MAC protocol which can efficiently accommodate multimedia traffic is required. One important MAC issue is the packet scheduling. As the capacity of a CDMA system is interference limited [1], [2], the actual achieved capacity heavily depends on the packet scheduling. As a result, packet scheduling is critical in such a system in order to fully exploit the system capacity.

Packet scheduling for wireless communications has been an active research area in recent years. Extension of the generalized processor sharing policy [3] from wireline networks to wireless networks are investigated in [4], [5]. Other packet scheduling strategies for wireless networks have been proposed with QoS differentiation in [6], with wireless channel adaptation in [7], and using the exponential rule for throughput optimality in [8]. The above works mainly focus on scheduling in time-division multiple access (TDMA). Packet scheduling in a CDMA system poses more significant technical challenges, due to the additional dimension of power allocation for the non-orthogonal transmissions. Medium access with packet scheduling has been investigated for CDMA systems without time-division (TD) component [9], [10], and for hybrid TD/CDMA [11], [12], all for a single-cell CDMA system. To the best of our knowledge, very few works are available in the open literature for packet scheduling in a multi-cell system, taking into account the dynamics of intercell interference.

In this paper, we propose a packet scheduling algorithm for a multi-cell TD/CDMA system which exploits statistical multiplexing in both time domain (time division) and code domain (code division) [13]. The algorithm combines fair packet loss sharing (FPLS) principle with a received power limited (RPL) power assignment scheme. Based on dynamic estimation of the intercell interference, with FPLS, fair resource allocation and QoS satisfaction to all the mobiles can be achieved at the same time. With RPL, the total maximum received power (including intercell interference) is limited at the base station (BS). Thus, when the interference increases, the total transmitted power from the mobiles in the cell decreases, which means that fewer packets can be scheduled. In this way, instead of competing with each other using higher power, high interference can be avoided and the capacity can be increased. Similar to the approach of the truncated power [14], transmissions of packets from mobiles with very bad channels have to be controlled. A packet from a mobile user with a poor channel condition has a higher chance to be scheduled in a low traffic load than in a high traffic load, so as to effectively control the interference level introduced to other cells. With FPLS, if the packet transmission from a mobile user is suspended due to a poor channel condition, the mobile user will be compensated with more resources for transmitting more packets at a later time when the channel condition is improved. This paper is organized as follows. Section II describes the system model, and Section III presents the proposed scheduling algorithm, including both FPLS scheduling and RPL power assignment. In Section IV, the performance of the proposed scheduling and the intercell interference estimation error model are studied, followed by the conclusions in Section V.

II. SYSTEM MODEL

Consider the uplink transmission of a hybrid TD/CDMA system with packetized transmission [15]. Time is partitioned into frames of a constant duration. Each frame is partitioned

---

Manuscript received October 2002; revised November 2003 and November 2004; accepted February 15, 2005. The editor coordinating the review of this paper and approving it for publication is L. Wang. This work was supported by a research grant from the Natural Science and Engineering Research Council (NSERC) of Canada.

Vincent Huang is with Ericsson, EAB/PDY/TN, Ericsson AB, 164 80 Stockholm, Sweden (e-mail: vincent.a.huang@ericsson.com).

Weihua Zhuang is with the Centre for Wireless Communications (CWC), Department of Electrical and Computer Engineering, University of Waterloo, Waterloo, ON N2L 3G1, Canada (e-mail: wzhuang@bbcr.uwaterloo.ca).

Digital Object Identifier 10.1109/TWC.2006.02013.

equally into time slots. Multiple access within each time slot is accomplished by assigning unique pseudo-random noise (PN) code sequence(s) to each mobile. The source information from the mobiles is segmented into packets of equal length, transmitted at a constant rate. Packet transmissions from the mobiles in the cell are synchronized in time. The transmission duration of each packet equals the length of a time slot. The decision on packet transmission is made at the BS frame by frame and is broadcast to the mobiles by a MAC protocol.

Intercell Interference Modeling — In the uplink transmission, the intercell interference comes from all the mobiles in other cells. Consider mobile user i connecting to the mth ($m \neq 0$) BS. The interference seen at the target ($m = 0$) BS from one packet transmission of this mobile user is [16]

$$f_i = p_i \left(\frac{r_m}{r_0}\right)^{\mu} 10^{(\chi_0 - \chi_m)/10} \tag{1}$$

where p_i is the average received power of one packet from mobile user i at its home BS, r_m and r_0 are the distances between the mobile and the mth BS, and the target BS (zero-th BS), respectively, μ is the path-loss exponent, χ_0 and χ_m are iid zero-mean Gaussian random variables with variance σ_χ^2. The component $10^{\chi_m/10}$ characterizes the shadowing effect at any instant. Eq. (1) describes the local average of the interference power, averaged over the relatively fast multipath fading component. If a mobile user is always connected to the BS with minimum path loss, $(r_m/r_0)^{\mu} 10^{(\chi_0-\chi_m)/10}$ is less than unity. The total intercell interference at the target BS in a time slot is given by

$$\sum_{\text{all packets}} p_i \left(\frac{r_m}{r_0}\right)^{\mu} 10^{(\chi_0-\chi_m)/10} \delta\left(\chi_0 - \chi_m, \frac{r_0}{r_m}\right)$$

where the summation is taken over all the packets transmitted in the time slot, and

$$\delta\left(\chi_0 - \chi_m, \frac{r_0}{r_m}\right) = \begin{cases} 1, & \left(\frac{r_m}{r_0}\right)^{\mu} 10^{(\chi_0-\chi_m)/10} \leq 1 \\ 0, & \text{otherwise.} \end{cases} \tag{2}$$

If a mobile does not transmit in the time slot, the corresponding received power p_i is zero. If the shadowing process does not change much over each frame duration (d_f), it can be modeled by a first-order autoregressive process given by [17]

$$\chi_m(l) = \zeta^{v_i d_f} \chi_m(l-1) + n(l) \tag{3}$$

where $\chi_m(l)$ is the χ_m value in the lth frame, ζ is correlation between two points separated by one meter, v_i is the velocity of mobile user i, and $n(l)$'s are iid Gaussian random variables with zero mean and standard deviation $\sigma_\chi \sqrt{\frac{1+\zeta^{v_i d_f}}{1-\zeta^{v_i d_f}}}$.

