Radio Resource Management for Cellular CDMA Systems Supporting Heterogeneous Services

Dongmei Zhao, *Member*, *IEEE*, Xuemin (Sherman) Shen, *Senior Member*, *IEEE*, and Jon W. Mark, *Fellow*, *IEEE*

Abstract—A novel radio resource management (RRM) scheme for the support of packet-switched transmission in cellular CDMA systems is proposed by jointly considering the physical, link, and network layer characteristics. The proposed resource management scheme is comprised of a combination of power distribution, rate allocation, service scheduling, and connection admission control. Power distribution allows individual connections to achieve their required signal-to-interference-plus-noise ratio, while rate allocation guarantees the required delay/jitter for real-time traffic and the minimum transmission rate requirement for non-real-time traffic. Efficient rate allocation is achieved by making use of the randomness and burstiness of the packet generation process. At the link layer, a packet scheduling scheme is developed based on information derived from power distribution and rate allocation to achieve quality of service (QoS) guarantee. Packet scheduling efficiently utilizes the system resources in every time slot and improves the packet throughput for non-real-time traffic. At the network layer, a connection admission control (CAC) scheme based on the lower layer resource allocation information is proposed. The CAC scheme makes use of user mobility information to reduce handoff connection dropping probability (HCDP). Theoretical analysis of the grade of service performance, in terms of new connection blocking probability, HCDP, and resource utilization, is given. Numerical results show that the proposed RRM scheme can achieve both effective QoS guarantee and efficient resource utilization.

Index Terms—Code division multiple access, radio resource management, grade of service, quality of service, packet scheduling, connection admission control.

1 Introduction

With the growing demands for multimedia services and the high degree of user mobility, radio resource management (RRM) plays a paramount important role in future code division multiple access (CDMA) systems to efficiently utilize the limited radio resources and to provide more mobile users with guaranteed quality of service (QoS) anywhere at any time. The effectiveness and efficiency of the RRM are affected by the system characteristics at the physical, link, and network layers.

Because of the wireless channel impairments, no practical system can utilize the entire system capacity. In the overall system design, it is important to maximize system resource utilization while satisfying QoS requirement. Power distribution and rate allocation are the basis to achieve this objective in a CDMA system where system resources are shared by all active users. It is shown in [1] that an algorithm, which achieves equal receive power for homogeneous services in a uniformly loaded CDMA system, is optimal in the sense that resource utilization is maximized. An approach which assigns different transmitted powers to achieve different transmission bit error rates (BERs) in CDMA systems is proposed in [2]. An optimal power control law is derived in [4] for multirate multimedia traffic to maximize the resource utilization. Power distribution and resource management are studied in [3] as an optimization problem to minimize the total interference or to maximize the transmission throughput. Provision of a large variety of transmission rates may be implemented in two ways: variable system processing gain CDMA [5] or multicode CDMA (MC-CDMA) [6]. The former allocates to each user one code channel with a variable spreading gain, and the latter allocates multiple codes to each user to avoid the need for high rate transmissions in each code channel. A combination of these two approaches is also feasible.

Medium access control (MAC) and packet scheduling at the link layer are designed to coordinate the packet transmissions from different users to efficiently utilize the system resources and guarantee the required QoS, including transmission delay/jitter and BER. A MAC protocol, referred to as MC-CDMA with distributed-queueing request update multiple access (DQRUMA), for multirate traffic is proposed in [9]. The MAC protocol is extended in [10] to incorporate different traffic classes. In [11], a packet scheduling scheme called WISPER is developed to support multimedia traffic based on a hybrid time division multiple access (TDMA) and MC-CDMA system. In both [10] and [12], real-time traffic, such as voice or video, is served in a circuit switching mode to reduce the request access overhead and ensure the required small delay, and non-real-time

- *D. Zhao is with McMaster University, Hamilton, ON L8S 4K1, Canada. E-mail: dzhao@mail.ece.mcmaster.ca.*
- *X. Shen and J.W. Mark are with the University of Waterloo, Waterloo, ON N2L 3G1, Canada. E-mail: {xshen, jwmark}@bbcr.uwaterloo.ca.*

Manuscript received 8 Jan. 2003; accepted 22 Apr. 2003.
For information on obtaining reprints of this article, please send e-mail to: tmc@computer.org, and reference IEEECS Log Number 4-012003.

traffic is served in a packet switching mode to improve the resource utilization and the packet throughput. Scheduling algorithms, which dynamically adjust transmission rates for non-real-time traffic based on current channel conditions and traffic load, can also be found in [13], [14].

