Available online at www.sciencedirect.com

SCIENCE DIRECT

Computer Communications 28 (2005) 1240–1250

computer communications

www.elsevier.com/locate/comcom

An adaptive handoff priority scheme for wireless MC-CDMA cellular networks supporting realtime multimedia applications

Jinfang Zhang, Jon W. Mark, Xuemin (Sherman) Shen*

Centre for Wireless Communications, Department of Electrical and Computer Engineering, University of Waterloo, 200 University Avenue West, Waterloo, Ont., Canada N2L 3G1

Received 21 July 2004; revised 21 July 2004
Available online 11 September 2004

Abstract

An adaptive handoff priority scheme is proposed for a packet switching multicode (MC)-CDMA cellular network supporting realtime multimedia applications. The scheme jointly considers the physical, link and network layer characteristics, and gives service priority to handoff calls by exploiting the transmission rate adaptability of multimedia calls to the available radio resources. Based on the scheme, a call admission region is derived with the estimated cell capacity and the proposed packet scheduling and partial packet integration scheme and used for call admission control and handoff management with the satisfaction of quality of service requirements for all multimedia traffic, where the QoS parameters include the wireless transmission bit error rate, packet loss rate and delay requirement. Simulation results are given to demonstrate the viability of the proposed scheme.

Keywords: MC-CDMA; Cellular networks; Handoff; Multimedia; Call admission control; Quality of service

1. Introduction

With the increasing demand for broadband multimedia services in wireless cellular networks, packet switching multicode-code division multiple access (MC-CDMA) is considered to be a suitable multiple access technology to support multimedia traffic by assigning multiple spreading codes to each user [1]. Packet switching for wireless transmission makes it possible to achieve a high statistical multiplexing gain for better radio resource utilization. Due to the heterogeneity and bursty nature of multimedia traffic, a flexible packet scheduling scheme, which can efficiently accommodate multimedia traffic, is desired. The order of packet transmission for multimedia traffic has a great impact on the efficiency and performance of wireless systems. The uplink call admission region of an MC-CDMA system is statistically defined as the number of calls of all traffic classes with different transmission and performance requirements that can be supported simultaneously. The call admission region can be enlarged by decreasing the cell size. However, shrinking the cell size will increase handoff rate during the call holding time. In a network supporting multimedia services, the increased handoff rate makes it difficult to guarantee the service requirements in a new cell agreed upon at call setup time. If a handoff is unsuccessfully executed, the connection of an ongoing call is forced to terminate. It is important to maintain service continuity of an ongoing call from a user's perspective. Many schemes have been proposed to prioritize handoff calls over new calls. In channel reservation schemes [2], a fixed amount of the channel capacity is reserved exclusively for handoff calls. Fixed channel reservation has the risk of increasing new call blocking probability. To decrease the probability of handoff call dropping and at the same time to keep the new call blocking probability as low as possible, dynamic channel reservation schemes based on user mobility are developed [3,4]. However, these schemes only consider homogeneous voice traffic. In Refs. [5,6], multimedia services are introduced where resources are reserved

* Corresponding author. Tel.: +1 519 888 4567x2691; fax: +1 519 746 3077.

E-mail addresses: jzhang@bbcr.uwaterloo.ca (J. Zhang), jwmark@bbcr.uwaterloo.ca (J.W. Mark), xshen@bbcr.uwaterloo.ca (X.(S.) Shen).

0140-3664/$ - see front matter © 2004 Elsevier B.V. All rights reserved.
doi:10.1016/j.comcom.2004.08.004

according to the potential handoff calls. However, channel reservation for broadband multimedia traffic has the risk of under-utilizing the radio resources. A dynamic bandwidth sharing scheme is proposed in Ref. [7] where realtime traffic can borrow the necessary bandwidth from existing non-realtime traffic when there is no sufficient available bandwidth in the target cell in handoff. A sub-rating channel assignment strategy for handoffs with homogeneous traffic is proposed by Lin et al. [8] which reduces handoff call dropping probability by temporarily dividing a full-rate channel into two half-rate channels, one to serve the existing call and the other to serve the handoff call. This scheme exploits the ability of mobile users to tolerate some degradation in quality of service (QoS). In fact, this is true for most multimedia services involving voice and video applications where multi-level or hierarchical source/channel coding schemes can be used. Two traffic classes are introduced in sub-rating schemes for handoff [9]. One traffic class is assumed to be QoS adaptive to the available resources, while the other cannot accept QoS degradation. However, a simple sub-rating scheme is not practical for a multimedia CDMA system due to its time-varying call admission region, neither is it suitable for traffic classes with different transmission rates and performance requirements.

In this paper, an adaptive handoff priority scheme is proposed for a packet switching MC-CDMA cellular network supporting multimedia services. The scheme jointly considers the physical, link and network layer characteristics, and gives service priority to handoff calls by exploiting the transmission rate adaptability of multimedia calls to the available radio resources. The overall error rate in the packet switching wireless system comprises both bit error rate (BER) in the physical layer wireless transmission caused by the interfering radio communication environments and link layer packet loss rate (PLR) caused by the failed scheduling of packets with delay exceeding the required delay bound. As the wireline network has much larger capacity compared to the wireless network, we ignore the error introduced in the wireline domain and focus on the last hop transmission, namely the transmission between the base station (BS) and the mobile station (MS). We first derive the uplink cell capacity, which limits the maximum number of packets that can be transmitted simultaneously while still satisfying the prescribed BER requirement. A predefined BER is applied to all traffic classes to guarantee the same performance in wireless transmission. Therefore, the uplink cell capacity derivation alleviates the power control complexity as it is traffic independent and accordingly depends solely on communication environments. For traffic types with different error rate requirements, additional error rate is converted from the link layer PLR due to failed scheduling of packets for transmission. Based on the uplink cell capacity, a packet scheduling and partial packet integration scheme is presented, which achieves different PLR requirements among different traffic types and provides fair packet loss constraint among ongoing calls with the same traffic type. The traffic characteristics at the call level and the packet loss constraint at the packet level are applied to derive the call admission region, which is used by the serving mobile switching center (MSC) to execute call admission control (CAC) and handoff management. An adaptive handoff priority scheme is proposed to adjust the transmission rate of ongoing rate-adaptive calls to the available radio resources to prioritize handoff calls over now calls. The scheme aims to decrease the handoff call dropping probability, and at the same time to maintain a low new call blocking probability. With the CAC and adaptive handoff management, better cell capacity utilization and satisfaction of QoS for each individual call can be achieved.

The novelty of this research is the constitution of a relation between the wireless transmission BER and the PLR in the derivation of call admission region with the proposed packet scheduling and partial packet integration scheme and the adaptive execution of handoff call prioritization. The remainder of the paper is organized as follows. Section 2 presents the system model. The proposed call admission region derivation and adaptive handoff priority scheme are presented in Section 3. Simulation results are given in Section 4, followed by conclusions in Section 5.

2. System model

We consider a Frequency Division Duplex/MC-CDMA (FDD/MC-DMA) system, in which the uplink and downlink use different frequency bands to provide frequency isolation, so there is no interference between the uplink and downlink transmissions. The network structure has two tiers. The first tier is a mesh connection of MSCs, which connect the wireless sub-networks with the backbone network. An MSC collects status information of all its serving BSs and performs CAC and handoff management based on traffic characteristics, the QoS requirements, and the resource availability. In order to effectively support soft handoff, a Selection/Distribution Unit (SDU) residing at the MSC schedules the packet transmission and reception for both downlink and uplink among the associated BSs. The second tier consists of a cluster of BSs, each connected to a serving MSC. A BS plays two roles. First, it takes part in the radio resource management under the control of the serving MSC. Second, it works as the interface between an MS and its serving MSC. An MS keeps an active connection with the serving BS by monitoring the received pilot signal strength (RPSS). An MS also measures the RPSS from all its neighboring BSs. Without loss of generality, we assume that only two BSs are involved in handoff.

