Available online at www.sciencedirect.com
SCIENCE DIRECT®
Computer Networks 46 (2004) 867–879
COMPUTER NETWORKS
www.elsevier.com/locate/comnet

QoS-aware radio resource management scheme for CDMA cellular networks based on dynamic interference guard margin (IGM)

Huan Chen [a], Sunil Kumar [b,*], C.-C. Jay Kuo [c]

[a] *Department of Electrical Engineering, National Chung Cheng University, Chiayi 621, Taiwan*
[b] *Department of Electrical and Computer Engineering, Clarkson University, Potsdam, NY 13699, United States*
[c] *Integrated Media Systems Center and Department of Electrical Engineering, University of Southern California, Los Angeles, CA 90089-2564, United States*

Received 13 September 2003; received in revised form 22 March 2004; accepted 16 June 2004
Available online 27 August 2004

Responsible Editor: E. Chong

Abstract

Efficient radio resource management (RRM) for CDMA-based cellular communication systems using the interference guard margin (IGM) scheme is investigated in this work. Two important concepts, i.e., the guard channel (GC) and the load curve (LC), are integrated to derive IGM. The resulting call admission control (CAC) scheme gives preferential treatment to higher priority handoff calls by pre-reserving a certain amount of resource in terms of IGM. A radio resource estimation (RRE) function is implemented in each base station (BS) to assist RRM module to adjust the level of IGM dynamically. RRE in each BS estimates the amount of IGM by considering traffic load in its current cell as well as traffic conditions in neighboring cells. A service model is adopted to support quality of service (QoS) demand of multiple services, which includes mobile terminal's data rate, different levels of priorities, mobility and rate adaptivity characteristics. Simulations are conducted by OPNET to study performance of the proposed IGM scheme in terms of a defined cost function, new call blocking probability, handoff dropping probability and system utilization, under different traffic conditions.

Keywords: CDMA; QoS; Resource reservation; Service model; Handoff; Cellular networks

* Corresponding author. Tel.: +1 315 268 6602.
E-mail addresses: huanchen@ieee.org (H. Chen), skumar@clarkson.edu (S. Kumar), cckuo@sipi.usc.edu (C.-C. Jay Kuo).

1389-1286/$ - see front matter © 2004 Elsevier B.V. All rights reserved.
doi:10.1016/j.comnet.2004.06.028

1. Introduction

The Wide-band CDMA (W-CDMA) technology has emerged as main air interface for 3G wireless systems, which promises to provide a transmission rate from 144 Kbps to 2 Mbps, enabling multimedia services as those provided by broadband wired networks [1,2]. To meet the large bandwidth requirement of multimedia traffic, it is important to utilize system resources efficiently, and provide preferential treatment according to mobile user's traffic profile when the system is congested. The RRM in cellular network systems is responsible for efficiently utilizing air interface resources and guarantee a certain QoS level to different users according to their traffic profiles. The CAC mechanism is an important component of RRM, as it affects the resource management efficiency and QoS guarantees provided to users.

In 2G TDMA/FDMA mobile systems, network accessibility, controlled by the RRM module, is typically designed based on the number of available channels. Due to the limited channel capacity, preferential treatment is given to high priority calls to support them with higher QoS guarantees when the system is congested. Since dropping an ongoing call during handoff is less tolerable than blocking a new call, a handoff call should get a higher priority than a new call request. One way to provide preferential treatment is to reserve a certain number of guard channels (GC) for higher priority calls such as handoff calls.

Various GC schemes have been extensively studied for 2G TDMA/FDMA wireless systems [3–8]. Recently, dynamic GC schemes have been discussed in the literature to improve system utilization while providing QoS guarantees to higher priority calls [5–10]. However, the GC approach is not completely suitable for CDMA systems, because their capacity is limited by the maximum tolerable interference in the system instead of available channels. Since, each new mobile user contributes to the overall level of interference in CDMA systems, a new call is admitted if it does not introduce excessive interference into the system.

The major factor in designing the CAC for CDMA is the system capacity, based on which an incoming call is admitted into system or rejected. Over the past decade, several CAC studies have been made for interference-limit systems based on either transmission power, SIR or other metrics. We refer interested readers to [11] where a comprehensive survey for channel assignment schemes can be found. Viterbi and Viterbi [12] studied Erlang capacity of a power controlled CDMA system and pointed out that the interference level increases rapidly when the system load reaches a certain level. Holma and Toskala [2] studied the effect of interference increase for users with different traffic parameters.

Park et al. [13] investigated the CAC scheme for CDMA forward link, taking into account both the number of codes and interference level. Knutsson et al. [14] also investigated the CAC scheme for downlink communication for CDMA systems. Due to asymmetric traffic conditions in the reverse link (from the mobile to BS) and the forward link (from BS to the mobile), the CAC scheme should admit a call only when the call admission requirements are met in both directions. However, the reverse link capacity is usually more constrained in CDMA systems [15,16,18], and should receive more attention. Huang and Yates [15], and Dimitriou and Tafazolli [16] presented CAC schemes based on transmission power. Grandhi et al. [17] proposed a distributed constrained power control scheme based on SIR measurements to achieve better system capacity. Liu and El Zarki [18] proposed a signal-to-interference ratio (SIR)-based CAC scheme for the reverse link in DS-CDMA systems to improve the system performance under heavy traffic conditions. They assumed that BS received the same signal power from each of its mobile users, and the CAC scheme was designed based on the variation in SIR value. However, such an assumption does not hold in practical systems, where power control is used to keep SIR close to a target value during the whole operation for each mobile user, according to link conditions [19]. Shin et al. [20] proposed an adaptive interference-based channel assignment scheme for DS-CDMA cellular systems. Rather than static channel assignment based on fixed link capacity, their scheme can improve channel capacity as well as service grade by reserving channels for handoff calls. However, their scheme fails to provide admission control to traffic with multiple bandwidth requirements and rate adaptive feature.

In this paper, we present a dynamic resource management scheme for multimedia traffic. A RRE function is implemented in each BS to assist RRM module to dynamically adjust the level of resource needed to be reserved (in terms of IGM) for the use of higher priority traffic, by considering traffic load in its current cell as well as traffic conditions in neighboring cells. The system resource is allocated efficiently by using dynamic resource reservation estimation and rate-adaptive CAC. In our proposed scheme, a constant target SIR value is assumed due to the use of power control in practical systems. This is based on the closed-loop power control in uplink where the BS performs frequent estimates of the received SIR and compares it to the target SIR. This allows the BS to command the mobile station to adjust its power level to maintain the target SIR nearly constant [2,19].

The total interference level in the system is computed by employing the load curve introduced by Holma and Laakso [21]. The use of load curve makes it possible to handle different levels of interference-increase introduced by heterogeneous traffic with various service rates. The salient features of our proposed scheme are summarized below. *First*, the proposed scheme bridges two important concepts, the GC and the load curve, that we call the IGM. *Second*, the proposed scheme supports rate adaptive characteristics for multiple services with flexible QoS guarantees. *Third*, it takes heterogeneous traffic mobilities into consideration to achieve better resource estimation. *Fourth*, by using adaptive RRE, the amount of reserved resource (i.e., IGM) can be dynamically adjusted by referencing the traffic condition in neighboring cells. *Finally*, the resulting CAC scheme gives preferential treatment to higher priority calls by reserving a certain amount of IGM.