Connection admission control (CAC) at the network layer can decide whether a new or handoff connection should be admitted into the system. A CAC scheme for homogeneous services is proposed in [15], where the admission decisions are determined by the interference level. An SIR-based CAC scheme is proposed in [16] for homogeneous services. A connection is accepted if, after the admission, system-wide outage probability can still be guaranteed. Two different CAC schemes are considered in [17] for homogeneous services: number-based CAC (NCAC) and interference-based CAC (ICAC). The NCAC admits a new connection if the total number of existing connections in the system is less than a predefined value, while the ICAC admits a new connection if the total interference in the system is less than a certain threshold. A CAC scheme for power controlled CDMA system is proposed in [18]. For cellular systems, the CAC should give handoff connections a higher priority than new connections to achieve lower handoff connection dropping probability (HCDP) than new connection blocking probability (NCBP) because interrupting an on-going connection is much more undesirable than refusing to admit a new connection from the user's point of view. This issue becomes increasingly important because frequent handoffs may occur in future wireless cellular networks due to reduced cell size for improving the system capacity. The guard channel approach [19], virtual connection tree (VCT) approach [20], shadow cluster approach [21], etc. are proposed to reserve resources for handling potential handoff connections in order to achieve reduced HCDP. A CAC scheme which reserves resources for potential handoff connections based on user mobility information is proposed in [22] for TDMA-based cellular networks. A framework which integrates mobility prediction into CAC for QoS guarantee is proposed in [23].

Effective and efficient power distribution and rate allocation provide the basis for the higher-layer resource management. Packet scheduling at the link layer is necessary to coordinate the transmission rate and power of different connections in the system to satisfy the packet-level QoS requirements and to achieve high packet throughput. CAC at the network layer makes an admission decision for a connection based on the transmission capability of the lower layer. Admitting more connections than the packet scheduler can handle will result in network congestion and the inability to guarantee QoS performance; but, admitting fewer connections than the capacity of the scheduler will underutilize the system resources. An effective and efficient CAC should also reserve an appropriate amount of resources for potential handoff connections in order to achieve low HCDP. However, reserving more resources leaves less resources for new connections and may increase NCBP. In general, different system characteristics at the physical, link, and network layers may be correlated. Consideration of the individual layers in isolation can lead to inefficient resource utilization and problems in QoS provisioning. In this paper, a novel RRM scheme is proposed for cellular CDMA systems by jointly considering the system characteristics from the physical, link, and network layers. The objective of the proposed RRM is to efficiently utilize the available system resources while providing more mobile users with guaranteed QoS. Specifically, the power distribution only allocates the necessary amount of power to each connection in order to achieve its required signal-to-interference-plus-noise ratio (SINR). The rate allocation guarantees the required delay/jitter for real-time traffic and the minimum transmission rate requirement for non-real-time traffic. Efficient rate allocation is achieved by making use of the randomness and burstiness of the packet generation process. At the link layer, a packet scheduling scheme based on the power distribution and rate allocation is developed. It schedules the system resources on a time slot basis to efficiently utilize the system resources in every time slot and to improve the packet throughput for non-real-time traffic. A CAC scheme based on the lower-layer resource allocation information is proposed at the network layer. The CAC scheme also makes use of user mobility information to reserve resources for handoff connections. Theoretical analysis of the grade-of-service (GOS) performance, in terms of NCBP, HCDP, and resource utilization, is given.

The remainder of this paper is organized as follows: Section 2 describes the system model. The rate allocation and the power distribution are studied in Sections 3 and 4, respectively. Section 5 describes the proposed CAC scheme. A packet scheduling scheme is developed in Section 6. The GOS performance is derived in Section 7. Numerical results are given in Section 8, followed by conclusions in Section 9.

2 System Model

Consider a cellular CDMA system which is connected to a wireline backbone network through a mobile switching center (MSC). The service area of the MSC consists of several radio cells, each of which is the coverage area of one base station (BS). A mobile station (MS) selects the nearest BS as its associated BS and is power controlled by the BS. This work is confined to the coverage area of one MSC and focuses on the uplink, i.e., the link from the MSs to the BSs. Different frequency bands are used for uplink and downlink, i.e., frequency-division-duplexing (FDD), so there is no interference between the downlink and the uplink transmissions.