QoS Provisioning — The QoS parameters under consideration are transmission accuracy and delay requirements over the wireless link. The overall transmission accuracy requirement is represented by packet loss and transmission bit error rate (BER) requirements. QoS satisfaction is achieved when a required percentage of packets is accurately received within the delay bound. The minimum bandwidth allocation derived from the delay bound and packet loss rate are often very loose [18]–[20], resulting in low resource utilization. Therefore, our proposed scheduling is directly based on the delay bound and packet loss rate requirements. The difference between the required delay bound and the total accumulated queuing delay up to the time of packet scheduling is referred to as *time-out value* of the packet. Packet loss due to buffer overflow or exceeding the delay bound is characterized by packet loss probability (PLP). The BER requirements are to be guaranteed by properly arranging simultaneous packet transmissions and controlling their received power levels, while the delay and PLP requirements are to be guaranteed by proper packet scheduling.

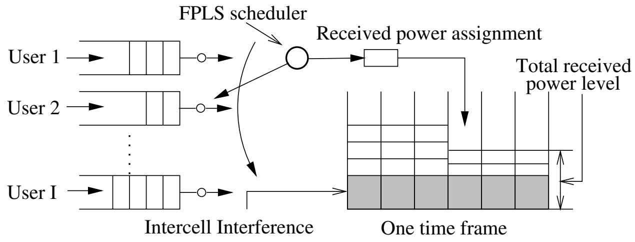


Fig. 1. The packet scheduling procedure, under the assumption that the intercell interference is constant over the time frame.

III. THE MULTI-CELL FAIR PACKET SCHEDULER

The main objectives of the packet scheduler are to guarantee the QoS requirements of all the mobiles and to maximize the resource utilization so that the system can support as many satisfied mobiles as possible. There is a trade-off between the two objectives. High QoS requirements usually results in low resource utilization. The scheduler will provide each mobile with just enough resources to satisfy its QoS requirements without over allocating resources. Since the CDMA system is interference limited, the power assigned to each mobile depends on the multiple access interference which in turn is the power from other mobiles. Scaling down the transmitted power of all mobiles leads to high resource utilization.

Consider a target cell with I active mobile users. By the end of the current frame, the packet scheduler at the BS schedules the packet transmission for the next frame, taking into account the interference from all other cells. To achieve the objectives, the scheduler first decides the order (priority) for the mobiles to transmit their packets. In a high traffic load condition when the number of packets waiting for transmission is in the neighborhood of the system capacity, the priority for transmission should be a function of the delay requirements, packet loss requirements, and traffic flow characteristics. Once the priorities are determined, a packet from the mobile with the highest priority is chosen, as illustrated in Fig. 1. The scheduler then decides if the packet can be scheduled for transmission. If the packet cannot be scheduled due to a poor channel condition or exceeding maximum received power, the mobile user will be given a high priority for transmission in the next frame; Otherwise, the scheduler chooses a time slot for the packet and assigns a proper received power level to the packet. The required received power levels of the packets already scheduled for the same slot are then updated accordingly, due to the intracell interference introduced by the newly scheduled packet.

A. The FPLS Algorithm

When the instantaneous total traffic load exceeds the amount that the system can accommodate, some packets have to be dropped. To focus on the PLP requirement, we first assume that all the packets have the same BER requirements and, therefore, the capacity of the target cell (denoted by C) represents the maximum number of packets that can be transmitted in a frame. The packets with the time-out value equal to one are referred to as *most urgent packets* (MUPs). The MUPs will be dropped if not scheduled for transmission in the next frame. When each mobile has a large enough terminal buffer size, the packet loss happens only due to scheduling when the total number of MUPs exceeds the capacity. To guarantee the PLP requirements, we need to control the dropped MUPs from each and every mobile. Even though the overload of MUPs is caused by some bursty traffic sources during their bursty periods, it is fair to distribute the packet loss among all the mobiles who can tolerate some degree of packet loss, according to their PLP requirements. The packets from a mobile with more stringent delay requirement become MUPs sooner and therefore should be scheduled for transmission sooner. Only if not all the time slots in the frame are fully utilized after all MUPs are scheduled, will the scheduler consider non-MUPs in the order of sequentially increased time-out values, starting with those of the time-out value equal to two.

The number of MUPs from each mobile depends on its traffic rate and delay requirement. In order to determine the number of the MUPs to be dropped for each mobile based on the PLP requirements of all the mobiles in the cell, first we need to establish a relation between the overall PLP requirements (with respect to all the packets including both MUPs and non-MUPs) and the packet loss probabilities with respect only to the MUPs. Let the integer random variable R denote the rate of MUPs in packets/frame from all the mobiles. Let $P_L^{(i)}$ (> 0) denote the PLP upper bound required by mobile user i, $1 \leq i \leq I$. Given the MUP traffic load in a frame, $R = n$, the conditional MUP packet loss probability for mobile i is denoted by $\hat{P}_{M|R}^{(i)}(n)$. Thus, the actual PLP for mobile i, $\hat{P}_L^{(i)}$, is the average number of lost MUPs divided by the average number of generated packets in each frame and is given by

$$\hat{P}_L^{(i)} = \frac{\sum_{n=C+1}^{R_{\max}} \bar{r}_{M_i|R}(n)\hat{P}_{M|R}^{(i)}(n)P(R=n)}{\bar{r}_i} \tag{4}$$

where C is an integer, $R_{\max}$ is the maximum value of the rate R, $\bar{r}_i$ is the average traffic generation rate in packets/frame of mobile i, $\bar{r}_{M_i|R}(n)$ is the conditional average transmission rate of the MUPs from mobile i given $R = n$, and $P(R = n)$ is the probability distribution function of R.