The rest of this paper is organized as follows. We provide an overview of capacity and load estimation for CDMA systems in Section 2. In Section 3, the IGM scheme to provide preferential treatment to mobile users in CDMA systems is proposed. It includes the CAC scheme and the associated dynamic RRE method. Section 4 shows the simulation results conducted with OPNET (the OPtimized Network Engineering Tool) by using a service model. Several performance metrics are used to study the comparative performance of the proposed scheme, in terms of the cost function (J), the handoff dropping probability (P_h) and the new call blocking probability (P_n). Finally, concluding remarks and future work are discussed in Section 5.

2. Overview of capacity and load estimation in CDMA systems

As mentioned earlier, the capacity of a CDMA system is limited by the total interference it can tolerate, which is why it is called the interference-limit system. In CDMA systems, each new mobile user contributes to the overall level of interference, and call blocking occurs when the overall interference reaches some level above background noise [12]. Normally, the interference level increases rapidly when the system load reaches a certain level. Users with different traffic profiles and attributes (e.g., service rate, SIR requirement, etc.), introduce different amounts of interference to the system. These factors are especially important in 3G cellular networks that support multimedia services.

Holma and Toskala [2], Viterbi and Viterbi [12], and Liu and El Zarki [18] have studied the effect of interference increase for traffics with the service rate R_i and target SIR requirement $\epsilon_i \equiv E_b/N_0$ for user i. Here E_b is the energy per user bit, and N_0 is the (background) noise plus interference spectral density. ϵ_i is determined by the QoS requirement such as the bit error rate (BER) for a specific media type.

In this paper, we consider the resource management for uplink of W-CDMA system with chip-rate of $W = 3.84$ Mcps. The target ϵ_i is specified for a user i carrying traffic with data rate R_i. The processing gain of user i is denoted as G_i, which can be written as $G_i = W/R_i$. We then define P_N and S_i at BS, as the background noise power and the received power from user i, respectively. Under these conditions, the target ϵ_i can be written as

$$\epsilon_i \equiv \left(\frac{E_b}{N_0}\right)_i = \frac{G_i \cdot S_i}{\sum\limits_{j=1, j\neq i}^{N} S_j + P_N}. \tag{1}$$

Let I_{total} denote the total received power at BS from N active users in the cell, i.e.,

$$I_{\text{total}} = \sum_{i=1}^{N} S_i + P_N. \tag{2}$$

Here, the interference only from the own cell is considered. If inter-cell interference effect is considered, the "other cell interference factor" f has to be taken into account. The standard value of f is 0.55 in IS-95 as described in [2,22]. I_{total} is the maximum planned power, which can be determined by the maximum planned load as well. The value of I_{total} is restricted to be smaller than the upper-bound I_{Th}, otherwise the system becomes unstable and overall interference increases dramatically.

Rewriting Eq. (1) by plugging the value of I_{total}, we get

$$S_i = \left(1 + \frac{G_i}{\epsilon_i}\right)^{-1} \cdot I_{\text{total}} \equiv \rho_i \cdot I_{\text{total}}, \tag{3}$$

where

$$\rho_i = \left(1 + \frac{G_i}{\epsilon_i}\right)^{-1} = \frac{\epsilon_i}{\epsilon_i + G_i} \tag{4}$$

is called the load factor for user i [2,21].

The total system load factor ρ is defined as the sum of load factors from N active mobile users, i.e., $\rho = \sum_{i=1}^{N} \rho_i$.

From Eqs. (1) and (2), we can write

$$I_{\text{total}} - P_N = \sum_{i=1}^{N} S_i = \sum_{i=1}^{N} \rho_i \cdot I_{\text{total}} = \rho \cdot I_{\text{total}}. \tag{5}$$

Holma et al. [2,21], and Shapira and Padovani [23] estimated the interference increase by taking into account the load curve as shown in Fig. 1. They further defined the noise-rise, η, as the ratio of I_{total} to background noise P_N.

The noise-rise η in Fig. 1 can be written as

$$\eta \equiv \frac{I_{\text{total}}}{P_N} = \frac{\sum_{i=1}^{N} S_i + P_N}{P_N} = (1-\rho)^{-1}. \tag{6}$$

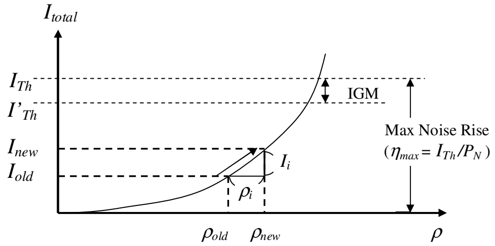

Fig. 1. The load curve and the load estimation.

By taking the partial derivative of I_{total} with respect to ρ, the relationship between I_i and ρ_i can be derived as

$$I_i = \frac{\rho_i}{1-\rho} \cdot I_{\text{total}} = \frac{(1 + G_i/\epsilon_i)^{-1}}{1-\rho} \cdot I_{\text{total}} = \frac{(1 + W/(\epsilon_i \cdot R_i))^{-1}}{1-\rho} \cdot I_{\text{total}}. \tag{7}$$

This equation implies that when a mobile user i is admitted into the cell, the total interference increases by the amount of I_i, which can be expressed in term of its data rate R_i and target ϵ_i.

3. Proposed IGM scheme

In this section, we discuss an efficient radio resource management scheme based on the concept of IGM to provide preferential treatment to higher priority calls. Some parts of this scheme were briefly presented in [24].

3.1. Service model

In a mobile communication system with N active mobile users, we describe the ith ($i < N$) user's traffic profile, as

$$\Im(i) = \{r_i, (R_{\max}, R_{\min})_i, \Pi_i, M_i\}, \tag{8}$$

where r_i, $(R_{\max}, R_{\min})_i$, Π_i and M_i in $\Im(i)$, denote ith user's rate adaptivity, service rate range, priority and mobility, respectively. The proposed service model is designed to take advantage of modern coding schemes and advanced mobile communication technologies as described below.

First, r_i is a binary indicator that indicates whether the user can be serviced at reduced bit-rates when the system is congested. To maintain a specified QoS level, a wireless system should adapt to varying traffic conditions.

Our proposed CAC scheme can achieve this goal, by exploiting the rate adaptive features (please see next paragraph) of modern multimedia coding schemes.