This work is concerned with the support of multirate transmissions in an MC-CDMA system. Here, each packet over the radio channel is transmitted at a basic rate, R_b. An individual connection can only transmit at rate mR_b, i.e., transmit m packets simultaneously, where m is a nonnegative integer. Because of its unified architecture, when using MC-CDMA to integrate heterogeneous traffic, traffic streams with significantly different transmission rates can be easily integrated with all the transmission channels having the same bandwidth and spread spectrum processing gain. Because of the subcode concatenation [6] in an MC-CDMA transmitter, orthogonality is achieved between the different simultaneously transmitted packets from the same user, and there is no interference between the simultaneously transmitted packets from the same user.

We assume that the end-to-end transmissions are bottlenecked by the radio transmission link. Therefore, the QoS considered in this work is for the radio transmission link only. We consider that each MS can carry only one connection. Each connection has one ready-to-transmit (RTT) buffer located at the MS to temporarily store its generated packets before transmission. The time difference between the instant a packet arrives to and the instant it departs from the RTT buffer is its experienced delay. The delay variation with respect to the mean delay is called jitter. There are four types of traffic under consideration: voice, video, data traffic with QoS requirement (Type I data), and data traffic without QoS requirement (Type II data). Each type of the traffic is further divided into classes. There are K_1 voice classes, indexed by $i = 1, 2, \ldots, K_1$; K_2 video classes, indexed by $i = K_1 + 1, K_1 + 2, \ldots, K_1 + K_2$; K_3 Type I data classes, indexed by $i = K_1 + K_2 + 1, K_1 + K_2 + 2, \ldots, K_1 + K_2 + K_3$; and one Type II data class, indexed by $i = K + 1$, where $K = K_1 + K_2 + K_3$.

The above traffic types have the following attributes where the subscript i takes on a different range of values that distinguish the different traffic types:

1. Each voice connection is characterized by a 3-tuple: $(\rho_i, \delta_i, \mathrm{BER}_i^*)$, where ρ_i is the packet generation rate, δ_i is the maximum jitter tolerance, and BER_i^* is the maximum tolerable BER. Since the maximum jitter is limited by the maximum delay, D_i, only the maximum tolerable delay is considered. A voice connection may contain both silent and active intervals. In the active intervals, packets are generated at constant rate ρ_i, while in the silent intervals, no packets are generated.
2. Each video connection is characterized by a 4-tuple: $(\rho_i, \sigma_i, D_i, \mathrm{BER}_i^*)$, where ρ_i is the average packet generation rate, σ_i is the maximum tolerable burstiness, D_i is the maximum tolerable delay, and BER_i^* is the maximum tolerable BER. Each video connection is regulated by a leaky bucket with token generation rate ρ_i and token buffer size $(\sigma_i - 1)$.
3. For Type I data, each connection is characterized by a 2-tuple: $(\rho_i, \mathrm{BER}_i^*)$, where ρ_i is the minimum transmission rate and BER_i^* is the maximum tolerable BER. The data connection may require a higher transmission rate than ρ_i, but the extra requests are served by best effort.
4. All Type II data connections are grouped as one class. A Type II data connection requires only best effort service. The system can assign a BER to the connection with any value between 0 and 1.

Propagation delay in cellular networks is usually negligibly small (from a few μs to a few 10s of μs) compared to the queueing delay (a few ms or more). As far as the uplink transmission is concerned, packet delay is mainly due to queueing delay in the RTT buffer. Transmission errors may be caused by channel noise and interference and buffer overflow. Let p_i^{L} be the packet loss rate due to buffer overflow. For a fixed packet size, the BER, p_i^{b}, caused by buffer overflow is the same as p_i^{L}. Since the overflowed packets are not transmitted in the channel, the total BER for connection i is: $\mathrm{BER}_i = p_i^{\mathrm{b}} + (1 - p_i^{\mathrm{b}})p_i^{\mathrm{w}}$, i.e., $p_i^{\mathrm{w}} = \frac{\mathrm{BER}_i - p_i^{\mathrm{b}}}{1 - p_i^{\mathrm{b}}}$, where p_i^{w} is the BER due to channel noise and interference. For an ith class connection, its p_i^{w} performance corresponds to the required SINR at the BS despread output, γ_i^*, for given modulation and coding schemes.