Letting $\xi_{M_i|R}(n) = \frac{\bar{r}_{M_i|R}(n)}{\bar{r}_i}$, $\hat{P}_{M|R}^{(i)}(n)$ is chosen in such a way that the following relation is satisfied

$$\frac{\xi_{M_i|R}(n)}{\xi_{M_j|R}(n)}\frac{\hat{P}_{M|R}^{(i)}(n)}{\hat{P}_{M|R}^{(j)}(n)} = \frac{P_L^{(i)}}{P_L^{(j)}}, \qquad \forall i, \forall j \in \{1, \cdots, I\}. \tag{5}$$

From (4) and (5), it can be observed that when the PLP requirement is satisfied for mobile i, i.e., $\hat{P}_L^{(i)} \leq P_L^{(i)}$, it will be satisfied for all other mobiles in the cell, i.e., $\hat{P}_L^{(j)} \leq P_L^{(j)}$, $1 \leq j \leq I$. This means that, if the BS can guarantee the PLP requirement for one mobile, it can guarantee the PLP requirements for all the mobiles. With FPLS, each mobile has a fair share of packet loss among all the mobiles. The fairness here means that the packet losses are arranged according to the PLP requirements of all the mobiles. Only when the QoS requirements of all mobiles are satisfied at the same time, the number of satisfied mobiles in service is maximized.

Letting n_i denote the MUP load from mobile i given $R = n$, we have $\sum_{i=1}^{I} n_i = n$. With $\hat{P}_{M|R}^{(i)}(n)$, the number of dropped packets from mobile i is $n_i\hat{P}_{M|R}^{(i)}(n)$. The total number of the lost packets is equal to the sum of the lost packets from all the mobiles and is equal to $n - C$ where $n > C$, i.e.,

$$\sum_{i=1}^{I} n_i\hat{P}_{M|R}^{(i)}(n) = n - C, \qquad n > C. \tag{6}$$

From (5)-(6), we can obtain $\hat{P}_{M|R}^{(i)}(n)$ given by

$$\hat{P}_{M|R}^{(i)}(n) = \frac{\frac{1}{\xi_{M_i|R}(n)}P_L^{(i)}}{\sum_{j=1}^{I}\frac{1}{\xi_{M_j|R}(n)}n_jP_L^{(j)}}(n - C), \qquad n > C. \tag{7}$$

Given $R = n$, from the above analysis, we can decide the number of dropped packets for each and every mobile in the frame for a given C value. Taking into account that packets from different mobiles may require different BERs represented in terms of signal to noise plus interference ratio (SINR), the actual value of the capacity C changes with the SINR requirements of the scheduled packets and with the intercell interference and background noise, due to the soft capacity of CDMA systems [1], [2]. This is one main reason why packet scheduling has a great impact on the system capacity. With the unknown capacity C before the scheduling, in the following, we propose a bin-packing scheduling algorithm for guaranteeing both BER and PLP requirements.

The original bin packing problem is a well-known combinatorial problem which deals with the way of packing a set of indivisible blocks into the minimum number of bins. It is known to be NP-complete [21]. In the packet scheduling for each time frame, we consider the time slots as bins and the packets as blocks. The size of each block is the received power level of the packet, and the size of each bin is the capacity of the time slot and is unknown before the scheduling. We want to pack as many blocks as possible in the bins without splitting and without exceeding the size of each bin (i.e., to satisfy the BER requirements of all the simultaneously transmitted packets in the time slot). We first discuss the order of packet transmission here, and then discuss the time slot and power allocation in the next subsection. For the scheduling, we define a priority index κ_i for mobile i, which determines the order of transmission for the mobile among all the mobiles. From (7), when the system capacity C is decreased by one packet, the share of packet loss for mobile i is given by

$$\Delta\kappa_i = \frac{n_i\frac{1}{\xi_{M_i|R}(n)}P_L^{(i)}}{\sum_{j=1}^{I}\frac{1}{\xi_{M_j|R}(n)}n_jP_L^{(j)}}. \tag{8}$$

Correspondingly, if we can increase the system capacity by one packet, the transmitted packet number for mobile i will also be increased by $\Delta\kappa_i$. With the capacity C unknown, we schedule the packets by first assuming $C = 0$ and then increasing the value of C one by one until we cannot schedule any more packets. Under the assumption that all the n MUPs are to be dropped (i.e., with the initial value for C being 0), the initial value of κ_i for the first time frame is given by

$$\kappa_i = n_i \left(1 - \frac{n \frac{1}{\xi_{M_i|R(n)}} P_L^{(i)}}{\sum_{j=1}^{I} \frac{1}{\xi_{M_j|R(n)}} n_j P_L^{(j)}} \right) \tag{9}$$

where $\sum_{i=1}^{I} \kappa_i = 0$. When the value of C is increased by 1, the value of κ_i is increased by $\Delta\kappa_i$, for $i = 1, 2, \ldots, I$. If mobile i has the maximum κ value, i.e., $i = \{i | \kappa_i = \max_{1 \le j \le I} \kappa_j\}$, then the capacity increase of one packet is used to schedule an MUP from the mobile. After that, κ_i is decreased by 1. As a result, the value of κ_i indicates the difference between the calculated number of packets for transmission according to FPLS and the actual number of scheduled packets for mobile i. Mobile i has been over-scheduled if $\kappa_i < 0$ and under-scheduled if $\kappa_i > 0$. Note that $\sum_{i=1}^{I} \Delta\kappa_i = 1$ where $\Delta\kappa_i \in [0, 1]$, and $\sum_{i=1}^{I} \kappa_i = 1$ before scheduling the packet and $\sum_{i=1}^{I} \kappa_i = 0$ after that. When the mobile with highest κ value does not have any more MUP to transmit or the channel condition does not allow it to transmit, MUPs from the mobile with next highest κ value will be scheduled. If there are resources available after all the MUPs are chosen for scheduling, the same scheduling algorithm will be used to schedule the non-MUPs of the time-out value equal to two, and so on. In order to avoid over-scheduling or under-scheduling for each mobile in the long run, the information whether each mobile user is over-scheduled or under-scheduled in the current frame should be carried over for scheduling in the next frame. As a result, for the subsequent time frames, the initial κ_i value is the summation of two components, one given by (9) and the other being the κ_i value from the previous frame.

B. RPL Power Assignment

For a multi-cell system, intercell interference must be considered in the scheduling for both QoS provisioning and high resource utilization. With a higher interference level, the transmitted power for each packet should be increased to maintain a desired SINR, thus causing even higher interference to other users and reducing overall system capacity. One way to control the intercell interference is to assign proper transmitted power to each packet so that a minimum power is used to guarantee the SINR requirement. Given the SINR requirement, the transmitted power is inversely proportional to the time-varying path gain. A truncated power control is proposed in [14], where the transmission is suspended when the path gain is below a cutoff level. This can effectively reduce the interference level and increase the system capacity. However, for a variable traffic load, a fixed cutoff level is not effective. As the required power level depends on other simultaneously transmitted packets in CDMA, we consider joint power assignment and packet scheduling in the following.