Second, the service rate range $(R_{\max},R_{\min})_i$ describes the target bandwidth consumption. If the network has enough resources, the request can be admitted at $R_{\max}$. If the cellular system is overloaded (congested), a rate-adaptive user can be serviced at a lower rate (down to $R_{\max,i}/2$ or even $R_{\min}$) with degraded QoS. Adaptation takes place only at the time of admitting new calls or at handoff epochs. *Third*, the priority tag Π_i helps the system to identify high priority users, who are likely to receive better QoS guarantees. *Finally*, three mobility types, M_i, are considered in our service model (high, moderate and low mobility). The speed for high, moderate and low mobility traffic are 1, 2 and 4 unit speeds. For traffic of each mobility type, we use a different weighting factor to estimate the amount of resources necessary to be reserved. This is discussed in our proposed resource reservation estimator in Section 3.3.

Modern coding schemes such as MPEG-2 [25] , MPEG-4 [26] and JPEG-2000 [27], have rate adaptive abilities for data communications. In MPEG-2 video/audio compression standard [25], different layers and profiles are defined to achieve target SNR and the spatial scalability. The base layer (with lowest bit-rates) consists of critical information for decoding the sequence at its lowest visual quality. Additional layers provide increasingly better quality. Applications using this kind of codes can adapt to available network resources by transmitting bit streams coded at different layers. Similarly, MPEG-4 [26], which is a new generation multimedia coding standard, has the fine-granular scalability (FGS) mode. Another promising approach for adaptation is the use of embedded coding schemes, such as the wavelet-based JPEG2000 image coding standard [27]. Instead of a few discrete coding rates provided by a layered coding scheme, continuous bit rates can be achieved in JPEG2000 by cutting a single coded bit stream at almost any bit. Better quality can be obtained by transmitting more bits in the bit stream.

3.2. Preferential treatment in IGM scheme

IGM is a natural extension of the GC concept developed in the context of TDMA/FDMA systems, and considers the load factor for system capacity estimation in CDMA systems. As illustrated in Fig. 1, it uses the following two operations. *First*, the load curve is used to estimate the load increase as well as the interference increase. *Second*, a certain amount of resources (in terms of IGM), instead of guard channels, is reserved for high priority calls. The IGM is dynamically adjusted by the RRE function.

For an incoming call j to be admitted, it should satisfy two constraints: (1) The total interference level after admitting this call should not exceed the upper bound of the interference (i.e., the threshold I_{th}) that the system can tolerate. (2) In order to provide preferential treatment for those potential calls whose priority is higher than that of call j, $IGM(j)$ is reserved. Therefore, call j should comply with an augmented constraint $I'_{\text{th}} = I_{\text{th}} - IGM(j)$.

Thus the margin between I_{th} and I'_{th} is exactly the *IGM*. The value of $IGM(j)$ will be defined in Eq. (9). Please note that, the proposed system reserves the resource in terms of $IGM(j)$ for other potential higher priority calls by limiting call j's access up to the interference level of I'_{th}. If call j is new call, the $IGM(j)$ represents IGM_{new}. Similarly, $IGM(j)$ represents IGM_{handoff}, if call j is a handoff call. To put it more precisely, if call j is a new call, it will be admitted into the cell only when the resulting net interference after admitting it is less than $I'_{\text{th}} = I_{\text{th}} - IGM_{\text{new}}$. Similarly, if call j is a handoff call, it will be admitted into the cell only when the resulting net interference after admitting this call is less than $I'_{\text{th}} = I_{\text{th}} - IGM_{\text{handoff}}$.

3.3. Dynamic resource-reservation estimation

When a mobile terminal moves away from its BS toward cell boundary, some of the neighboring BS will receive a stronger signal from it. The mobile terminal will likely handoff to one of these cells. In our simulation system, the information about these neighboring cells is recorded in a handoff candidate registration (HCR) table, that is assumed to be maintained at the corresponding mobile switching center (MSC).

The HCR table provides useful information for estimating the mobile terminals that are likely to handoff to a given cell from its neighboring cells. We use this information to estimate the amount of resource, in terms of interference margin *IGM*(*j*), needed to be reserved when admitting a low priority call *j* in the cell. Call *j* will be admitted into the cell only when resulting net interference of the system is less than $I_{\mathrm{Th}}-IGM(j)$ after admitting this call.

The IGM is estimated based on the weighted sum of minimum interference-increments, $I_{\min,i}$, according to the traffic profile, for each potential handoff call from neighboring cells as

$$\begin{aligned} IGM(j) &= \alpha \cdot \sum_{i\in S(j)} \omega_i \cdot I_{\min,i} \\ &= \alpha \cdot \sum_{i\in S(j)} \omega_i \cdot \left(\frac{\rho_{\min,i}}{1-\rho} \cdot I_{\mathrm{total}}\right) \\ &= \alpha \cdot \sum_{i\in S(j)} \omega_i \cdot \frac{(1+W/(\epsilon_i \cdot R_{\min,i}))^{-1}}{1-\rho} \cdot I_{\mathrm{total}}, \end{aligned} \tag{9}$$

where α, $0 \leqslant \alpha \leqslant 1$, is an empirical scaling factor that takes into account the fact that either some calls from neighboring cells which are likely to handoff in the current cell do not arrive (i.e., terminate or handoff to other cells) or ongoing calls in the current cell terminate (or handoff to other cells).

In Eq. (9), *S*(*j*) is a set consisting of all neighboring active calls who are moving toward current cell and whose priority is higher than the current call request. An example is given later in this section, to illustrate how set *S*(*j*) is determined.

Furthermore, the weighting factor ω_i for user *i* in (9) is proportional to the ratio of its mobility (i.e., speed) M_i to its distance d_i from BS, i.e.,

$$\omega_i \propto (M_i/d_i) \equiv T_i^{-1},$$

where T_i is represented in units of time. The factor ω_i implies that a high speed mobile user at a certain distance from the current BS is more likely to handoff into the current cell, as compared to a low speed mobile user, at the same distance. Let us define

$$\omega_i = \begin{cases} T_{\mathrm{Th}}/T_i, & \text{if } T_i > T_{\mathrm{Th}}, \\ 1, & \text{if } T_i < T_{\mathrm{Th}}, \end{cases} \tag{10}$$

where the threshold, T_{Th}, is an empirical value (in time unit) determined by a critical distance d_{Th} of a call from the target cell boundary with a typical mobile speed of 30 miles/h. For example, if we use a value of $d_{\mathrm{Th}} = 400$ meter and a mobile speed of 30 miles/h (or 800 m/min), the value of T_{Th} is approximately 0.5 min. Therefore, any mobile user whose estimated arrival time T_i is less than 0.5 min is very likely to handoff into the current cell and we use $\omega_i = 1$, to reserve minimum resources requested by it. For T_i greater than T_{Th}, we reserve resources partially. For another example, if T_i is estimated as 1.0 min, we use $\omega_i = 0.5$ and reserve one half of the requested minimum resources.

The distance d_i for user *i* is measured based on the received signal power at the target BS. The radio propagation model is designed based on the analysis given in [28]. It is however difficult to measure distance d_i correctly based on the radio propagation model alone, due to varying effects of shadowing, fading, etc. The Global Positioning System (GPS) technology is becoming popular at a fast pace. The recent E-911 ruling issued by Federal Communications Commission (FCC) in USA requires that the cellular operators must be able to accurately locate mobile callers requesting emergency services via 911 call. We can thus assume that cellular systems will soon be equipped with the technology that will accurately measure the distance d_i between the BS and mobile terminal. When the positioning technology becomes more advanced, accurate estimation of user velocity will also become possible. Chiu and Bassiouni [29] have recently proposed a scheme that predicts the handoff requests based on mobile positioning.