In this research, we assume calls are uniformly located in a cell. It is assumed that fast fading has been taken care of by the physical layer functions so that the channel is characterized by path loss and shadowing. Given the location of an MS as (r,θ), the channel gain $G(r,\theta)$ is modeled as

$$G(r,\theta) = d(r,\theta)^{-n}10^{\zeta/10}, \tag{1}$$

where $d(r,\theta)=d_r/d_0$, d_r is the distance between the BS and the MS and d_0 is the reference distance from the BS; n is the path loss exponent; and $10^{\zeta/10}$ is the lognormal shadowing effect with ζ being a Gaussian distributed random variable with mean zero and standard deviation σ.

Suppose there are K classes of traffic with different transmission rates, BER, PLR and delay requirements. Let B bits/s be the basic transmission rate of a single code channel, and a type i call have a transmission rate of c_iB, where c_i is a positive integer representing the number of subcode channels used for class i calls. For each traffic class, the wireless transmission BER is defined by the upper bounded outage rate $(P_{\text{out}})_i$, which is the probability that the BER exceeds the prescribed value.

For a given transmission environment, coding and detection schemes, the BER requirement corresponds to the bit energy-to-interference-density ratio E_b/I_0 requirement. It is assumed that all calls of the same traffic class have the same E_b/I_0 requirements, denoted as $(E_b/I_0)_i$. It is noted that for a class i call requiring c_i subcode channels, each subcode channel will have the same $(E_b/I_0)_i$ requirement. For realtime service, a user requires its packets to be delivered to the destination within a delay bound; otherwise, the packets will be worthless and discarded. Here we only consider the access delay over wireless links, which is denoted as a time-out value. For each session, a buffer is required at the MS side to accommodate bursty information generation. Each arrival packet at the buffer is labeled with a time-out value (delay bound at the first arrival) in frames. This value is decreased by one for every frame interval till the packet is scheduled to transmit. The rate at which packets are discarded because they exceed the delay bound is referred to as the PLR.

Voice and video are two types of realtime traffic requiring immediate resource reallocation during handoff to maintain service continuity. Fig. 1 shows an ON–OFF model with a the probability of transition from OFF (silent) state to ON (talk spurt) state and b from ON state to OFF state. The activity factor, which is defined as the fraction of average time that a call is in the ON state, is given by $a/(a+b)$. When in ON state, information is transmitted at a constant rate of R bits/s to achieve small delay jitter requirement. The number of subcodes needed is set to be $\lceil R/B \rceil$, which usually has the value of one, where $\lceil \cdot \rceil$ is a ceiling function. Video packets are generated at a rate according to their characteristics and the coding scheme used. Hierarchical source coding, which adapts the packet rate to the available resources, is often used for video coding [10]. Two types of video calls are defined according to the packet generation rate, namely high-quality coding video call and low-quality coding video call. Each video call can be approximately represented by a superposition of N ON–OFF minisources as in Fig. 1 with parameters a, b and R different from those of the voice call model. An approach to calculate the parameters a, b and R can be found in Ref. [11]. If the values of a and b, which can be used to decide the activity factor of a minisource, are given, different quality video calls can be represented by different values of R. It is noted that the number of subcodes needed for high- and low-quality video calls are time-varying.

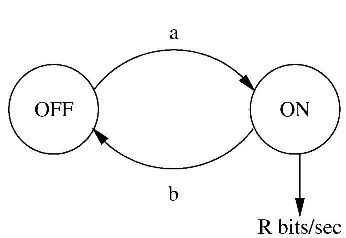

Fig. 1. Two-state ON–OFF model.

3. Adaptive handoff priority scheme

In this section, we first estimate the uplink cell capacity, then present an efficient packet scheduling and partial packet integration scheme for multimedia packet transmission. Based on the uplink cell capacity and the scheduling scheme, a call admission region for MC-CDMA systems supporting voice, high- and low-quality video calls is derived. An adaptive handoff priority strategy is proposed to decrease handoff call dropping probability with satisfaction of new call blocking probability and resource utilization.

3.1. Uplink cell capacity estimation

In a packet switching multimedia wireless network, the overall packet error rate for user i, P_e^i, includes error rate due to wireless transmission, and packet loss

$$P_e^i = P_{\text{PLR}}^i + (1-P_{\text{PLR}}^i)\times P_{\text{WTE}}^i, \tag{2}$$

where P_{PLR}^i is the packet error rate due to packet loss and P_{WTE}^i is the packet error rate due to wireless transmission error. To simplify complicated power control algorithms for different traffic, especially for variable data rate video traffic, all the uplink channels are specified to provide the same P_{WTE}^* value so that the power control algorithm can be traffic independent. It is note that P_{WTE}^* should not be greater than the most stringent overall packet error rate requirement of the traffic in the system. In this way, different error rate requirements for different users can be achieved by providing different PLR with a proper packet scheduling scheme. For user i, the required P_{PLR}^i to satisfy a predefined P_e^i with the specified P_{WTE}^* is expressed as

$$P_{\text{PLR}}^i = \frac{P_e^i - P_{\text{WTE}}^*}{1-P_{\text{WTE}}^*}. \tag{3}$$

We define the BER caused by radio transmission as P^0_{BER}. If the physical layer error detection and correction are taken into consideration, the actual BER seen by link layer is P^*_{BER}. Therefore, the PLR caused by BER at link layer can be expressed as

$$P^*_{\mathrm{WTE}} = 1-(1-P^*_{\mathrm{BER}})^L, \tag{4}$$

where L is the packet length.

Let P^1_{PLR}, $P^{2\mathrm{h}}_{\mathrm{PLR}}$ and $P^{2\mathrm{l}}_{\mathrm{PLR}}$ denote the PLR for voice, high- and low-quality video calls, respectively. A large P_{PLR} value implies a small P_{WTE} requirement and vice versa. It is expected that more efficient resource utilization can be achieved by choosing a less stringent P_{WTE} value because a more stringent P_{WTE} value indicates a small cell capacity and a degradation of multiplexing gain at packet scheduler. Also the noisy communication environment makes a stringent P_{WTE} requirement not appropriate for practical implementation. The effect of different P_{PLR} and P_{WTE} pairs on performance measures will be demonstrated through simulations.