Let p_f denote the sum of intercell interference and thermal noise in the time slot, where the intercell interference is the dominant term. The received SINR after despreading for the packet from mobile i, denoted by γ_i, is given by

$$\gamma_i = \frac{G p_i}{\sum_{k=1, k \neq i}^{I} n_k p_k + p_f} \tag{10}$$

where G is the processing gain, p_i is the received power of the desired signal as defined earlier, and n_i is total scheduled packets from mobile i. For presentation clarity, consider two mobiles (mobile i and mobile j) belonging to two different traffic classes, class-I and class-II, respectively. Mobile i has n_i packets and mobile user j has n_j packets for transmission in the time slot. The SINRs are given by

$$\begin{aligned} \gamma_i &= \frac{G p_i}{(n_i - 1) p_i + n_j p_j + p_f} \\ \gamma_j &= \frac{G p_j}{(n_j - 1) p_j + n_i p_i + p_f}. \end{aligned}$$

Solving for p_i and p_j, we have

$$\begin{aligned} p_i &= \frac{(G\gamma_i + \gamma_i \gamma_j) p_f}{[G - \gamma_i (n_i - 1)][G - \gamma_j (n_j - 1)] - \gamma_i \gamma_j n_j n_i} \\ p_j &= \frac{(G\gamma_j + \gamma_i \gamma_j) p_f}{[G - \gamma_i (n_i - 1)][G - \gamma_j (n_j - 1)] - \gamma_i \gamma_j n_j n_i} \end{aligned}$$

which leads to

$$\frac{p_i}{p_j} = \frac{G\gamma_i + \gamma_i \gamma_j}{G\gamma_j + \gamma_i \gamma_j} \triangleq \alpha_{ij}. \tag{11}$$

From (11), the ratio of p_i to p_j is a constant depending only on processing gain and SINR requirements, but independent of the number of packets scheduled in the time slot. Also, from (11), we have $\alpha_{ji} = \frac{1}{\alpha_{ij}}$ and $p_j = \alpha_{ji} p_i$.

For the interference limited system, we want to minimize the transmitted power and thus to reduce interference under the QoS constraints. Given a path gain, minimizing the transmitted power is equivalent to minimizing the received power. Due to the interference limited nature, the number of packets that can be scheduled in each time slot depends on the QoS requirements of the packets waiting for transmission, resulting in a very high complexity for a minimal power solution of all the scheduled packets. As an alternative, we minimize the total power increase for each scheduled packet via time slot allocation. From (11), the received powers for packets from different classes are proportional to each other and independent of the number of scheduled packets. Traffic with different SINR requirements can be viewed as a single traffic class if each assigned power is proportionally scaled. For example, a packet of class-II is equivalent to α_{ji} packets of class-I. Let k denote the equivalent number of class-I packets for the time slot. For a new class-I packet of mobile i, from (10), the received power in the time slot can be calculated from $\gamma_i = \frac{G p_i}{k p_i + p_f}$, resulting in

$$p_i = \frac{\gamma_i p_f}{G - \gamma_i k}, \qquad k < G/\gamma_i,\ G/\gamma_i > 1. \tag{12}$$

The total received power increase due to the newly scheduled packet is $\Delta p_{tot} = p_i + k \Delta p_i$, where Δp_i is the required received power increase of each previously scheduled packet because of the interference generated by the new packet and is given by

$$\Delta p_i = \gamma_i p_f \left[\frac{\gamma_i}{(G-\gamma_i k)^2 + \gamma_i (G - \gamma_i k)} \right]. \tag{13}$$

From the preceding analysis, Δp_{tot} decreases with k. Therefore, the total power increase is minimum if the new packet is scheduled in the time slot with the minimum total power.

Given the processing gain and SINR requirements, the received power levels of the packets in the same time slot are always proportional to each other. The intracell interference caused by one class-II packet is equal to the intracell interference caused by α_{ji} class-I packets. Therefore, we have $p_i = \{\frac{G-\gamma_i[(n_i-1)+n_j\alpha_{ji}]}{\gamma_i p_f}\}^{-1}$. This means, if we increase n_i by one, the increase of the inverse power for each class-I packet is given by $\Delta(\frac{1}{p_i}) = -\frac{1}{p_f}$ and, if we increase n_j by one, the increase of the inverse power for each class-I packet is given by $\Delta(\frac{1}{p_i}) = -\frac{\alpha_{ji}}{p_f}$, where $p_i < p_f$ to avoid a negative power assignment. Thus, the received power levels of the packets already scheduled for transmission in the time slot should be updated accordingly and the received power of a new packet is the corresponding updated power level.

When the intercell interference is increased, the assigned power to each packet has to be increased to maintain the required SINR. This, in turn, will cause more increase in the intercell interference. Therefore, we upper limit the total received power plus interference, $\sum_{i=1}^{I} n_i p_i + p_f$, by a preset threshold in the scheduling. When the intercell interference increases, the number of scheduled packets is reduced, resulting in a decrease of the intercell interference. For each scheduled packet, we want to ensure that both the received power in its own cell and the interference it introduces to other cells are controlled. We first examine whether the total received power is below the threshold when the packet is scheduled, and then check if the sum of intercell interference caused to all the other cells is below a certain level. Instead of a fixed cutoff level for the transmitted power of each mobile, we use a limit on the total intercell interference. Hence, a packet with a large path loss is less likely to be scheduled during a high traffic load period than during a low traffic load period. In this way, we have flexibility in the scheduling and keep the interference level low. Fig. 2 illustrates the flow chart of the proposed scheduling and power assignment procedure, where i and j are the mobile user (MU) indices, $1 \leq i, j \leq I$; l is the time slot index in each frame and L_u the total number of time slots in each frame for the uplink transmission; p^l is the total received power of the scheduled packets in time slot l, Δp^l is the total increased power for already scheduled packets if a new packet is scheduled, $p_{\max}$ is received power threshold for a time slot, $\mathbf{p} = (p^1, p^2, \ldots, p^{L_u})$ which has an initial value $\mathbf{0} = (0, 0, \ldots, 0)$, f_j is the total intercell interference to other cells introduced by a packet from mobile j (as defined earlier), f^l is total intercell interference caused by all already scheduled packets in time slot l, and $f_{\max}$ is intercell interference threshold from the cell to its neighboring cells. With the packet scheduling algorithm, each mobile receives a fair share of the system resources in terms of the number of scheduled packets (i.e., transmission rate) to meet the transmission delay requirement and the received power level to meet the transmission accuracy (BER) requirement. As long as there are sufficient resources, each and every mobile is guaranteed with a minimum amount of resources for QoS satisfaction. In the sense of QoS provisioning, the packet scheduling algorithm is a fair CDMA resource (in both transmission rate and power) allocation scheme [22]–[24].