Next, we describe an efficient way to determine set *S*(*j*) in Eq. (9). This set consists of all neighboring active calls that meet two criteria. *First*, the priority $\Pi(i)$ of call *i* is higher than that of the incoming call *j* denoted by $\Pi(j)$. *Second*, the target cell $\Lambda^*(j)$ of call j is in the set of the handoff candidate cells $\Lambda(i)$ of call i in the HCR table. Here, the target cell $\Lambda^*(j)$ is defined as the neighboring BS with maximum received signal power among the handoff candidate BS for call j. Note also that the current cell of call i is not the same as the target cell of incoming call j. To conclude, we have

$$S(j) = \{i \mid \Pi(i) > \Pi(j), \Lambda^*(j) \in \Lambda(i)\}. \quad (11)$$

Figs. 2(a) and (b) illustrate criteria $\Pi(i) > \Pi(j)$ and $\Lambda^*(j) \in \Lambda(i)$ for set $S(j)$, respectively. In Fig. 2(a), we show an incoming call j that requests a handoff from its current base station BS_3 toward its target cell BS_0. We will find out neighboring calls wrt BS_0 whose priority is higher than that of call j. From this operation, we get $\{i|\Pi(i) > \Pi(j)\} = \{i_2,i_3,i_5\}$. In Fig. 2(b), the target cell BS_0 is chosen as the best candidate cell from j's HCR table. Let us denote cell BS_0 by $\Lambda(j)^*$. In this figure, we see several dotted lines associated to each call i, which represent handoff candidate cells. For example, i_2, i_3 and i_5 calls have $\{BS_5\}$, $\{BS_0,BS_6\}$ and $\{BS_0,BS_2\}$, respectively, as their candidate cells. Here, only calls i_3 and i_5 satisfy the condition $\{\Lambda^*(j) \in \Lambda(i)\}$. From the results of (a) and (b) in Eq. (11), we get $S(j) = \{i_3,i_5\}$.

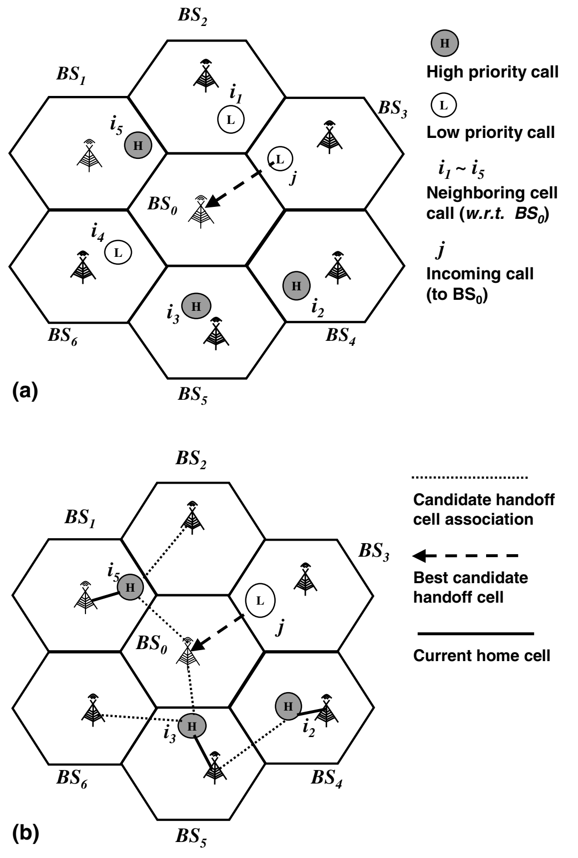

Fig. 2. Set $S(j)$ in resource-reservation estimation: (a) $\Pi(i) > \Pi(j)$ and (b) $\Lambda(j)^* \in \Lambda(i)$.

3.4. Call admission control algorithm

The pseudocode of the CAC algorithm for a media type with three scalable rates is given in Fig. 3, where a new call or a handoff call can be admitted into the system with three data rates: $R_{\max}$, R_{half} and $R_{\min}$. It can be generalized to a media type consisting of even more rates. Note that IGM_{new} and IGM_{handoff} are the estimated bandwidths required to be reserved for new and handoff calls, respectively.

The basic concept behind CAC is to test whether there is enough system resource left to serve the current call request at a certain rate, after reserving the necessary resource for preferential

```
If INCOMING CALLS ARE NEW CALLS
  If CALLS ARE NON-RATE ADAPTIVE
    If (I_current + ΔI_i) < (I_Th − IGM_new)
        ADMIT CALL REQUEST WITH RATE R_i
    Else
        REJECT CALL REQUEST
  Else /*CALLS ARE RATE ADAPTIVE*/
    If (I_current + ΔI_max,i) < (I_Th − IGM_new)
        ADMIT CALL REQUEST WITH RATE R_max,i
    Else If(I_current + ΔI_half,i) < (I_Th − IGM_new)
        ADMIT CALL REQUEST WITH RATE R_half,i
    Else If(I_current + ΔI_min,i) < (I_Th − IGM_new)
        ADMIT CALL REQUEST WITH RATE R_min,i
    Else
        REJECT CALL REQUEST
Else /*INCOMING CALLS ARE HANDOFF CALLS*/
  If CALLS ARE NON-RATE ADAPTIVE
    If (I_current + ΔI_i) < (I_Th − IGM_handoff)
        ADMIT CALL REQUEST WITH RATE R_i
    Else
        REJECT CALL REQUEST
  Else /*CALLS ARE RATE ADAPTIVE*/
    If (I_current + ΔI_max,i) < (I_Th − IGM_handoff)
        ADMIT CALL REQUEST WITH RATE R_max,i
    Else If(I_current + ΔI_half,i) < (I_Th − IGM_handoff)
        ADMIT CALL REQUEST WITH RATE R_half,i
    Else If(I_current + ΔI_min,i) < (I_Th − IGM_handoff)
        ADMIT CALL REQUEST WITH RATE R_min,i
    Else
        REJECT CALL REQUEST
```

Fig. 3. The proposed call admission control algorithm.

treatment to higher priority calls. The CAC test is performed according to the following steps:

Step 1. *New or handoff call test*. An incoming call is first identified as a new or a handoff call type to decide its priority.

Step 2. *Rate adaptivity test*. The rate adaptivity of a new call (handoff call) is tested to decide whether it can be serviced at a lower data rate if the system is congested.

Step 3. *Non-rate adaptive call test*. If the call is rate-adaptive, go to *Step 4*. Otherwise, test whether the amount of interference after admitting the current call and reserving the estimated IGM will exceed the maximum interference level I_{th} that the system can tolerate.