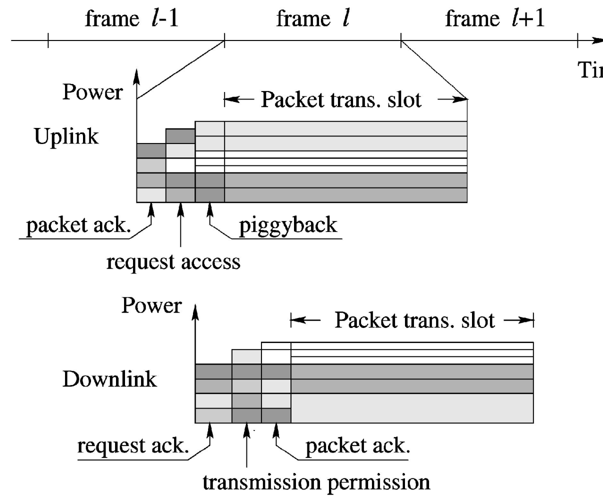


Fig. 1. Uplink and downlink frame architecture.

Channel time is divided into frames of equal length T_{f}. The frame structure of uplink and downlink is shown in Fig. 1. Each uplink frame is divided into a minislot for acknowledgment of downlink packet transmissions, a request access minislot for contention-based transmission requests, a piggyback rate-request minislot, and a packet transmission slot. Each downlink frame is divided into a request acknowledgment minislot for acknowledgment of contention-based transmission requests in the uplink, a transmission permission minislot, a minislot for acknowledgment of uplink packet transmissions, and a downlink packet transmission slot.

A new voice, video, or Type I data connection needs to request admission through the request access minislot before it can transmit any packets. Upon receiving an admission request from the voice, video, or Type I data connection, the BS forwards it to the MSC which makes an admission decision. If the connection can be admitted, the admission decision is sent to the MS through the BS via the downlink request acknowledgment minislot. It also includes the allocated transmission rate for the voice or video connection. A constant transmission rate is allocated to the voice or video connection to ensure its delay/jitter requirement. A dedicated code channel and the corresponding receiver are allocated to each admitted voice and video connection and released when the connection is completed, or handed off to a new cell. Therefore, no packet transmission requests are needed for an admitted voice or video connection. The admitted voice, video, and Type I data connections, and any Type II data connections may transmit packets. A newly admitted Type I data connection, or a Type II data connection whose buffer was empty after the previous transmission, or did not transmit packets in the previous packet transmission slot, may make a packet

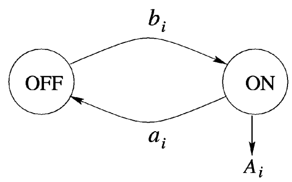


Fig. 2. ON-OFF model.

transmission request via the request access minislot in the uplink. When a data connection is permitted to transmit packets in the packet transmission slot of a frame, it can use the piggyback minislot to update the required transmission rate for the next scheduled frame. Successful packet transmission requests are broadcast in the downlink request acknowledgment minislot before the next uplink request access minislot. A packet scheduler at the MSC determines the transmission rate allocated to each connection, and the required power for each active connection in the system. The allocated transmission rate for each data connection and the distributed power for all active connections are broadcast at each BS via the transmission permission minislot in the downlink. Then, the BS assigns a number of code channels and the corresponding receivers to the scheduled data connections. Uplink packet transmissions are acknowledged through the packet acknowledgment minislot in the downlink. Transmissions at the request access minislot may be received erroneously at the BS due to interference or noise or due to the fact that different transmission requests may select the same code channels. Approaches such as harmonic backoff [7] and Binary Stack Algorithm [8] can be used to achieve reasonably high throughput and low request access delay.

3 Rate Allocation

A voice connection can be modeled by an ON-OFF source as is shown in Fig. 2, where the ON and OFF states represent the active and the silent conditions of the voice connection, respectively. a_i and b_i are the transition probabilities from the ON state to the OFF state and from the OFF state to the ON state, respectively. A_i is a constant packet generation rate in the ON state, and $A_i = \rho_i$ for an ith class voice connection. Both the ON and OFF intervals are assumed to be exponentially distributed. In order to satisfy its strict jitter (delay) constraint, a constant transmission rate equal to ρ_i is assigned to each voice connection during its active intervals. Such a rate allocation results in packet queueing delay (and jitter) bounded by the frame length of the system, T_{f}. If the buffer size B_i is equal to $T_{\mathrm{f}}\rho_i$, there is no packet loss due to buffer overflow, i.e., $p_i^{\mathrm{b}} = 0$. Therefore, this rate allocation is based on the condition $T_{\mathrm{f}} \leq D_i$, for all $1 \leq i \leq K_1$. Since the typical delay tolerance for a voice connection is about 30 ms, while the value of T_{f} is no more than 10 ms, the delay, and, therefore, the jitter constraint of the voice connection can be satisfied. During the silent intervals of the voice connection, the allocated resources can be utilized for best effort data transmissions.