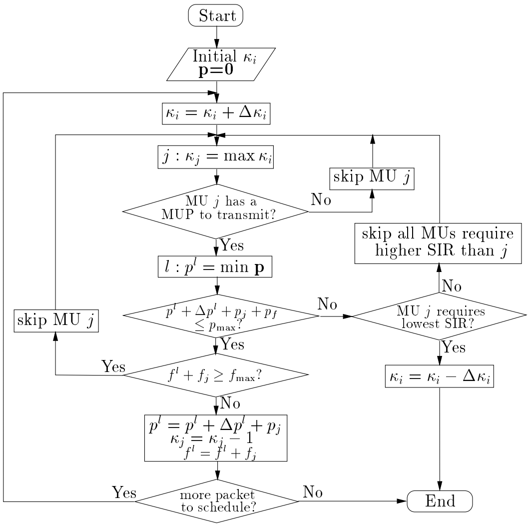


Fig. 2. The proposed resource allocation algorithm for each frame based on FPLS packet scheduling and RPL power assignment.

The knowledge of intercell interference in the next frame is needed to schedule packets for transmission in that frame. In the distributed resource allocation where each BS independently schedules packet transmission in its cell, the knowledge of intercell interference is not available. Taking into account the correlation of path gain (distance and shadowing effect dependent) and assuming that the total traffic load changes slowly from frame to frame, the intercell interference can be predicted from the measured average interference level in the current frame.[1] The prediction error will result in situations where the required SINR cannot be satisfied. Let p_{out} denote the required outage probability (i.e., the probability of received SINR being less than the required threshold should not be larger than p_{out}), and let $\hat{p}_f$ denote the predicted intercell interference, then we have $p_f = \hat{p}_f + \Delta p_f$ and $P\left(\Delta p_f \geq \left(\frac{G}{\gamma_i} - k\right) p_i - \hat{p}_f\right) = p_{out}$. When there are many mobiles in other cells, the prediction error Δp_f can be approximately modeled by a Gaussian distribution $\mathcal{N}(0, \sigma_f^2)$. For a given p_{out}, the allowed value of Δp_f can be calculated numerically by $p_{out} = \frac{1}{2}\left[1 - \text{erf}\left(\frac{\Delta p_f}{\sqrt{2}\sigma_f}\right)\right]$, where $\text{erf}(x) = \frac{2}{\sqrt{\pi}} \int_0^x e^{-t^2} dt$. We can add the Δp_f value to the estimated $\hat{p}_f$ in the power assignment to guarantee the desired outage probability.

---

[1] The prediction is based on the fact that, even for bursty traffic flows, there is a strong correlation in packet transmission from frame to frame as each burst consists of a large number of packets [25], [26]. However, the assumption needs to be further justified based on field measurements on wireless packet data transmission.

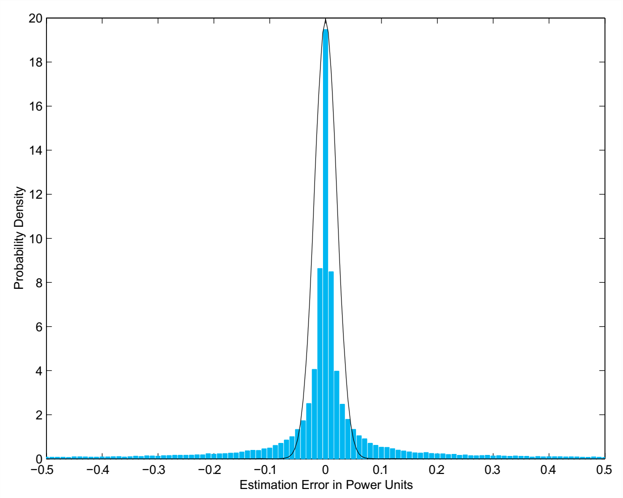
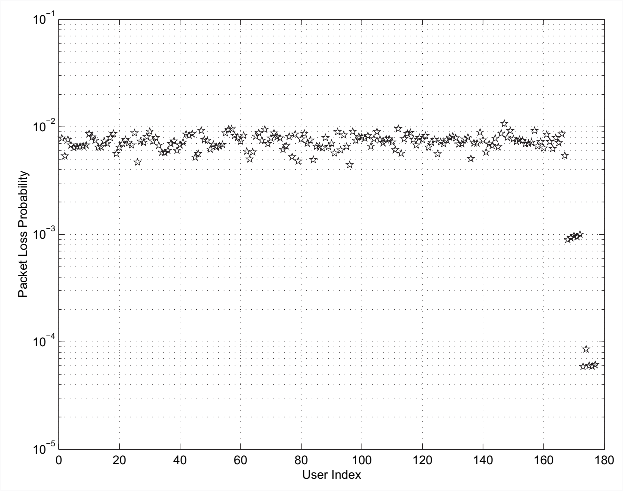


Fig. 3. The intercell interference prediction error (shaded area) and the Gaussian distribution (solid line).