Step 4. *Rate adaptive call test*: If the call is rate-adaptive, the current call could be serviced at rates of $R_{\max,i}$, $R_{\text{half},i}$ and $R_{\min,i}$, depending on the system traffic condition. The amount of interferences introduced by a call are $\triangle I_{\max,i}$, $\triangle I_{\text{half},i}$ and $\triangle I_{\min,i}$ when it is serviced at rates $R_{\max,i}$, $R_{\text{half},i}$ and $R_{\min,i}$, respectively. Then, we test admission criteria by the order of data rates from the highest to the lowest. The call is served at its highest admissible rate.

4. Simulation results

4.1. System model and link characteristics

Simulations were conducted by using the OPtimized Network Engineering Tool (OPNET) [30]. The link characteristics of the CDMA system used in simulation are given below. A network topology with seven cells is used. Normally, the noise rise would be 2–4 [2]. When system is heavily loaded, e.g., ρ close to 0.9, the noise could rise up to 10 as shown in [2]. An empirical value of the maximum interference level I_{Th} is thus set to ten times of background noise, e.g., $\eta_{\max} = 10$ for simulation purpose.

The same radio frequency band is reused for every cell, and separate frequency bands are used for reverse and forward links. There are 420 mobile terminals with three types of mobility (equally distributed). Each cell has its own home BS. Several neighboring BSs together are connected with a centralized center such as the MSC (or RNC) via a wired link. There are a number of mobile users with their own traffic profiles in each cell, which can move across two or more cells according to their predetermined trajectories. Along its trajectory, a mobile user can originate connection requests randomly at its call generation rate.

The Poisson call arrival rate and the exponentially distributed call holding time are assumed. They are controlled by two parameters, i.e., λ (mean request arrival rate measured in the number of connections per hour) and l (mean call holding time of each flow in minutes, which is set to 15 for each call connection). Increasing the value of λ increases network traffic load.

Values used in the traffic profile $\Im(i)$ of user i are listed below:

1. $r_i \in \{YES, NO\}$.
2. $R_{\max,i}$ are, respectively, set to 19.2 Kbps, 38.4 Kbps and 76.8 Kbps for voice, audio and video traffics and $R_{\min,i}$ is set to be $R_{\max,i}/2$.
3. $\Pi \in \{new, handoff\}$.
4. $M_i \in \{HIGH, MODERATE, LOW\}$.
5. Communication system parameters used in simulation include: CDMA chip rate W = 3.84 Mcps and target SIR ε_i = 7 dB.

We use a weighted cost function J, new call blocking probability (P_n), handoff dropping probability (P_h) and system utilization (U) for performance comparison, where

$$J = w_n \cdot P_n + w_h \cdot P_h = w_n \cdot \frac{N_{\text{new\_block}}}{N_{\text{new\_request}}} + w_h \cdot \frac{N_{\text{handoff\_block}}}{N_{\text{handoff\_request}}}. \quad (12)$$

Here $N_{\text{new\_block}}$ is the total number of new calls blocked, $N_{\text{new\_request}}$ the total number of new calls requested, $N_{\text{handoff\_block}}$ the total number of handoff calls blocked, and $N_{\text{handoff\_request}}$ the total number of handoff calls requested. Given a new-call blocking weighting factor $\omega_n = 1$, we used the handoff-call dropping weighting factor $\omega_h = 10$, to reflect the higher cost for dropping a handoff call.

To illustrate the advantage of dynamic IGM, we compare its QoS performance with non-priority scheme (also referred to as the complete sharing scheme) and fixed IGM 20% scheme (i.e., IGM is fixed to 20% I_{Th}).

4.2. Non-rate adaptive traffic

Fig. 4(a)–(c) shows the QoS performance under light to heavy traffic load with λ varying from 0.1 to 1.3 (calls per hour per user). Fig. 4(a) shows that dynamic IGM scheme has the best performance in terms of J. Fig. 4(b) and (c) shows that the dynamic IGM scheme significantly reduces P_h without much increase in P_n as compared to the non-priority scheme. We have used scaling factor $\alpha = 1$ to maximize the value of J. Lower value of J will reduce the reserved bandwidth, resulting in relatively lower P_n and higher P_h.

4.3. Rate adaptive traffic

The rate adaptive traffic can be admitted into a congested system with a lower data rate. Fig. 5(a) and (b) shows the system performance under light to heavy traffic load with λ varying from 0.1 to 1.3. The performance comparison in terms of J is given in Fig. 5(a). Results show that non-priority scheme is better than 20% fixed IGM, for light and moderate traffic conditions. This is because a fixed scheme can not adapt to traffic conditions and possibly reserves excessive resources, which leads to higher P_n.

However, dynamic IGM scheme can adapt well to varying traffic load, and has the best performance in term of J. The P_n and P_h are shown in Figs. 5(b) and (c). Again, dynamic IGM scheme has the best performance.

We compare the cost function J for non-rate adaptive as well as rate adaptive cases in Fig. 6. It is clear that the proposed dynamic IGM scheme outperforms the non-priority scheme for both the

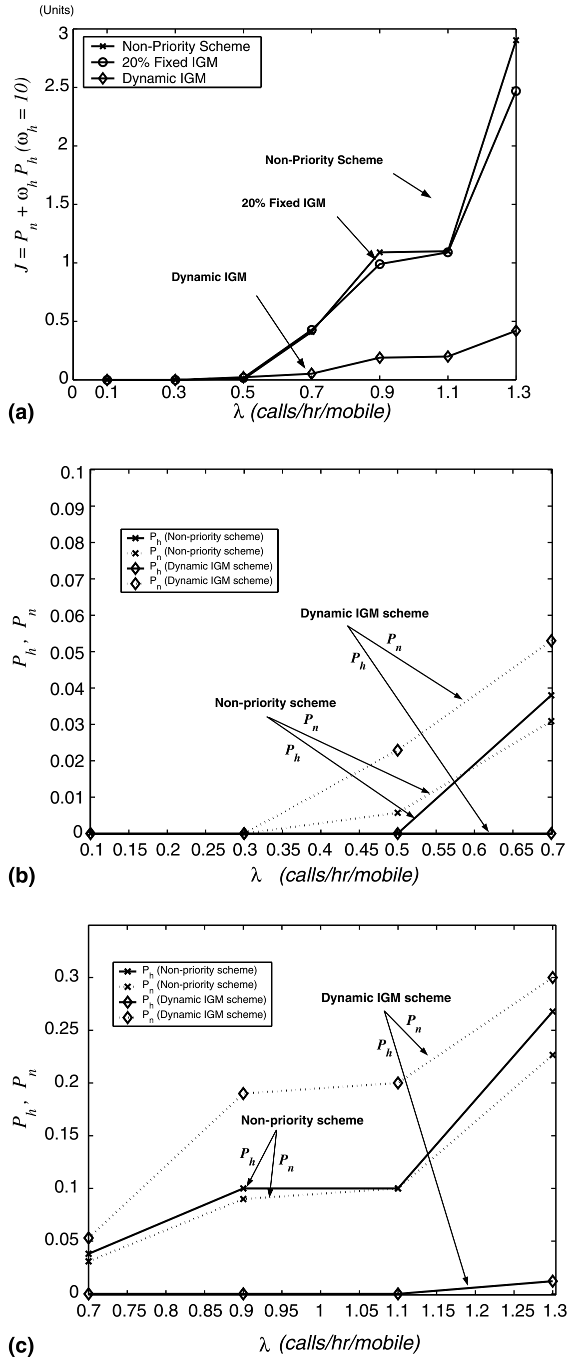

Fig. 4. Performance comparison for non-rate adaptive users under light to heavy traffic load: (a) the cost function J, (b) the new call blocking rate P_n and the handoff dropping rate P_h under light to moderate traffic, and (c) the P_n and P_h under moderate to heavy traffic.