For a video connection, a limited buffer size is needed due to its bursty packet generation process and stringent delay requirement. If the buffer size B_i is chosen to be $D_i V_i$, where V_i is the constant packet transmission rate assigned to the video connection, then the maximum experienced packet transmission delay is less than D_i for that connection. Let each video connection be approximately represented by a superposition of N_i ON-OFF minisources, where i is the class index. The effective transmission rate [27] V_i for the video connection is given by:

$$V_i = A_i N_i \left[\frac{1 - z_i}{2} + \sqrt{\left(\frac{1 - z_i}{2}\right)^2 + z_i \eta_i} \right], \tag{1}$$

where $\eta_i = \frac{b_i}{a_i + b_i}$ is the activity factor of the ON-OFF source, and $z_i = \frac{a_i B_i}{A_i (1 - \eta_i) \ln(1/p_i^{\mathrm{b}})}$. Allocating a transmission rate V_i according to (1) can guarantee that with the buffer size of B_i, the loss rate due to buffer overflow is less than p_i^{b}.

The assigned transmission rate for each Type I data connection is at least ρ_i. It is assumed that a Type I data connection can tolerate any large delay, and its buffer size is infinite, then the loss rate due to buffer overflow, p_i^{b}, for the connection is zero.

In general, let $m_i R_{\mathrm{b}}$ be the required transmission rate for an ith class connection in order to achieve its required QoS. Then, m_i is given by

$$m_i = \begin{cases} \lceil \rho_i / R_{\mathrm{b}} \rceil, & \text{when } i \leq K_1 \ \text{(voice)} \\ \lceil V_i / R_{\mathrm{b}} \rceil, & \text{when } K_1 < i \leq K_1 + K_2 \ \text{(video)} \\ \lceil \rho_i / R_{\mathrm{b}} \rceil, & \text{when } K_1 + K_2 < i \leq K \ \text{(Type I data)} \\ 0, & \text{when } i = K + 1 \ \text{(Type II data)}. \end{cases} \tag{2}$$

The ceiling function $\lceil \cdot \rceil$ is required because the actual transmission rate in an MC-CDMA system is an integral multiple of R_{b}. The buffer size for each voice connection is $B_i = T_{\mathrm{f}} m_i R_{\mathrm{b}}$, $i \leq K_1$, and for each video connection is $B_i = D_i m_i R_{\mathrm{b}}$, $K_1 < i \leq K_1 + K_2$. For each data connection, its actual transmission rate is determined according to the current traffic load through a packet scheduling scheme slot by slot and may be larger than $m_i R_{\mathrm{b}}$.

4 Power Distribution

Let S_i be the required power at the BS receiver input for an ith class connection in order to guarantee the required transmission rate, $m_i R_{\mathrm{b}}$, for the connection. It should be noted that the power distribution considered in this section is for CAC purpose, and the objective is to support as many simultaneous connections as possible so long as their QoS can be guaranteed. The instananeous rate for each data connection may change from one slot to another, and the required power for all connections may be calculated accordingly. This will be discussed in Section 6. In the analysis below, we assume perfect power control, which gives an upper bound on the connection-level performance. As it will be seen from the numerical results, with a reasonably good power control algorithm (e.g., a power control algorithm with standard deviation of power control errors of about 1 dB or less), the connection-level GOS performance is very close to that under perfect power control.

The objective here is to find the minimum S_i so that the required SINR, γ_i^*, can be satisfied for all $i = 1, 2, \ldots, K$ given the allocated transmission rate $m_i R_b$, where m_i is given in (2). Consider an ith class connection in a reference cell 0. The multiple access interference (MAI) experienced by any transmitted packets from the ith class connection is given by

$$MAI_i = \sum_{l=1}^{K} m_l k_l S_l - m_i S_i + Y, \qquad (3)$$

where k_i is the number of connections in the ith class, Y is the experienced MAI from all the transmissions in other cells, and can be expressed by

$$Y = \sum_{b=1}^{N_b - 1} \sum_{l=1}^{K} \sum_{j=1}^{k_{bl}} m_{blj} S_{blj} \left(\frac{d_{blj}}{d_{blj0}}\right)^{\alpha} e^{\beta(X_b - X_0)}, \qquad (4)$$