A typical hexagonal cell structure with dashed lines representing inner and outer cells of radii R_{I} and R_{O} is shown in Fig. 2. The radius of the cell is R_{C}. The relative handoff area is defined as $\alpha=(R_{\mathrm{O}}-R_{\mathrm{I}})/R_{\mathrm{C}}$. It has been experimentally verified that imperfectly received power can be modeled as a random variable, S_{e}, with lognormal distribution [12]

$$S_{\mathrm{e}} = S10^{\xi_{\mathrm{e}}/10}, \tag{5}$$

where S is the required received power if perfect power control can be achieved; ξ_{e} is a Gaussian random variable with mean zero and standard deviation σ_{e}. Because of the subcode concatenation [1] in an MC-CDMA transmitter, orthogonality is achieved between the different simultaneously transmitted packets from the same user, so there is no interference between the simultaneously transmitted packets from the same user. For user i, the BS will receive a composite signal including the desired signal power, $S_{\mathrm{e}}(i)$, interference from other users in the same cell, I_{I}, interference from users in the soft handoff area, I_{s}, interference from outer cell users, I_{O} and background noise, η. During a snapshot, the received bit energy-to-interference density ratio, $E_{\mathrm{b}}/I_0(i)$, is given by [13]

$$E_{\mathrm{b}}/I_0(i) = \frac{W/B}{I_{\mathrm{I}}/S_{\mathrm{e}}(i)+I_{\mathrm{S}}/S_{\mathrm{e}}(i)+I_{\mathrm{O}}/S_{\mathrm{e}}(i)+\eta/S_{\mathrm{e}}(i)}, \tag{6}$$

where W is the spread spectrum bandwidth. Let $C=[N_{\mathrm{I}}+P_{\mathrm{R}}N_{\mathrm{S}}]$ denote the maximum number of packets that can be transmitted simultaneously, where N_{I} represents the number of packets from the inner cell users; N_{S} represents the number of packets from soft handoff users; and P_{R} is the probability of an MS in soft handoff and power controlled by the reference BS [13]. C is called the uplink cell capacity with the satisfaction of the specified wireless transmission error requirement P^0_{BER}.

$$\begin{aligned} &P_{\mathrm{r}}\{P_{\mathrm{BER}} \ge P^0_{\mathrm{BER}}\} \\ &= P_{\mathrm{r}}\left\{ I_{\mathrm{I}}/S_{\mathrm{e}} + I_{\mathrm{S}}/S_{\mathrm{e}} + I_{\mathrm{O}}/S_{\mathrm{e}} + \eta/S_{\mathrm{e}} \ge \frac{W/B}{(E_{\mathrm{b}}/I_0)^0} \right\} \\ &= \int_0^\infty \int_0^\infty Q\left(\frac{\frac{W/B}{(E_{\mathrm{b}}/I_0)^0} - \eta/S_{\mathrm{e}} - y - z - E[I_{\mathrm{I}}/S_{\mathrm{e}}]}{\sqrt{\mathrm{Var}(I_{\mathrm{I}}/S_{\mathrm{e}})}} \right) \\ &\quad \times f_{I_{\mathrm{S}}/S_{\mathrm{e}}}(y) f_{I_{\mathrm{O}}/S_{\mathrm{e}}}(z)\,\mathrm{d}y\,\mathrm{d}z \le P^*_{\mathrm{out}}, \end{aligned} \tag{7}$$

where P^*_{out} is the outage requirement and $(E_{\mathrm{b}}/I_0)^0$ is associated with P^0_{BER}; $f_{I_{\mathrm{S}}/S_{\mathrm{e}}}(y)$ and $f_{I_{\mathrm{O}}/S_{\mathrm{e}}}(z)$ are the probability density functions of $I_{\mathrm{S}}/S_{\mathrm{e}}$ and $I_{\mathrm{O}}/S_{\mathrm{e}}$, respectively, and both of them are Gaussian distributed with different means and variances:

$$Q(x) = \int_x^\infty \frac{1}{\sqrt{2\pi}} \mathrm{e}^{-y^2/2}\,\mathrm{d}y.$$

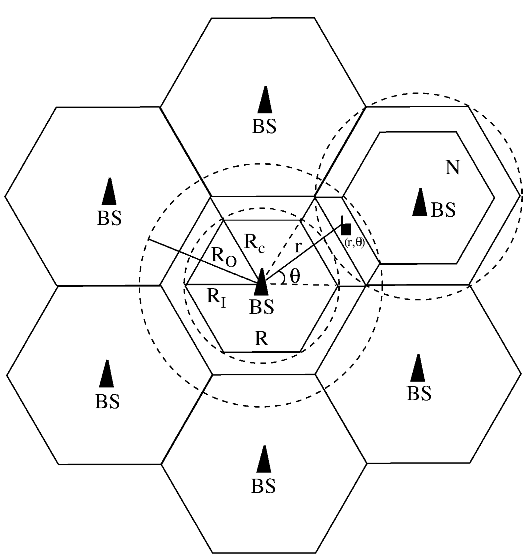

Fig. 2. Cell structure.

3.2. Call admission region

For a practical packet switching network, resource allocation can be performed on a frame basis. The frame duration is determined so that each voice source generates one packet in each frame during a talk spurt. A video call has a variable packet arrival rate in each frame duration. It is noted that buffering at the MS smoothes out the burstiness of the traffic in each connection. Waiting packets with time-out value equal to one are referred to as the most urgent packets (MUPs). And waiting packets with time-out value equal to two referred to as the second MUPs and so on. With the assumption of infinite buffer size, packet loss happens when the accumulated number of MUPs exceeds the cell capacity. Let N_1, $N_{2\mathrm{h}}$, and $N_{2\mathrm{l}}$ denote the number of ongoing voice, high- and low-quality video calls in the system. The number of MUPs from each call depends on its data generation rate, the delay bound and the wireless transmission capacity.

Let the integer random variable M, $M \in [0, M_{\max}]$, denote the accumulated MUPs in packets/frame, and $M_{\max}$ be the maximum value of M. Let $\bar{r}_i$ denote the average number of generated data packets for user i. Conditioned on the MUP load in a frame, $M=m$, the number of lost packets for user i is denoted by $r_{\mathrm{loss}_i|m}$. Thus, the actual PLR for user i, P^i_{PLR}, is the average number of lost MUPs divided by the average number of generated packets and is given by

$$P^i_{\mathrm{PLR}} = \frac{\sum_{m=C+1}^{M_{\max}} P_r\{M=m\} r_{\mathrm{loss}_i|m}}{\bar{r}_i}, \tag{8}$$

where $P\{M=m\}$ is the probability distribution function of M. It is observed that whatever M is, in each packet scheduling process, if $r_{\mathrm{loss}_i|m}$ is chosen in such a way that the following relation is satisfied

$$\frac{r_{\mathrm{loss}_i|m}}{r_{\mathrm{loss}_j|m}} = \frac{\bar{r}_i P^i_{\mathrm{PLR}}}{\bar{r}_j P^j_{\mathrm{PLR}}} \quad \forall\, i, \forall j \in \{1, \ldots, N_1 + N_{2\mathrm{h}} + N_{2\mathrm{l}}\}, \tag{9}$$

then when the required P^i_{PLR} is achieved for user i, the PLR for all the other users can be satisfied. It is noted that $r_{\mathrm{loss}_i|m}$ can be a non-integer which is not appropriate in wireless transmission. A high priority is given to the traffic with stringent PLR requirement for implementation simplicity. For instance, during a frame only two connections, i and j, have information to transmit and i has a more stringent PLR requirement than traffic j. The MUP loads from both connections are accumulated to be m. The actual packet loss for i and j is determined to be $\tilde{r}_{\mathrm{loss}_i|m} = \lfloor r_{\mathrm{loss}_i|m} \rfloor$ and $\tilde{r}_{\mathrm{loss}_j|m} = m - C - \tilde{r}_{\mathrm{loss}_i|m}$, where $\lfloor \cdot \rfloor$ is a floor function. The PLR requirement for j is guaranteed by limiting the admitted number of calls as shown in the following call admission region derivation. Given that the total number of MUPs from all the users is m, the packet loss in a frame is constrained by

$$\sum_{i=1}^{N_1+N_{2\mathrm{h}}+N_{2\mathrm{l}}} \tilde{r}_{\mathrm{loss}_i|m} = m - C \quad m \geq C. \tag{10}$$

Combining (9) and (10), the packet losses for each user in a frame can be obtained.