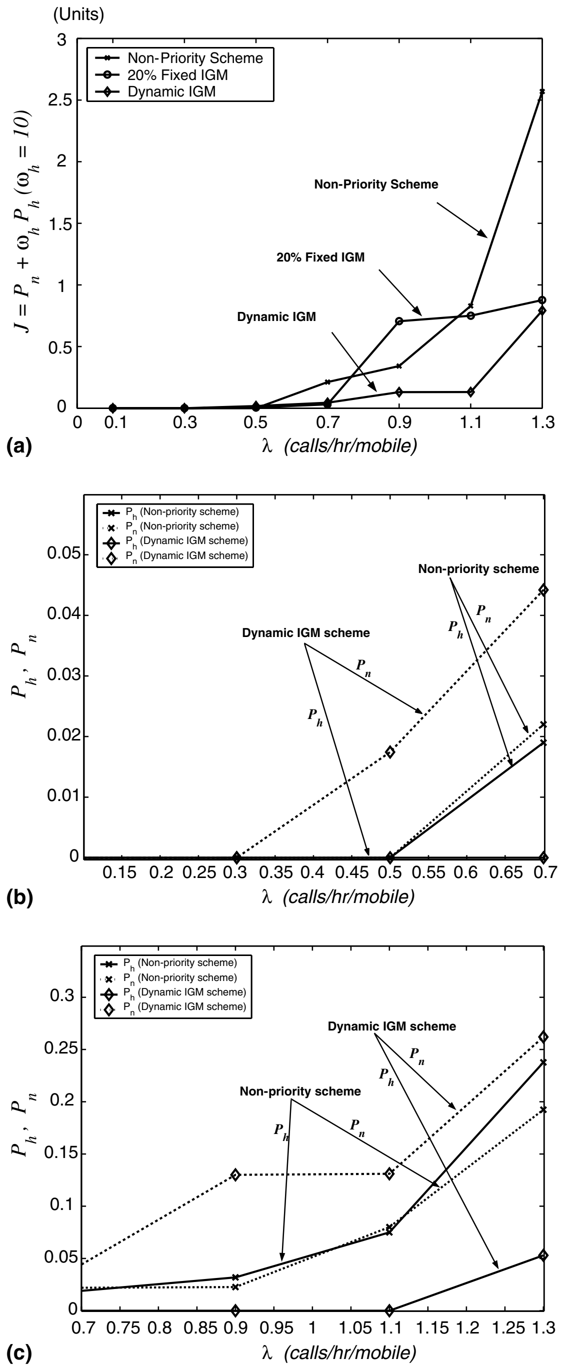

Fig. 5. Performance comparison for rate adaptive users under light to heavy traffic load: (a) the cost function J, (b) the new call blocking rate P_n and the handoff dropping rate P_h under light to moderate traffic, and (c) the P_n and P_h under moderate to heavy traffic.

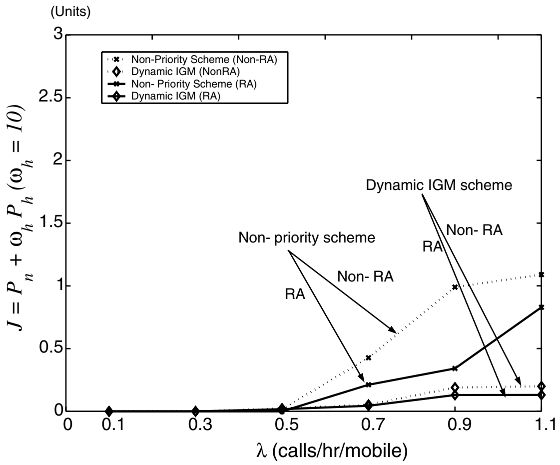

Fig. 6. Performance comparison between rate adaptive (RA) and non-rate adaptive (Non-RA) schemes for the cost function J.

cases. Furthermore, the dynamic IGM scheme achieves more improvement when rate adaptive mechanism is used.

4.4. System utilization

Let system utilization U be defined as a ratio of the time average of occupied bandwidth to total bandwidth, i.e., $U = \mathbf{E}[BW_{\text{occu}}/BW_{\text{total}}]$, where BW_{occu} and BW_{total} are instantaneous occupied and total system bandwidth, respectively. $\mathbf{E}[\cdot]$ denotes time average over 5 h simulation time.

System utilizations for all four schemes with respect to non-rate adaptive and rate-adaptive cases is compared in Fig. 7 under moderate and heavy traffic load. We see that the system utilization is higher for traffic with rate-adaptive capability than that without rate-adaptive capability. This can be explained by the fact that the congested system can support calls at a reduced data rate, thus increasing the overall system utilization.

Due to resource reservation adopted by the proposed dynamic IGM scheme (to provide preferential treatment to higher priority calls), it cannot fully utilize the system resources. The use of scaling factor $\alpha = 0.7$ in IGM increases the system utilization for non-rate adaptive traffic at the expense of dropping relatively more handoff calls.

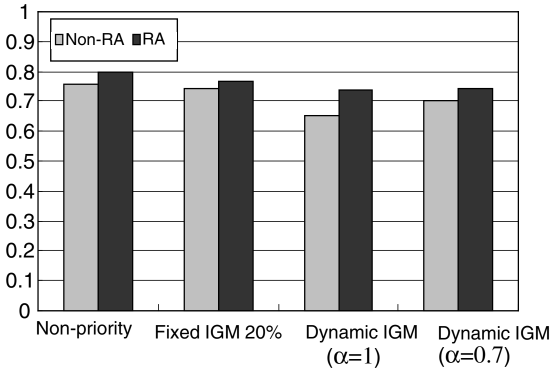

Fig. 7. Comparison of system utilization wrt rate-adaptive and non-rate-adaptive traffic.

As shown in the figure, non-priority scheme has the best system utilization performance since it does not reserve any resources for handoff calls, and accepts the calls on the first-come first-serve basis. The fixed IGM scheme gives higher system utilization as compared to dynamic IGM scheme because the latter reserves more resources to serve handoff calls under heavy load condition.

Finally, we would like to point out that, under light traffic condition, system utilization is about the same for non-priority scheme, fixed IGM and dynamic IGM schemes.