where N_b is the total number of radio cells, k_{bl} is the number of lth class connections in cell b, S_{blj} is the target receive power for the jth connection of the lth class in cell b, d_{blj} is the distance between the BS in cell b and the MS carrying the jth connection of the lth class in cell b, d_{blj0} is the distance between the MS and the BS in the reference cell, α is the path loss exponent, and $\beta = \ln 10/10$ is a constant. Typical values of α, usually obtained by measurement, are between 2.7 and 5. Both X_b and X_0 are Gaussian distributed random variables representing the channel shadowing effect in cell b and the reference cell, respectively, and are assumed to be independent of each other. The variance of X_b and X_0 is $\mathrm{Var}[X_b] = \mathrm{Var}[X_0] = \sigma_x^2$. The values of Y are the same for all connections in the same cell. Because Y is a sum of several log-normally distributed random variables, it can be approximated as a log-normally distributed random variable [25]. Let $Y = e^y$, where y is a normally distributed random variable with mean value μ_y and standard deviation σ_y. When the traffic load is uniformly distributed in all the neighboring cells, and there are many connections in the system, the values of μ_y and σ_y can be considered as approximately the same for all connections in the system. The values of μ_y and σ_y can be found [25], given the information about the MS locations, transmission rate, and power level of each connection. Alternatively, the MSC can obtain the measurements of Y for the connections through the BSs. Based on the measurements, μ_y and σ_y can be estimated. The instant measurements of Y can be used for packet scheduling in order to make a good use of the current channel conditions, and the values of μ_y and σ_y are used for CAC in order to guarantee a long term QoS for each connection.

For a rectangular pulse shaped baseband signal, the SINR at the BS receiver output for any one of the m_i packets can be expressed as

$$\gamma_i = \frac{3GS_i}{2\left(\sum_{l=1}^{K} m_l k_l S_l - m_i S_i + Y\right) + 3GN_0 B_b}, \qquad (5)$$

where N_0 is the one-sided power spectral density (PSD) of the background additive white Gaussian noise (AWGN), $G = B_{ss}/B_b$ is the spread spectrum processing gain, B_b is the baseband bandwidth, and B_{ss} is the spread spectrum bandwidth [26]. In order to achieve successful transmissions with high probability, the following condition must hold:

$$\Pr\{\gamma_i < \gamma_i^*\} \le p^{\mathrm{y}} \qquad (6)$$

for $i = 1, 2, \ldots, K$, where p^{y} is the maximum outage probability due to random MAI. When equality holds in (6), the required receive power for each connection is minimized, and the number of simultaneous connections can be maximized. Then, the following relationship is satisfied after replacing γ_i by the right-hand side of (5) and manipulating:

$$\Pr\left\{e^y > \left(\frac{3G}{2\gamma_i^*} + m_i\right) S_i - \sum_{l=1}^{K} m_l k_l S_l - \frac{3}{2} G N_0 B_{\mathrm{b}}\right\} = p^{\mathrm{y}} \qquad (7)$$

for $i = 1, 2, \ldots, K$. Acceptable communication quality requires that $p^{\mathrm{y}} < 0.5$. Then, (7) can be further written as

$$\left(\frac{3G}{2\gamma_i^*} + m_i\right) S_i - \sum_{l=1}^{K} m_l k_l S_l = \frac{3}{2} G N_0 B_{\mathrm{b}} + e^{\mu_y + \sigma_y \mathrm{Q}^{-1}(p^{\mathrm{y}})} \qquad (8)$$

for $i = 1, 2, \ldots, K$, where $\mathrm{Q}^{-1}(x)$ is the inverse function of $\mathrm{Q}(x)$, and $\mathrm{Q}(x) = \frac{1}{2\pi}\int_x^{\infty} e^{-\frac{t^2}{2}} dt$ for $x \ge 0$. Solving this system of linear equations yields

$$S_i = \frac{3GN_0 B_{\mathrm{b}} \gamma_i^* + 2\gamma_i^* e^{\mu_y + \sigma_y \mathrm{Q}^{-1}(p^{\mathrm{y}})}}{\left(3G + 2m_i \gamma_i^*\right)\left(1 - \sum_{l=1}^{K} \frac{2m_l k_l \gamma_l^*}{3G + 2m_l \gamma_l^*}\right)}. \qquad (9)$$

It is seen from (9) that the target receive power for a connection is determined by its required transmission rate and SINR, and the transmission rates and SINR values of other connections in the system.