A scheduling procedure, which aims at providing fair packet transmissions among voice, high- and low-quality video calls with the strictly bounded PLR P_{PLR} for each individual call, can proceed as follows. At the beginning of each frame, all the waiting packets in the buffer have their time-out values decreased by one. Transmission requests for each connection, including the number of waiting packets and the tagged time-out values, are sent to the MSC via the serving BS. For instance, a call with a delay bound of D in frames can submit a transmission request as $(n_1,1)$, $(n_2,2),\ldots,(n_D,D)$ representing the number of MUPs, the second MUPs,…, and the least urgent packets. Based on (9) and (10), the serving MSC determines the transmission permissions of MUPs. If there are still resources available after all the MUPs are scheduled for transmission, the remaining cell capacity is used to accommodate the less urgent packets. The same criterion is applied to the less urgent packets until all the waiting packets in the buffer are scheduled or there is no cell capacity left. The transmission permissions are passed via the serving BSs to each user for uplink packet transmissions.

A burst of traffic arrivals in a frame duration can be divided into packets with fixed length. It is noted that due to the randomly varying data generation rate, the last packet may not be in full length for a video connection. This kind of packets are referred to as partial packets. If a partial packet is filled with redundant bits and occupies a full channel for transmission, resources are wasted. To make efficient use of the precious radio resources, a partial packet integration scheme is proposed to combine the partial packets with different time-out values for transmission. Due to the identification header in each packet for right order and correct reassembly [14], it is possible to implement packets integration at the sender and the disintegration at the receiver. When there is a partial MUP scheduled for transmission, the MSC will check if there is a second MUP transmission request so that it can be combined with the MUP for a possible transmission in the following frame. Fig. 3 demonstrates three cases of partial packet integration where 1 and 2 represent the time-out values of the consecutive waiting packets. The second MUPs are disassembled and reassembled with the partial MUP to form a new packet. In case (a), the partial MUP and the second MUP are combined into an exact full packet and occupy a full channel. Case (b) shows that after combination

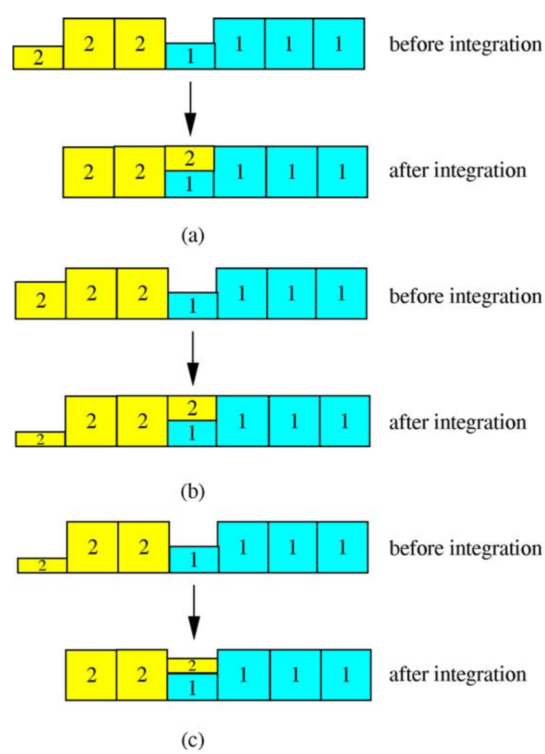

Fig. 3. Partial packet integration.

there is still a partial second MUP packet left, while in case (c), the combined packet is still a partial packet. In both cases (b) and (c), the same partial packet integration scheme can be applied to the third MUP and even less urgent packets to efficiently utilize the limited radio resources. In our simulation, we only combine the partial MUP and the second MUP for simplicity.

With the estimated uplink cell capacity and the proposed scheduling and partial packet integration scheme, a call admission region can be derived as a vector (N_1, $N_{2\mathrm{h}}$, $N_{2\mathrm{l}}$) subject to the PLR constraints for all users:

$$\begin{cases} \mathrm{Min}(P^i_{\mathrm{PLR}} - P^j_{\mathrm{PLR}}) \text{ and } P^i_{\mathrm{PLR}}, P^j_{\mathrm{PLR}} \le P^1_{\mathrm{PLR}} & 1 \le i,j \le N_1 \\ \mathrm{Min}(P^i_{\mathrm{PLR}} - P^j_{\mathrm{PLR}}) \text{ and } P^i_{\mathrm{PLR}}, P^j_{\mathrm{PLR}} \le P^{2h}_{\mathrm{PLR}} & 1 \le i,j \le N_{2\mathrm{h}} \\ \mathrm{Min}(P^i_{\mathrm{PLR}} - P^j_{\mathrm{PLR}}) \text{ and } P^i_{\mathrm{PLR}}, P^j_{\mathrm{PLR}} \le P^{2l}_{\mathrm{PLR}} & 1 \le i,j \le N_{2\mathrm{l}} \end{cases} \tag{11}$$

The first item represents a fair packet scheduling among users with the same traffic class, and the second item guarantees the strict PLR requirement for each traffic type. A call admission region regulates the admitted number of voice, high-quality and low-quality video calls under the constraint that a call admission takes place only within the call admission region.

3.3. Adaptive handoff priority scheme

Call admission region limits the admitted number of calls of all classes in the system By adaptively transforming between high- and low-quality video calls at heavy traffic load environments to prioritize handoff call admission, handoff call dropping probability can be improved.

The proposed adaptive handoff priority scheme prioritizes handoff calls by reducing the transmission rate of ongoing high-quality video calls and admitting handoff calls with the gained resources. When a call admission request, which can either be a voice call or a video call served with either high or low quality, arrives, call admission decision is made based on the derived call admission region. For new and handoff calls, different CACs are applied. New voice and video calls are accepted only when the transmission rate and the overall BER can be satisfied. On the other hand, if a voice handoff call arrives, the MSC will first check to see if the spare resource can be used to accept the handoff voice call. When the spare resource cannot support a satisfactory voice handoff call or there is no spare resource available, and if there is a rate-adaptive high-quality video call in progress, the rate renegotiation will be triggered. After both the correspondents agree to reduce the transmission rate, the MSC will do the CAC again. Service degradation of more than one rate-adaptive high-quality video calls may be necessary to admit a handoff call according to the traffic characteristics. A voice handoff call can be accepted if the overall BER requirements of the handoff call and of all other ongoing calls in the system can be achieved; otherwise, it is dropped. Fig. 4 shows the voice handoff CAC diagram.