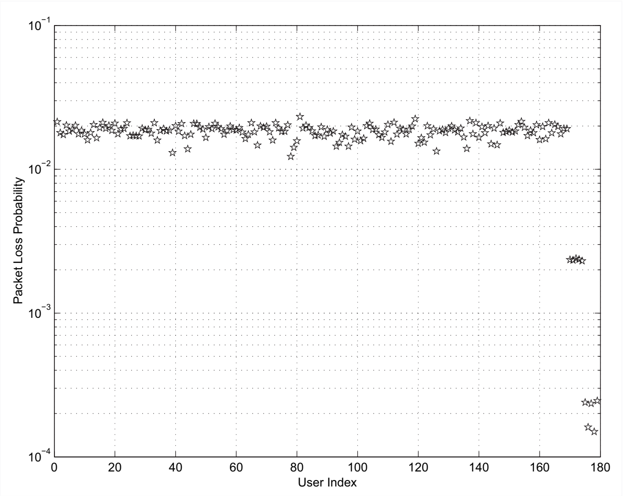
IV. PERFORMANCE EVALUATION

For a single-cell CDMA system, it has been shown that the FPLS algorithm outperforms the previously proposed scheduling algorithms [27]. In the following, we evaluate the performance of the proposed scheduling scheme in a multi-cell environment via computer simulation. For simplicity, consider a target cell surrounded by 6 neighboring cells in the first tier. Each time frame is 10 ms in length and is partitioned into 8 time slots. Three traffic classes (voice, video, and data) are considered. The voice traffic is generated according to the on-off model [15]. During the on-state, one packet is generated in each frame. On average, an on-state lasts ten frames and an off-state lasts 15 frames. The video traffic has a variable rate within zero and 34 packets/frame [28], following a truncated Gaussian distribution with mean 17 packets and standard deviation 0.3×17 packets. The rate is correlated from frame to frame with an autocorrelation coefficient between two consecutive frames of 0.6. The short message transfer protocol (SMTP) is used to simulate the data traffic. The data size of SMTP traffic can be modeled by $\log_2$-normal distribution and data burst arrival rate follows a Poisson distribution [29]. Using the data from Table X in [29], we simulate the SMTP traffic with an average arrival interval of 10 frames and size of each data burst following a $\log_2$-normal distribution with a geometric mean of 10 packets and a geometric standard deviation 3. The SINR requirements are 7 dB for voice and video traffic and 8.75 dB for data. The delay tolerance for voice, video and data packets is 5, 20 and 40 frames, respectively. The PLP requirements for voice, video and data traffic classes are set to 10^{-2}, 10^{-3}, and 10^{-4}, respectively.

The initial mobile location is uniformly distributed over the seven cells. A mobile can move away from the starting point in any direction which is uniformly distributed in $[0, 360^\circ]$. The initial velocity of each mobile is a Gaussian random variable with a mean of 50 km/h and truncated between zero and 100 km/h with a probability of 99%, and the velocity increment from frame to frame is a uniformly distributed random variable within the range $\pm 1\%$ of the current velocity. The variation of the mobile movement direction from frame to frame is uniformly distributed within the range of $\pm 45^\circ$. The cell radius is 2 km, $\mu = 4$, $\sigma_\chi = 8$ dB, and $\zeta = 0.9$. The maximum received power (plus interference and noise) is set at 60 units, with the background thermal noise power being 0.1 units. The interference level is limited at 60 percent of the maximum received power. Each simulation is run for 20000 frames.



(a) With 167 voice users.

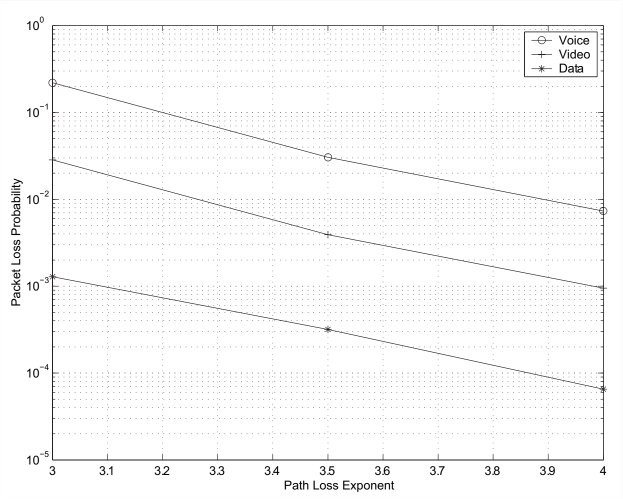


(b) With 169 voice users.

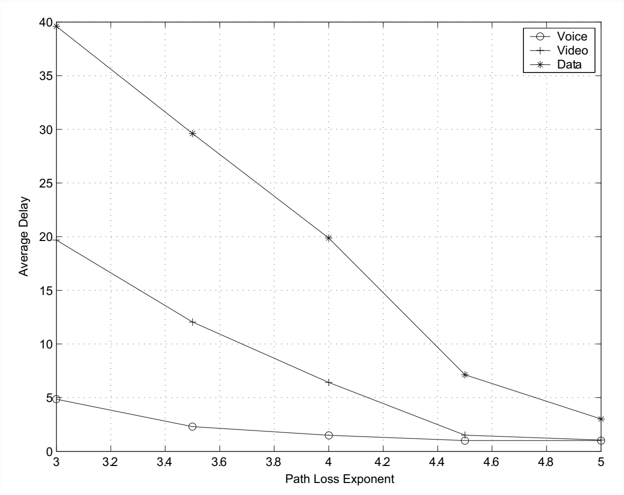
Fig. 4. The average PLP for 167 voice users and 169 voice users, respectively, in the presence of 5 video users and 5 data users.

Simulation results in Fig. 3 demonstrate that the interference prediction error in each frame can be modeled as a Gaussian random variable with a very small variance. This is because of (a) the strong correlation of the voice, video, and data traffic flows from frame to frame, (b) the strong correlation of path gain for each mobile from frame to frame, and (c) a large number of mobiles in the system (180 mobiles per cell). The facts (a)-(b) facilitates the prediction with a small error using the current measured result, and the fact (c) leads to a Gaussian prediction error based on the central limit theorem.

Fig. 4(a) shows the simulation results of PLP for 167 voice mobiles, 5 video mobiles and 5 data mobiles. The PLP requirements for all the mobiles are satisfied at the same time.

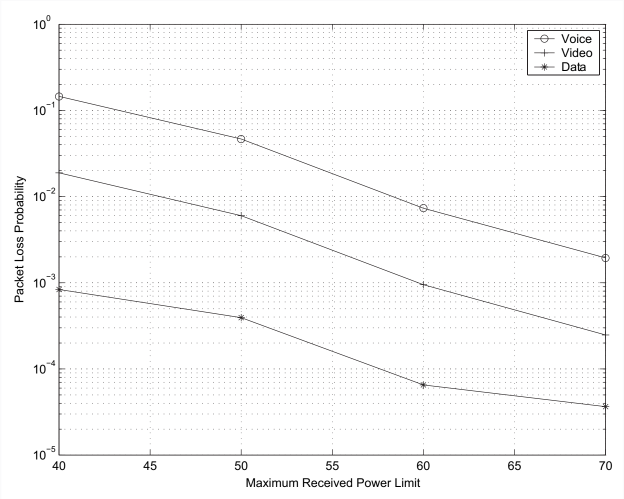


(a) The average PLP.