5. Conclusion and future work

Effective radio resource management schemes, including dynamic RRE and CAC, based on the concept of IGM for CDMA systems were presented. We considered a service model that included mobile terminals' service rate, their different priority levels, rate adaptivity as well as their mobility. The proposed dynamic IGM scheme reserves a certain amount of interference margin for high priority calls by referencing the traffic condition and mobile users' traffic profile in neighboring cells. The mobility-aware weighted sum plays an important role in the RRE process so that the effect of different mobility is taken into consideration. It was shown by computer simulation that the proposed dynamic IGM scheme outperforms the fixed IGM and non-priority schemes in achieving a smaller cost function J under light as well as heavy traffic conditions. It is worthwhile to point out that the cost function J was primarily introduced for performance comparison. It should be interesting to study (in future research) how J can be used to fine-tune parameters in the algorithm, e.g., α, to achieve the optimal value of J.

MSC in CDMA-based cellular systems (e.g., IS-95) monitors the received signal strength from each user at several neighboring base stations for implementing the handoff mechanism. We assume that this information is stored in HCR tables. This process does not contribute additional overhead to our scheme as it is already being implemented by the cellular system. The extra overhead in our scheme comes from the following: (i) these tables need to be scanned for the users who are likely to handoff in a given cell; (ii) IGM is computed using Eq. (9) while accepting a new call. The number of users that belong to $S(j)$ will be only a small fraction of the calls active in the neighboring cells because $S(j)$ includes only those ongoing calls (in the neighboring cells) that are likely to handoff to the current cell. Since the computational power of the processor-controlled systems is rapidly increasing and MSC is a fairly powerful system, we feel that the computation of IGM in Eq. (9) should not introduce excessive computational overhead even for users moving at a high speed. Thus, the whole operation should be manageable in real time.

In modern CDMA systems, soft handoff is employed to provide a better transition process than hard handoff. The benefit of soft-handoff is that a mobile user can connect to two or more base stations at the same time, thus greatly reducing the probability of call-dropping due to severe channel impairments [31–34]. However, soft-handoff should be used only up to a certain extent because an excessive amount of soft-handoff connections increases the down-link interference. It should be interesting to generalize the proposed IGM scheme to CDMA systems with hybrid hard- and soft-handoff schemes to achieve an efficient radio resource management mechanism.

Finally, the performance of the IGM scheme was mainly verified by simulation modeling in this work due to the complex and dynamic environments that involve many variables such as the presence of users with multi-class traffic parameters (including adaptive data rates, mobility, priority, distance from the handoff cell), the cost function, the blocking probabilities, and their relative weights. It may be challenging but worthwhile to perform mathematical analysis under some simplifying conditions to shed light on the performance of the IGM scheme.

References

[1] J. Zander, S.-L. Kim, M. Almgren, O. Queseth, Radio Resource Management for Wireless Networks, Artech House, Norwood, MA, 2001.

[2] H. Holma, A. Toskala, WCDMA for UMTS—Radio Access for Third Generation Mobile Communications, Wiley, New York, 2000 (Fig. 2 in Chapter 8).

[3] D. Hong, S.S. Rappaport, Traffic model and performance analysis for cellular mobile radio telephone systems with prioritized and nonprioritized handoff procedures, IEEE Trans. Veh. Technol. 35 (1996) 77–92.

[4] H. Chen, S. Kumar, C.-C. J. Kuo, Differentiated QoS aware priority handoff in cell-based multimedia wireless network, in: Proceedings of the IS&T/SPIE's 12th International Symposium, Electronic Imaging, San Jose, CA, January 2000, Vol. 3974, pp. 940–948.

[5] M. Naghshineh, M. Schwartz, Distributed call admission control in mobile/wireless networks, IEEE J. Sel. Areas in Commun. 14 (1996) 711–717.

[6] P. Ramanathan, K.M. Sivalingam, P. Agrawal, S. Kishore, Dynamic resource allocation schemes during handoff for mobile multimedia wireless networks, IEEE J. Sel. Areas in Commun. 17 (1999) 1270–1283.

[7] A.S. Acampora, M. Naghshineh, Control and quality of service provisioning in high-speed micro-cellular networks, IEEE Personal Commun. 1 (2) (1994) 36–43.

[8] A. Sutivong, J.M. Peha, Novel heuristics for call admission control in cellular systems, in: Proceedings of the IEEE 6th International Conference Universal Personal Communications Record, 1, 1997, pp. 129–1331.

[9] B.M. Epstein, M. Schwartz, Predictive QoS-based admission control for multiclass traffic in cellular wireless networks, IEEE J. Sel. Area in Commun. 18 (3) (2000) 523–534.

[10] L. Huang, S. Kumar, C.-C.J. Kuo, Adaptive resource allocation for multimedia QoS management in wireless networks, IEEE Trans. Veh. Technol. 53 (2) (2004) 547–558.

[11] I. Katzela, M. Naghshineh, Channel assignment schemes for cellular mobile telecommuication systems: a comprehensive survey, IEEE Personal Commun. 3 (3) (1996) 10–31.

[12] A.M. Viterbi, A.J. Viterbi, Erlang capacity of a power controlled CDMA system, IEEE J. Sel. Areas in Commun. 11 (1993) 892–900.

[13] W. Park, Y.H. Kwon, D.C. Lee, A CAC scheme with code and interference limits on the forward link in CDMA cellular network, in: Proceedings of IEEE MILCOM, Vol. 2, 2001, pp. 1461–1466.

[14] J. Knutsson, P. Butovitsch, M. Persson, R.D. Yates, Downlink admission control strategies for CDMA systems in a Manhattan environment, in: Proceedings of the IEEE VTC, Vol. 2, May 1998, pp. 1453–1457.

[15] C.Y. Huang, R.D. Yates, Call admission in power controlled CDMA systems, in: Proceedings of IEEE VTC, Vol. 3, 1996, pp. 1665–1669.

[16] N. Dimitriou, R. Tafazolli, Resource management issues for UMTS, in: Proceedings of the 1st IEEE International Conference of 3G Mobile Communications and Technology, March 2000, pp. 401–405.

[17] S.A. Grandhi, R.D. Yates, D.J. Goodman, Resource allocation for cellular radio systems, IEEE Trans. Veh. Technol. 46 (3) (1997) 581–587.

[18] Z. Liu, M. El Zarki, SIR-based call admission control for DS-CDMA cellular systems, IEEE J. Sel. Areas Commun. 12 (1994) 638–644.

[19] K.S. Gilhousen, I.M. Jacobs, R. Padovani, A.J. Viterbi, L.A. Weaver Jr., C.E. Wheatley III, On the capacity of a cellular CDMA system, IEEE Trans. Veh. Technol. 40 (1991) 303–312.

[20] S.M. Shin, C.-H. Cho, D.K. Sung, Interference-based channel assignment for DS-CDMA cellular systems, IEEE Trans. Veh. Technol. 48 (1999) 233–239.

[21] H. Holma, J. Laakso, Uplink admission control and soft capacity with MUD in CDMA, in: Proceedings of the IEEE VTC, Amsterdam, Netherlands, September 1999, Vol. 1, pp. 431–435.