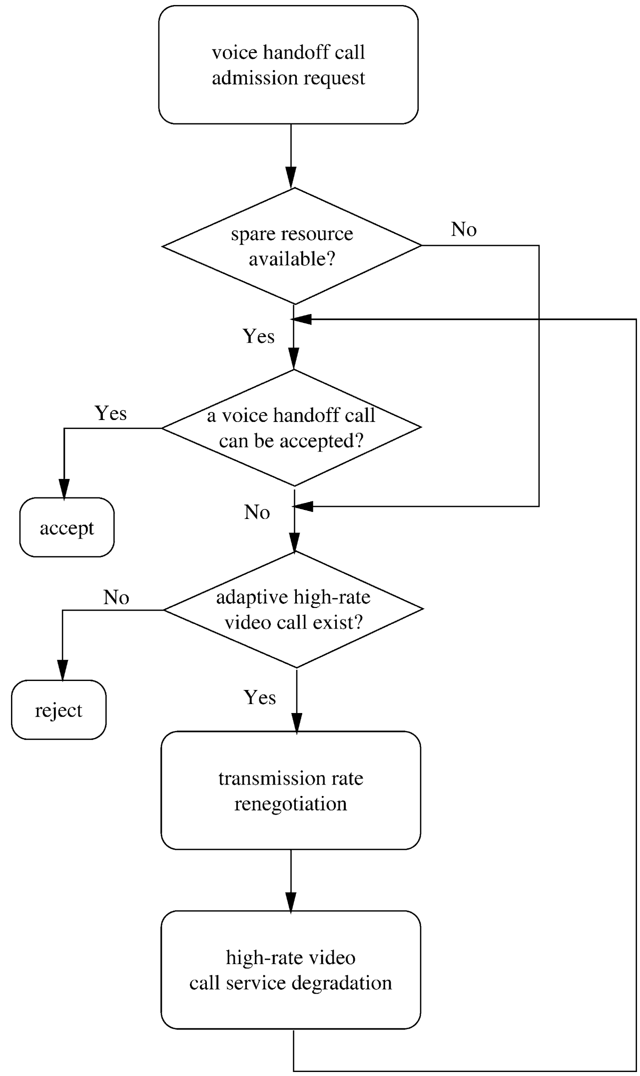

Fig. 4. Voice handoff call admission control diagram.

Due to the fact that there are two types of video calls, high- and low-quality, in the system, handoff and CAC schemes for video calls are of a bit different from that for voice calls. Fig. 5 shows the video handoff CAC flow chat. No matter whether the coming handoff video call is served with high or low quality, the MSC will first try to provide the handoff video call with high quality service with the spare resource. If the available resource is not enough, then low quality service will be provided. When the spare resource can neither provide high- nor low-quality service for the video handoff call, or when at a heavy traffic load environment with no spare resource available, the same rate adaptation process is executed. In this way, the admitted video handoff call is served with low transmission rate. It is seen in Fig. 5 that the gained resource from rate adaptation is only used to support a low-quality video handoff call. There are four possibility of service quality for a video call before and after handoff. A high-quality video call can be

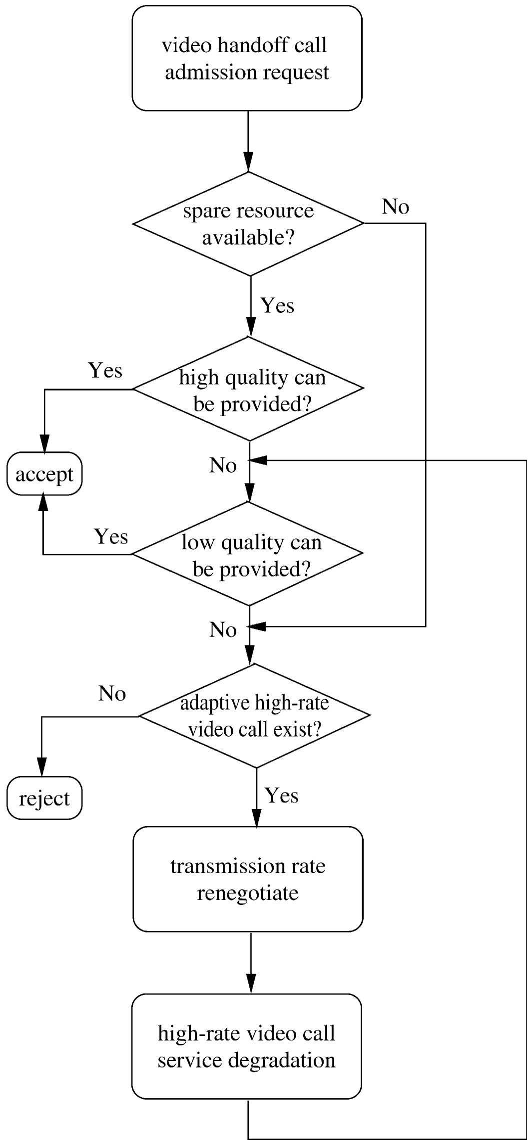

Fig. 5. Video handoff call admission control diagram.

provided with either high or low quality after a successful handoff, and the same for a low-quality video handoff attempt depending on the traffic load and resource utilization in the new target cell.

In summary, let the number of ongoing calls of all classes in the cell be (N_1, N_{2h}, N_{2l}), a voice or video handoff is successfully executed if and only if

$$(N_1+1, N_{2h}-n_2, N_{2l}+n_2), \quad N_{2h} \geq n_2 \tag{12}$$

for a voice handoff call or

$$\begin{cases} (N_1, N_{2h}+1, N_{2l}) & \text{or} \\ (N_1, N_{2h}-n_2, N_{2l}+n_2+1), & N_{2h} \geq n_2 \end{cases} \tag{13}$$

for a video handoff call is still within the derived call admission region. Here $n_2 \geq 0$ is the total number of rate-adaptive high-quality video calls that have to suffer quality degradation. $n_2=0$ indicates that there is no service degradation for high-quality video calls for handoff call priority. Eq. (13) also indicates that either high or low QoS can be provided to a video handoff call, and only low QoS is provided if service degradation is needed.

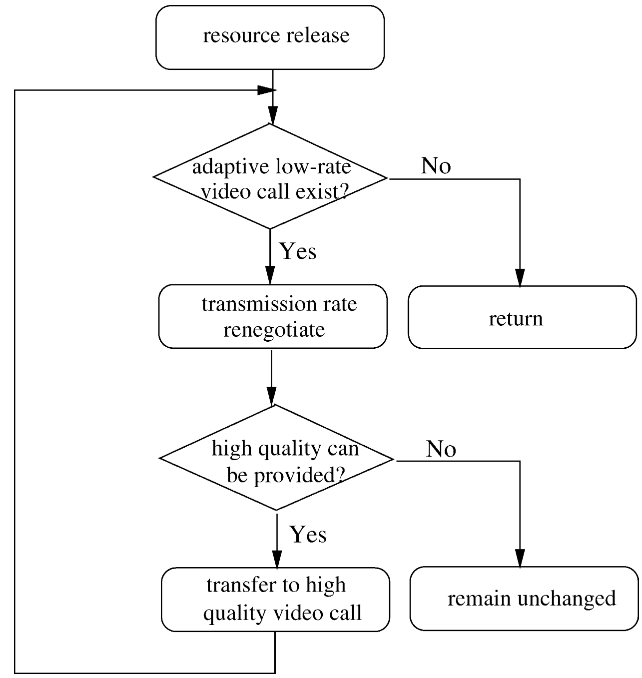

Fig. 6. Low-quality video call transmission rate adjustment.

From both user's and service provider's perspective, the ratio of the number of low-quality video calls over the total number of video calls in the system should be kept as low as possible for better resource utilization and user satisfaction. Therefore, whenever there is resource release due to call termination or call handoff to a neighboring cell, a low-quality video call will try to improve its transmission rate. Fig. 6 shows the flow chart of transmission rate adjustment of low-quality video calls at resource release. If the released resource is not sufficient for a low-quality video call to be provided with high QoS, the video call remains with low QoS till next time resource release happens.