5 Connection Admission Control (CAC)

The formulation of CAC is based on the rate allocation and power distribution. The decision to accept a connection (new or handoff) can be made if, after the admission, guaranteed QoS for the particular connection and all the existing connections can be satisfied. The CAC is for voice, video, and Type I data connections only. There is no admission control for Type II data connections since only best effort services are required.

In order to achieve guaranteed QoS, the required packet transmission rate and target receive power are $m_i R_{\mathrm{b}}$ and S_i, respectively, for an ith class connection, where m_i and S_i are given by (2) and (9), respectively. The required transmission power at the MS carrying the jth connection in the ith class can be calculated as $P_{ij} = S_i d_{ij}^{\alpha} e^{\beta X_0}$, where d_{ij} is the distance between the MS and the BS. Because of the random movement of the MS, d_{ij} is a random variable. Communication outage occurs when P_{ij} is larger than the maximum transmission power limit of the MS. In order to guarantee the required communication outage probability, p^{o}, during the entire duration of the connection, the following condition should be satisfied:

$$\Pr\left\{P_{ij} > P^{\mathrm{m}}\right\} = \Pr\left\{S_i d_{ij}^{\alpha} e^{\beta X_0} > P^{\mathrm{m}}\right\} \le p^{\mathrm{o}}, \qquad (10)$$

where P^{m} is the maximum transmission power for the MS. Because d_{ij} and X_0 are independent, the outage probability can be further rewritten as

$$\Pr\{P_{ij} > P^{\mathrm{m}}\} = \int_0^{d^{\mathrm{m}}} f_d(d_{ij}) dd_{ij} \int_{\frac{1}{\beta}\left[\ln(P^{\mathrm{m}}) - \ln(S_i) - \alpha \ln(d_{ij})\right]}^{\infty} f_X(x) dx, \tag{11}$$

where d^{m} is the maximum distance between the MS and its serving BS, $f_d(d_{ij})$ and $f_X(x)$ are the probability density functions (pdf's) of d_{ij} and X_0, respectively, $f_d(d_{ij})$ is determined by the MS's movement pattern, and

$$f_X(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2\sigma_x^2}}.$$

For a CDMA system supporting K traffic classes (voice, video, and Type I data), an admissible state, $\mathbf{k}^{\mathrm{a}}$, which is a combination of the numbers of connections in each traffic class that can coexist in the system with guaranteed QoS, and can be obtained based on the following calculations:

$$\begin{array}{ll} & \mathbf{k}^{\mathrm{a}} = (k_1^{\mathrm{a}}, k_2^{\mathrm{a}}, \ldots, k_K^{\mathrm{a}}), \quad k_i^{\mathrm{a}} \geq 0, \quad i = 1, 2, \ldots, K \\ \text{s.t.} & S_i|_{(k_1, k_2 \ldots, k_K) = (k_1^{\mathrm{a}}, k_2^{\mathrm{a}}, \ldots, k_K^{\mathrm{a}})} > 0, \\ \text{and} & \Pr\{P_{ij} > P^{\mathrm{m}}\} \leq p^{\mathrm{o}} \\ \text{for all} & i = 1, 2, \ldots, K, \quad \text{and} \quad \mathrm{j} = 1, 2, \ldots, \mathrm{k_i}, \end{array} \tag{12}$$

where S_i is calculated from (9), which is the required power level for the allocated rate $m_i R_b$ in (2), and $\Pr\{P_{ij} > P^{\mathrm{m}}\}$ is found from (11). Therefore, a CAC scheme based on (12) takes the information of both the power distribution and transmission rate allocation into consideration. The admissible region of the system, $\mathbf{K}^{\mathrm{a}}$, is the set of all possible $\mathbf{k}^{\mathrm{a}}$, i.e., $\mathbf{K}^{\mathrm{a}} = \{\mathbf{k}^{\mathrm{a}}\}$.