[22] Vijay K. Garg, IS-95 CDMA and CDMA2000, Prentice-Hall PTR, Englewood Cliffs, NJ, 2000.

[23] J. Shapira, R. Padovani, Microcell engineering in CDMA cellular networks, IEEE Trans. Veh. Technol. 43 (1994) 213–216.

[24] H. Chen, S. Kumar, C.-C.J. Kuo, Dynamic call admission control and resource reservation with interference guard margin (IGM) for CDMA systems, in: Proceedings of the IEEE Wireless Communications and Networking Conference, New Orleans, LA, March 2003.

[25] ISO/IEC JTC1/SC29/WG11, Generic coding of moving pictures and associated audio information, ISO/IEC International Standard 13818, November 1994.

[26] ISO/IEC JTC1/SC29/WG11, Overview of the MPEG-4 standard, ISO/IEC N3747, October, 2000.

[27] ISO/IEC JTC1/SC29/WG1, JPEG2000 Part 1: final committee draft Ver. 1.0, ISO/IEC International Standard N1646R, March 2000.

[28] W.C.Y. Lee, Mobile Cellular Telecommunications—Analog and Digital Systems, second ed., McGraw-Hill, New York, 2000.

[29] M.-H. Chiu, M.A. Bassiouni, Predictive schemes for handoff prioritization in cellular networks based on mobile positioning, IEEE J. Sel. Areas Commun. 18 (2000) 510–522.
[30] I. Karzela, Modeling and Simulating Communication Networks: A Hands-On Approach Using OPNET, Prentice-Hall, Englewood Cliffs, NJ, 1998.
[31] D. Wong, T.G. Lim, Soft handoff in CDMA systems, IEEE Personal Commun. 4 (6) (1997) 6–17.
[32] A.J. Viterbi, A.M. Viterbi, K.S. Gilhousen, E. Zehavi, Soft handoff extends CDMA cell coverage and increases reverse link capacity, IEEE J. Sel. Areas Commun. 12 (8) (1994) 1281–1288.
[33] N. Zhang, J.M. Holtzman, Analysis of a CDMA soft-handoff algorithm, IEEE Trans. Veh. Technol. 47 (2) (1998) 710–714.
[34] J.W. Chang, J.H. Chung, D.K. Sung, Admisison control scheme for soft handoff in DS-CDMA cellular systems supporting voice and stream-type data services, IEEE Trans. Veh. Technol. 51 (6) (2002) 1445–1459.

Huan Chen received his B.S. and M.S. degrees in Electrical Engineering from National Tsing Hua University (NTHU), Taiwan in 1993 and 1995, respectively. He received his Ph.D. degree in Electrical Engineering from University of Southern California (USC), Los Angeles, in December 2002. Since February 2003, he has been an Assistant Professor with the Electrical Engineering department at National Chung Cheng University (NCCU), Taiwan. His research interests include QoS support, optimal resource management, power management for wireless cellular and ad hoc networks.

Sunil Kumar received B.E. (Electronics Engineering) degree from S.V. National Institute of Technology, Surat (India), in 1988 and the M.E. (Electronics and Control Engineering) and Ph.D. (Electrical and Electronics Engineering) degrees from the Birla Institute of Technology and Science (BITS, Pilani, India) in 1993 and 1997, respectively. He also served as a Lecturer in the Electrical and Electronics Engineering department at BITS from January 1993 to July 1997. From August 1997 to August 2002, he was a postdoctoral researcher and adjunct faculty in Signal and Image Processing Institute, Integrated Media Systems Center and Electrical Engineering department at University of Southern California, Los Angeles. From July 2000 to July 2002, he was also an expert consultant in industry on JPEG2000 and MPEG4-based projects and participated in JPEG2000 standardization activities. Since August 2002, he has been with Electrical and Computer Engineering department at Clarkson University, Potsdam, NY.

He is a Senior Member of IEEE. He is a co-author of more than 55 technical publications in international conferences and journals as well as a book on *Radio Resource Management for Multimedia QoS Support in Wireless Networks* (Kluwer Academic Publishers, 2003). He is Guest Editor of Special Issue of *Journal of Visual Communications and Image Representations* on 'Emerging H.264/AVC Video Coding Standard' to be published during June–October 2005.

His research interests include QoS support for multimedia traffic in wireless cellular, ad hoc and sensor networks, Error resilient multimedia compression techniques, MPEG-4, H.264/AVC and JPEG2000 image/video compression standards.

C.-C. Jay Kuo received the B.S. degree from the National Taiwan University, Taipei, in 1980 and the M.S. and Ph.D. degrees from the Massachusetts Institute of Technology, Cambridge, in 1985 and 1987, respectively, in Electrical Engineering. He is with the Department of Electrical Engineering, the Signal and Image Processing Institute (SIPI) and the Integrated Media Systems Center (IMSC) at the University of Southern California (USC) as Professor of Electrical Engineering and Mathematics. His research interests are in the areas of digital media processing, multimedia compression, communication and networking technologies, and embedded multimedia system design. He is a Fellow of IEEE and SPIE. He received the National Science Foundation Young Investigator Award (NYI) and Presidential Faculty Fellow (PFF) Award in 1992 and 1993, respectively.

He has guided 54 students to their Ph.D. degrees and supervised 15 postdoctoral research fellows. Currently, his research group at USC consists around 30 Ph.D. students and five postdoctors (please visit website http://viola.usc.edu), which is one of the largest academic research groups in multimedia technologies. He is a co-author of more than 600 technical publications in international conferences and journals as well as the following seven books: *Content-based Audio Classification and Retrieval for Audiovisual Data Parsing* (with Tong Zhang, Kluwer Academic Publishers, 2001), *Semantic Video Object Segmentation for Content-based Multimedia Applications* (with Ju Guo, Kluwer Academic Publishers, 2001), *Intelligent Systems for Video Analysis and Access over the Internet* (with Wensheng Zhou, Prentice Hall, 2002), *Video Content Analysis Using Multimodal Information* (with Ying Li, Kluwer Academic Publishers, 2003) *Quality of Service for Internet Multimedia* (with Jitae Shin and Daniel Lee, Prentice Hall, 2003), *Radio Resource Management for Multimedia QoS Support in Wireless Cellular Networks* (with Huan Chen, Lei Huang and Sunil Kumar, Kluwer Academic Publishers, 2003) and *High Fidelity Multichannel Audio Coding* (with Dai Tracy Yang and Chris Kyriakakis, Hindawi Publishing Corporation, 2004).

He is Editor-in-Chief for the *Journal of Visual Communication and Image Representation*, and Editor for *the Journal of Information Science and Engineering* and *the RURASIP Journal of Applied Signal Processing*. He is also on the Editorial Board of the *IEEE Signal Processing Magazine*. He served as Associate Editor for *IEEE Transactions on Image Processing* in 1995–1998, *IEEE Transactions on Circuits and Systems for Video Technology* in 1995–1997 and *IEEE Transactions on Speech and Audio Processing* in 2001–2003.