4. Simulation results

Consider an MC-CDMA system with a spread spectrum bandwidth of 5 MHz. The relative soft handoff area is chosen to be $\alpha=0.4$, i.e. the soft handoff area corresponds to 36% of a normal cell area [15]. The basic transmission rate B is 16 kbps for one single code channel. The path loss exponent n is set to 4 and the standard deviation for shadowing effect is $\sigma=8$ dB. The standard deviation due to imperfect power control is set to $\sigma_e=2$ dB. Because CDMA is an interference dominant system, we ignore background noise for simplicity. Without consideration of packet error rate imposed from wireline domain, the overall packet error rate requirements for voice and low-rate video

Table 1
Simulation parameters

Parameters	Value
Spread spectrum bandwidth W (MHz)	5
Basic transmission rate B (kbps)	16
Path loss exponent n	4
Lognormal shadowing ζ	$\zeta - N(0, 8\text{ dB})$
Imperfect power control ξ_e	$\xi_e - \text{N}(0, 2\text{ dB})$
Frame length	10 ms (160 bits)
Outage probability P^*_{out}	10^{-2}
Voice traffic	
Average talk spurt duration (s)	0.35
Average silent period (s)	0.65
Tolerable time-out value in frames	2
Packet error rate requirement	10^{-3}
Video traffic	
Number of minisources N	20
ON duration (s)	0.2
OFF period (s)	0.8
Transmission rates during ON state (kbps)	8, 4
Tolerable time-out value in frames	3
Packet error rate requirement	10^{-4}

(for applications such as video conferencing) communications are 10^{-3} and 10^{-4}, respectively. The upper bound of the outage probability P^*_{out} is set to 10^{-2}. Each frame is 10 ms in length which is equivalent to $L = 160$ bits.

The voice traffic is simulated by the ON–OFF model. During ON state, one packet is generated in each frame. On average, an ON state lasts 0.35 s and an OFF state lasts 0.65 s. The video traffic has a variable transmission rate, which is simulated by 20 minisources with average ON and OFF durations of 0.2 and 0.8 s. During the ON state of a minisource, the transmission rates are 8 and 4 kbps for high- and low-quality video calls, respectively. The delay bound for voice and video packets are 2 and 3 frames, respectively. The simulation parameters are tabulated in Table. 1.

To achieve maximum revenue for service providers, and at the same time to meet user satisfaction, a grade of service (GoS) function is defined by considering both the new call blocking probability P_{b} and the weighted handoff call dropping probability P_{d}

$$\text{GoS} = P_{\text{b}} + \gamma P_{\text{d}}, \tag{14}$$

where γ is a weighting factor to emphasize more on handoff call dropping probability. γ equals 10 is recommended in Ref. [16]. It is noted that P_{b} and P_{d} are a pair of conflicting performance measures. The more new calls are blocked, the more resources will be available for handoff calls, which results in an increased new call blocking probability and a decreased handoff call dropping probability, and vice versa. To minimize the GoS function, the ratio of the variation of new call blocking probability and handoff call dropping probability should be kept the same as γ [17]. In the simulation the ratio of new call blocking and handoff call dropping probabilities are approximately kept at value γ to achieve the best GoS performance.

Tables 2 and 3 give the call admission regions (N_1, $N_{2\text{h}}$, $N_{2\text{l}}$) estimated with the traditional circuit switching transmission and the proposed packet switching transmission, respectively. The row numbers represent the number of low-quality video calls and the column numbers represent the number of high-quality video calls that can be admitted. The table entries represent the number of voice calls. The call admission region is derived with $P^*_{\text{WTE}} = 10^{-4}$. It can be seen from the tables that the proposed packet switching transmission enlarges the call

Table 2
Call admission region with circuit switching transmission ($P^*_{\text{WTE}} = 10^{-4}$)

		$N_{2\text{l}}$									
		0	1	2	3	4	5	6	7	8	9
$N_{2\text{h}}$	0	38	35	31	28	25	21	17	14	11	7
	1	32	29	25	21	18	14	11	8	3	
	2	26	23	19	15	12	8				
	3	20	16	12	9	2					
	4	12	9	5							
	5	6									

Table 3
Call admission region with packet switching transmission ($P^*_{\text{WTE}} = 10^{-4}$)

		$N_{2\text{l}}$												
		0	1	2	3	4	5	6	7	8	9	10	11	12
$N_{2\text{h}}$	0	40	35	32	28	25	22	19	16	12	8	5	1	
	1	33	31	27	24	21	18	15	12	8	5	3		
	2	28	25	22	19	16	13	10	7	3				
	3	23	20	17	13	10	7	3	1					
	4	17	14	11	7	4	2							
	5	11	7	5	1									
	6	6	1											

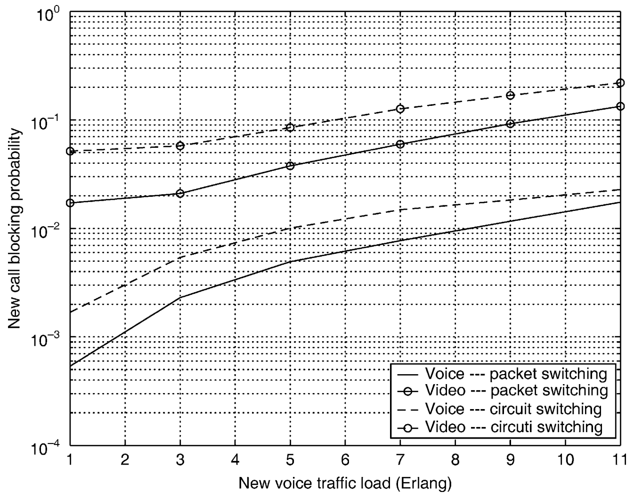

Fig. 7. New call blocking probability with the packet switching and circuit switching transmissions.

admission region significantly compared to that from circuit switching transmission for the same BER requirement and cell capacity. The enlarged call admission region can accommodate more calls, which improves the new call blocking probability, the handoff call dropping probability and the radio resource utilization. Figs. 7 and 8 show the new call blocking and handoff call dropping probabilities with the call admission regions given in Tables 2 and 3. The new video traffic load is fixed at 2 Erlang. The handoff traffic load to the new traffic load for voice and video are set to be 1/3 and 1/10, respectively. It can be seen that the new call blocking and the hand-off call dropping probabilities are significantly reduced with the enlarged call admission region. Due to the much higher transmission rate and more stringent transmission error rate requirement, video calls suffer relatively higher new call blocking and handoff call dropping probabilities compared with voice calls. It can also

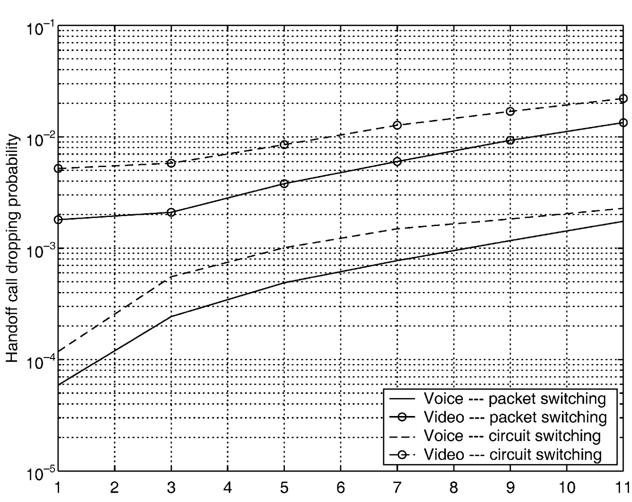

Fig. 8. Handoff call dropping probability with the packet switching and circuit switching transmissions.