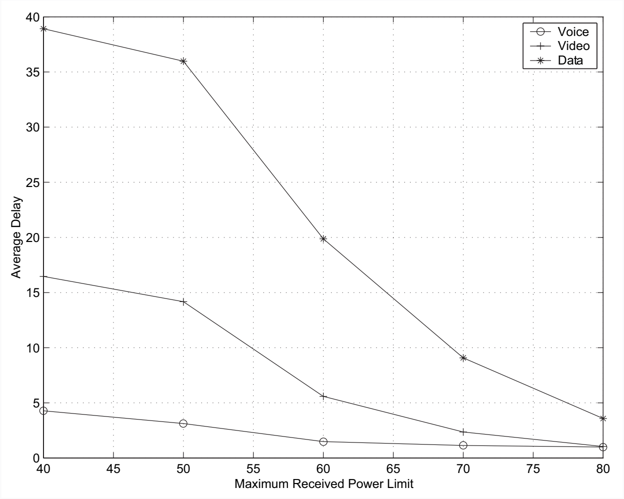


(b) The average delay.

Fig. 5. The average PLP and delay for 167 voice users, 5 video users and 5 data users, versus the path loss exponent μ.



(a) The average PLP.



(b) The average delay.

Fig. 6. The average PLP and delay for 167 voice users, 5 video users and 5 data versus the maximum received power limit.

If we increase the number of voice mobile users by two to 169, as shown in Fig. 4(b), none of the users' PLP requirements can be satisfied. It is observed that the PLP is increased for all the traffic classes. The average delay (in frames) for all the packets is also increased (from 1.49 to 1.87 for voice, from 5.59 to 10.32 for video, and from 19.87 to 27.89 for data). The results demonstrate that, with FPLS packet scheduling, all the mobiles have a fair share of system resources to meet their QoS requirements. With sufficient resources, when one mobile has the QoS satisfaction, so have the other mobiles; When the resources are not sufficient, the QoS of all the mobiles are degraded equally. The results provide a bound on the system capacity in terms of the mobile numbers with QoS satisfaction.

Fig. 5 shows the average PLP and delay as a function of the path loss exponent μ for 167 voice mobiles, 5 video mobiles and 5 data mobiles. The path loss exponent changes from 3 to 5. The average PLP and delay decrease when the path loss index increases. The PLP becomes zero when μ is above 4. When μ is 3, the average delay is very close to the maximum delay limit, which means that most scheduled packets are MUPs. When μ increases, the distance to the base station has more weight in the path gain. Since most mobiles are closer to their home BS, the intercell interference decreases and the system capacity increases. The results demonstrate the link between the QoS parameters and the intercell interference via the propagation parameter μ. With a larger μ value, the interference is smaller because of a larger path attenuation. This results in better QoS and/or higher system capacity. Such relations do not exist in a single-cell CDMA system where there is no intercell interference.

Fig. 6 shows the average PLP and delay as a function of maximum received power limit for 167 voice mobiles, 5 video mobiles and 5 data mobiles. The average PLP and delay decrease when the received power limit increases. When the received power limit is increased to 80, the PLP values decrease to zero. Since the total number of packets can be scheduled in a time slot is limited by the total received power, the capacity is increased when the total received power limit increases. However, the received power for each packet also increases. That is, the better QoS is achieved at the cost of mobile transmit power. The choice of the received power limit depends on the maximum transmit power of each mobile.

V. CONCLUSION

In this paper, we have proposed a multi-cell packet scheduling scheme which combines the FPLS principle with RPL power assignment. The scheduling scheme ensures that the QoS requirements of all the mobiles can be satisfied at the same time, which maximizes the number of satisfied mobiles in service. With the RPL power assignment, the received power is kept at a low level to reduce the intercell interference while meeting the SINR requirements of the mobiles. Simulation results demonstrate that the proposed packet scheduling is efficient in supporting the three traffic classes. The scheduling method requires the prediction of intercell interference. In computer simulation, the measured average interference in the current frame is used as the prediction for the next frame. It is demonstrated via simulations that the prediction error can be modeled by a Gaussian distribution. However, it is necessary to further investigate the intercell interference prediction based on field measurements.

REFERENCES

[1] A.M. Viterbi and A.J. Viterbi, “Erlang capacity of a power controlled CDMA system”, *IEEE J. Select. Areas Commun.*, vol. 11, no. 6, pp. 892-900, Aug. 1993.

[2] A. Sampath, N.B. Mandayam, and J.M. Holztman, “Erlang capacity of a power controlled integrated voice and data CDMA system,” in *Proc. IEEE VTC’97*, pp. 1557-1561.

[3] A. K. Parekh and R. G. Gallager, “A generalized processor sharing approach to flow control in integrated services networks: The single-node case,” *IEEE/ACM Trans. Net.*, vol. 1, pp. 344–357, June 1993.

[4] S. Lu, V. Bharghavan, and R. Srikant, “Fair scheduling in wireless packet networks,” *IEEE/ACM Trans. Networking*, vol. 7, pp. 473–489, Aug. 1999.

[5] T. S. E. Ng, I. Stoica, and H. Zhang, “Packet fair queuing algorithms for wireless networks with location-dependent errors,” in *Proc. IEEE INFOCOM’98*, pp. 1103–1111.

[6] M. Andrews, K. Kumaran, K. Ramanan, A. Stolyar, and P. Whiting, “Providing quality of service over a shared wireless link,” *IEEE Commun. Mag.*, pp. 150-154, Feb. 2001.

[7] X. Liu, E. K. P. Chong, and N. B. Shroff, “Transmission scheduling for efficient wireless utilization,” in *Proc. IEEE INFOCOM’01*, pp. 776-785.