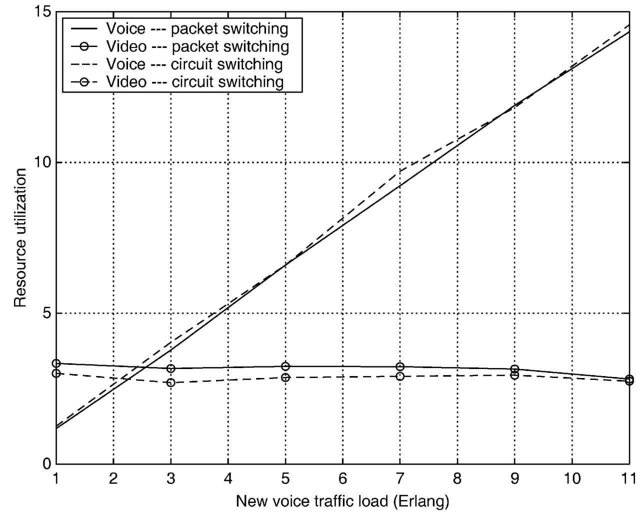

Fig. 9. Resource utilization with the packet switching and circuit switching transmissions.

be seen that the ratio of new call blocking probabilities and handoff call dropping probabilities is around 10.

Resource utilization in terms of the admitted voice and video calls vs. new voice traffic load is shown in Fig. 9. With the increasing voice traffic load, the resource utilization for voice traffic increases accordingly, and that for video traffic decreases gradually. This is due to the increasing new call blocking and handoff call dropping probabilities for video calls. The figure also shows that with the enlarged call admission region, the video traffic resource utilization is increased.

The ratio of the number of low-quality video calls over the total number of video calls in the system is shown in Fig. 10. It can be seen that the low-quality video call occupancy increases with the voice traffic load. For the maximum traffic load in the simulation, the ratio obtained

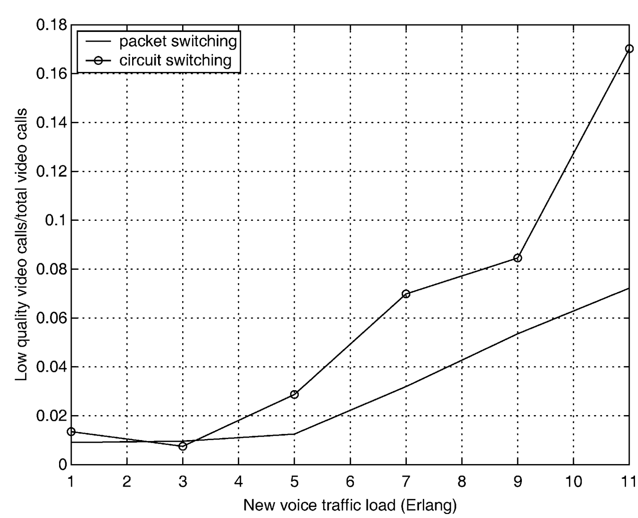

Fig. 10. Ratio of the number of low-quality video calls over the total number of video calls with the packet switching and circuit switching transmissions.

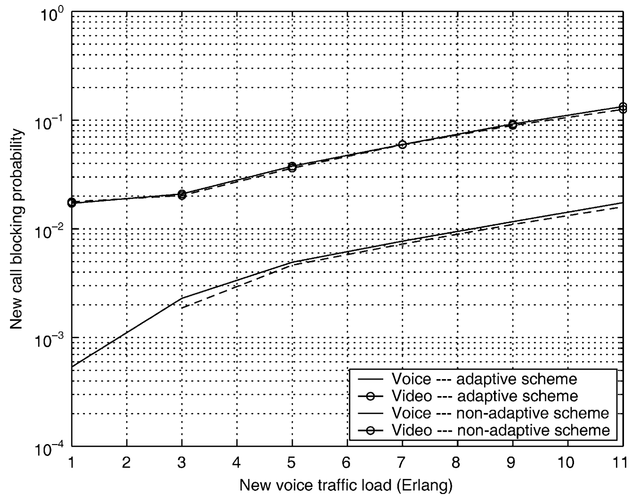

Fig. 11. New call blocking probability with the adaptive priority and non-adaptive handoff schemes.

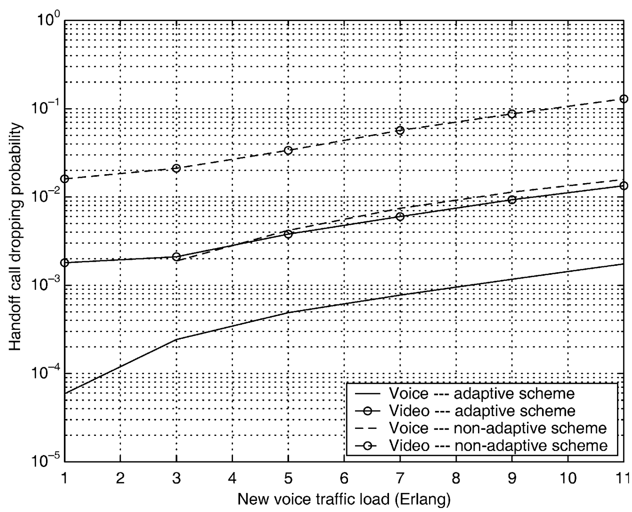

Fig. 12. Handoff call dropping probability with the adaptive priority and non-adaptive handoff schemes.

with the proposed packet switching wireless transmission scheme is 0.07, which significantly outperforms the ratio obtained with the traditional circuit switching wireless transmission scheme, which is 0.17. This result can be reasonably accepted with user satisfaction.

Figs. 11 and 12 show the new call blocking and the handoff call dropping probabilities with the proposed adaptive handoff scheme and the non-adaptive handoff scheme, respectively, where the transmission rate of video calls is fixed during call holding time. Due to the transmission rate reduction of adaptive rate video calls to accommodate handoff calls, both handoff voice calls and handoff video calls are prioritized with much lower dropping probabilities compared to the new call blocking probabilities. Due to the admission of more handoff calls into the system, the performance of new call blocking probabilities of both voice and video calls is slightly reduced compared to those obtained with the non-adaptive handoff schemes.

With different P^*_{WTE} settings, different multiplexing gain and different call level performance measures can be achieved. For demonstration purpose, we set P^*_{WTE} to 10^{-4} and 10^{-5}, respectively, for call admission region derivation. When P^*_{WTE} equals to 10^{-4}, additional error rate can be introduced in packet scheduling for voice traffic, while no more packet loss can be endured for video traffic. On the other hand, when P^*_{WTE} is set to 10^{-5}, both voice and video traffic can introduce packet loss according to (3). Table 4 gives the shrunk call admission region derived with $P^*_{\mathrm{WTE}} = 10^{-5}$ due to the degraded multiplexing gain in wireless transmission compared to that in Table 3. Figs. 13–15 show the performance comparisons with different prescribed P^*_{WTE} values. As expected, the performance in terms of new call blocking probability, handoff call dropping probability, and the ratio of the number of low-quality video calls over the total number of ongoing video calls obtained with a less stringent P^*_{WTE} value of 10^{-4} significantly outperforms that obtained with a more stringent P^*_{WTE} value of 10^{-5}.

5. Conclusions

An adaptive handoff priority scheme for MC-CDMA cellular networks supporting realtime multimedia services has been proposed. By jointly considering physical, link and network layer characteristics, the scheme can significantly reduce the handoff call dropping probability with a slight increase in new call blocking probability and an acceptable

Table 4
Call admission region with packet switching transmission ($P^*_{\mathrm{WTE}} = 10^{-5}$)

		N_{21}										
		0	1	2	3	4	5	6	7	8	9	10
N_{2h}	0	35	32	27	24	21	17	14	11	8	4	
	1	29	26	23	20	16	14	10	7	4	1	
	2	24	21	17	15	11	9	5	2			
	3	18	16	12	9	6	3					
	4	13	9	6	3							
	5	7	4	1								
	6	1										

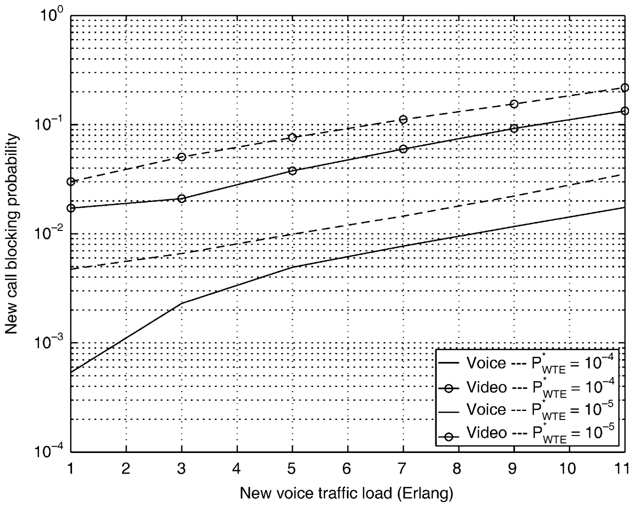

Fig. 13. New call blocking probability with P^{*}_{WTE} being 10^{-4} and 10^{-5}.

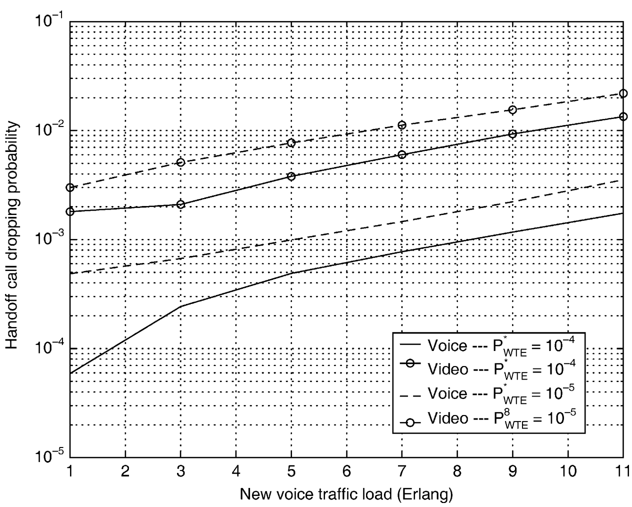

Fig. 14. Handoff call dropping probability with P^{*}_{WTE} being 10^{-4} and 10^{-5}.

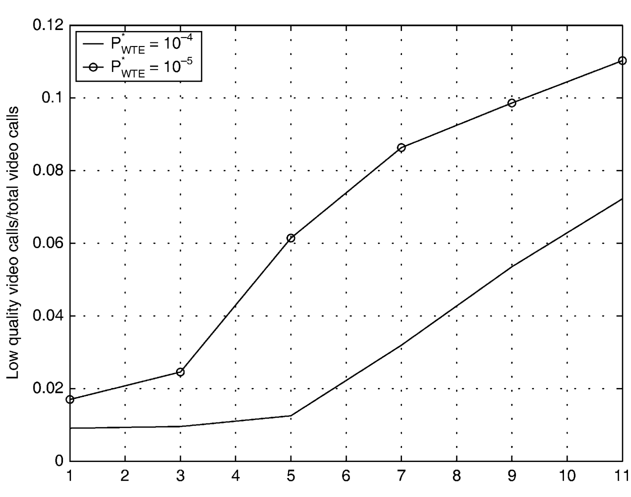

Fig. 15. Ratio of the number of low-quality video calls over the total number of video calls with P^{*}_{WTE} being 10^{-4} and 10^{-5}.

video service degradation. This research should be helpful for the development of resource and handoff management in future wireless communication networks.

Acknowledgements

This work has been supported by a grant from the National Science and Engineering Research Council (NSERC) of Canada under strategic project and a Bell/ORCDF Scholarship.

References

[1] C.-L. I, R.D. Gitlin, Multi-code CDMA wireless personal communications networks, Proc. IEEE ICC' 2 (1995) 1060–1064.
[2] W. Li, A.S. Alfa, Channel reservation for handoff calls in a PCS network, IEEE Trans. Veh. Technol. 48 (2000) 95–104.
[3] Y.C. Kim, D.E. Lee, B.J. Lee, Y.S. Kim, B. Mukherjee, Dynamic channel reservation based on mobility in wireless ATM networks, IEEE Commun. Mag. November (1999) 47–51.
[4] L. Ortigoza-Guerrero, A.H. Aghvami, A prioritized handoff dynamic channel allocation strategy for PCS, IEEE Trans. Veh. Technol. 48 (4) (1999) 1203–1215.
[5] C. Oliveira, J.B. Kim, T. Suda, An adaptive bandwidth reservation scheme for high-speed multimedia wireless networks, IEEE J. Select. Areas Cummun. 16 (1998) 858–874.
[6] P. Ramanathan, K.M. Sivalingam, P. Agrawal, S. Kishore, Dynamic resource allocation schemes during handoff for mobile multimedia wireless networks, IEEE J. Select Areas Commun. 17 (1999) 1270–1283.
[7] Q. Huang, S. Chan, K.-T. Ko, Dynamic bandwidth sharing for handoff traffic in multimedia mobile cellular networks, Electron. Lett. 39 (10) (2003) 801–802.
[8] Y.-B. Lin, A.R. Noerpel, D.J. Harasty, The sub-rating channel assignment strategy for PCS hand-offs, IEEE Trans. Veh. Technol. 45 (1) (1996) 122–130.
[9] W. Zhuang, B. Bensaou, K.C. Chua, Adaptive quality of service handoff priority scheme for mobile multimedia networks, IEEE. Trans. Veh. Technol. 49 (2000) 494–505.
[10] K. Lee, Supporting mobile multimedia in integrated services networkings, Wireless Networks 2 (3) (1996) 205–217.
[11] M. Schwartz, Broadband Integrated Networks, Prentice Hall, New Jersy, 1996.
[12] S. Ariyavisitakul, L. Chang, Signal and Interference statistics of a CDMA system with feedback power control, IEEE Trans. Commun. 41 (1993) 1626–1634.
[13] J. Zhang, J.W. Mark, X. Shen, Joint packet- and call-level soft handoff in CDMA wireless cellular networks, Proc. 14th IEEE PIMRC 2003 2 (2003) 2485–2489.
[14] F.H.P. Fitzek, B. Rathke, M. Schlager, A. Wolisz, Quality of service support for real-time multimedia applications over wireless links using the simultaneous MAC-packet transmission (SMPT) in a CDMA environment, Proc. MoMuC 36 (1998) 367–378.
[15] D.K. Kim, D.K. Sung, Characterization of soft handoff in CDMA systems, IEEE Trans. Veh. Technol. 48 (1999) 1195–1202.
[16] S. Boumerdassi, A.-L. Beylot, Adaptive channel allocation for wireless PCN, Mobile Netw. Appl. 4 (1999) 111–116.
[17] J. Zhang, J.W. Mark, X. Shen, A novel resource reservation sheme for handoff in CDMA wireless cellular networks, Proc. IEEE WCNC 3 (2003) 2069–2074.