[8] S. Shakkottai and A. L. Stolyar, “Scheduling for multiple flows sharing a time-varying channel: the exponential rule,” *Analytic Methods in Applied Probability*, Series 2, vol. 207, Y. M. Suhov ed., pp. 185-201, 2001.

[9] M. A. Arad and A. Leon-Garcia, “Scheduled CDMA: A hybrid multiple access for wireless ATM networks,” in *Proc. PIMRC’96*, pp. 913–917.

[10] G. D. Mandyam and Y.-C. Tseng, “Packet scheduling in CDMA systems based on power control feedback,” in *Proc. IEEE ICC’01*, pp. 2877-2881.

[11] A. E. Brand and A. H. Aghvami, “Multidimensional PRMA with prioritized Bayesian broadcast: a MAC strategy for multiservice traffic over UMTS,” *IEEE Trans. Veh. Technol.*, vol. 47, pp. 1148–1161, Nov. 1998.

[12] I. F. Akyildiz, D. A. Levine, and I. Joe, “A slotted CDMA protocol with BER scheduling for wireless multimedia networks,” *IEEE/ACM Trans. Networking*, vol. 7, pp. 146–158, Apr. 1999.

[13] J. Blanz, A. Klein, M. Nabhan, and A. Steil, “Performance of a cellular hybrid C/TDMA mobile radio system applying joint detection and coherent receiver antenna diversity,” *IEEE Trans. Veh. Technol.*, vol. 14, no. 4, pp. 568-579, May 1994.

[14] S. W. Kim and A. J. Goldsmith, “Truncated power control in code-division multiple-access communications,” *IEEE Trans. Veh. Technol.*, pp. 965–972, May 2000.

[15] V. Huang and W. Zhuang, “Optimal resource management in packet-switching TDD CDMA systems,” *IEEE Pers. Commun.*, pp. 26–31, Dec. 2000.

[16] A. J. Viterbi, A. M. Viterbi, and E. Zehavi, “Other-cell interference in cellular power-controlled CDMA,” *IEEE Trans. Commun.*, vol. 42, pp. 1501–1504, Feb./Mar./Apr. 1994.

[17] G. L. Stüber, *Principles of Mobile Communication*, Kluwer Academic Publishers, 1996, pp.89-90.

[18] A. Elwalid, D. Heyman, T. V. Lakshman, and D. Mitra, “Fundamental bounds and approximations for ATM multiplexers with applications to video teleconferencing,” *IEEE J. Select. Areas Commun.*, vol. 13, pp. 1004–1016, Aug. 1995.

[19] C.-S. Chang, K.-C. Chen, and M.-Y. Y. and J. F. Chang, “Guaranteed quality-of-service wireless access to ATM networks,” *IEEE J. Select. Areas Commun.*, vol. 15, pp. 106–118, Jan. 1997.

[20] D. Zhao, X. Shen, and J. W. Mark, “QoS performance bounds and efficient connection admission control for heterogeneous services in wireless cellular networks,” *Wireless Networks*, pp. 85–95, 2002.

[21] M. R. Garey and D. S. Johnson, *Computers and intractability: a guide to the theory of NP-completeness*. New York: W. H. Freeman and Company, 1979.

[22] O.-S. Shin and K. B. Lee, “Packet scheduling over a shared wireless link for heterogeneous classes of traffic,” in *Proc. IEEE ICC’04*, vol. 1, pp. 58-62.

[23] K. Kim, H. Kim, and Y. Han, “A proportionally fair scheduling algorithm with QoS and priority in 1xEV-DO,” in *Proc. IEEE PIMRC’02*, vol. 5, pp. 2239-2243.

[24] Q. Qiang, L. Jacob, R. R. Pillai, and B. Prabhakaran, “MAC protocol enhancements for QoS guarantee and fairness over the IEEE 802.11 wireless LANs,” in *Proc. 11th Int. Conf. Computer Communications and Networks*, pp. 628-633, Oct. 2002.

[25] W. Willinger, M. S. Taqqu, R. Sherman, and D. V. Wilson, “Self-similarity through high-variability: statistical analysis of Ethernet LAN traffic at the source level,” *IEEE/ACM Trans. Networking*, vol. 5, no. 1, pp. 71-86, Feb. 1997.

[26] M. Cheng and L. F. Zhang, ”Wireless dynamic channel assignment performance under packet data traffic,” *IEEE J. Select. Areas Commun.*, vol. 17, no. 7, pp. 1257-1269, July 1999.

[27] V. Huang and W. Zhuang, “Fair packet loss sharing (FPLS) bandwidth allocation in wireless multimedia CDMA communications,” in *Proc. 3GWireless’01*, pp. 198–203, 2001.

[28] M. Nomura, T. Fujii, and N. Ohta, “Basic characteristics of variable rate video coding in ATM environment,” *IEEE J. Select. Areas Commun.*, vol. 7, pp. 752–760, 1989.

[29] V. Paxson, “Empirically derived analytic models of wide-area TCP connections,” *IEEE/ACM Trans. Networking*, vol. 2, pp. 316–336, Aug. 1994.

Vincent Huang received the MSc degree in 1994 from Chalmers University of Technology in Gothenburg, Sweden, and the PhD degree in 2003 from the University of Waterloo, Waterloo, Ontario, Canada, both in electrical engineering. He is currently with Ericsson in Stockholm, Sweden. His research interests are MAC protocols, Radio Access Bearer, capacity management and mobility in wireless communications.

Weihua Zhuang (M’93-SM’01) received the B.Sc. and M.Sc. degrees from Dalian Maritime University, China, and the Ph.D. degree from the University of New Brunswick, Canada, all in electrical engineering. Since October 1993, she has been with the Department of Electrical and Computer Engineering, University of Waterloo, ON, Canada, where she is a Professor. She is a co-author of the textbook *Wireless Communications and Networking* (Prentice Hall, 2003). Her current research interests include multimedia wireless communications, wireless networks, and radio positioning. She received the Outstanding Performance Award in 2005 from the University of Waterloo for outstanding achievements in teaching, research, and service, and the Premier’s Research Excellence Award (PREA) in 2001 from the Ontario Government for demonstrated excellence of scientific and academic contributions. She is an Editor/Associate Editor of *IEEE Transactions on Wireless Communications*, *IEEE Transactions on Vehicular Technology*, and *EURASIP Journal on Wireless Communications and Networking*.