To achieve a lower HCDP compared to NCBP, an appropriate amount of system resources is reserved for potential handoff requests. Effective and efficient resource reservation for potential handoff connections can be achieved by making use of user mobility information. User mobility information is defined as the probability that a mobile user may reside in a particular cell at future moments, which is a function of the distance between the MS's current location and the cell boundary, speed, and direction. Intuitively, to keep the required HCDP, the more likely the users will handoff to a particular cell, the more resources should be reserved in that cell. Since user mobility information changes with time, the resource reservation process should be updated periodically based on the current information. To achieve this, system time is divided into equal length intervals beginning at $t = 0, \tau, 2\tau, \ldots$. A smaller value of τ leads to frequent updating, while a larger value of τ affects the accuracy of the reservation. The value of τ is chosen so that the probability of more than one handoff event for any MS in any interval of length τ is negligible. At the beginning of each interval, the handoff probabilities of each MS from its currently serving cell to its neighboring cells are assumed to be known to the MS's serving BS [28], [29], and all the neighboring cells can exchange the probabilistic information with each other. Without loss of generality, let cell 0 be the reference cell. Define p_{b0u} as the probability that MS u will handoff from its current cell b to cell 0 in the next time interval, then the accumulated handoff probability of all the MSs carrying an ith class connection from the neighboring cells to cell 0 in the next time interval is given by

$$h_i^{\tau} = \sum_{b \in B_0} \sum_{u \in \chi_{b,i}} p_{b0u}, \tag{13}$$

where B_0 is the set of the neighboring cells of cell 0, and $\chi_{b,i}$ is the set of the MSs carrying an ith class connection in cell b. Define $k_i^{\mathrm{h}} = \phi_i h_i^{\tau}$ as the equivalent number of potential handoff connections in the ith class. The amount of resources required by k_i^{h} ith class connections, $i = 1, 2, \ldots, K$, is reserved in cell 0 for the potential handoff connections. The coefficient ϕ_i is used to adjust the amount of reserved resources so as to achieve different HCDP values.

When a new or handoff connection request is received in cell 0 and it belongs to the ξth traffic class, the connection is first allocated to a guaranteed packet transmission rate $m_\xi R_{\mathrm{b}}$ according to (2). Define

$$\tilde{\mathbf{k}}^{\mathrm{a}} = \begin{cases} (k_1^{\mathrm{e}} + k_1^{\mathrm{h}}, \ldots, k_\xi^{\mathrm{e}} + k_\xi^{\mathrm{h}} + 1, \ldots, k_K^{\mathrm{e}} + k_K^{\mathrm{h}}), & \text{for a new connection admission} \\ (k_1^{\mathrm{e}}, \ldots, k_\xi^{\mathrm{e}} + 1, \ldots, k_K^{\mathrm{e}}), & \text{for a handoff connection admission,} \end{cases} \tag{14}$$

where k_i^{e} is the number of existing ith class connections in cell 0 when the new or handoff connection requests for admission. According to (12), if $\tilde{\mathbf{k}}^{\mathrm{a}} \in \mathbf{K}^{\mathrm{a}}$, the connection is accepted. Otherwise, the connection is rejected. From this process, it can be seen that the amount of resources required by k_i^{h} ith class connections, $i = 1, 2, \ldots, K$, is reserved for handoff connections, and cannot be used by new connections. In this way, the CAC gives a handoff connection a higher priority than a new connection.

6 Packet Scheduling

A constant packet transmission rate is allocated to each admitted voice connection during its active intervals and each video connection, and the rate is given by (2). No packet is transmitted for a voice connection during its silent intervals. Before each uplink packet transmission slot, the packet scheduler assigns a transmission rate to each data connection based on the current transmission requests, MSs' locations, and the current channel conditions. An appropriate amount of power should be distributed to each active connection in the system to overcome the MAI and achieve its required SINR. The packet scheduler should also efficiently utilize the system resources in each time-slot by improving the packet throughput of the data traffic.

Consider cell 0 as the reference cell. Let $\tilde{m}_{ij}$ be the current transmission rate for connection j in class i in the cell. The MAI currently experienced by any transmitted packet from the connection is given by

$$\tilde{MAI}_i = \sum_{l=1}^{K} \sum_{n=1}^{k_l} \tilde{m}_{ln} \tilde{S}_{ln} - \tilde{m}_{ij} \tilde{S}_{ij} + Y, \tag{15}$$

where $\tilde{S}_{ij}$ is the target receive power for the connection at the BS receiver input, and Y is the current intercell MAI given by (4), but can be obtained through measurement. Then, the SINR for the connection is given by

$$\frac{3G\tilde{S}_{ij}}{2\sum_{l=1}^{K+1}\sum_{n=1}^{k_l}\tilde{m}_{ln}\tilde{S}_{ln}-\tilde{m}_{ij}\tilde{S}_{ij}+Y+3GN_0B_{\rm b}}\geq\gamma_i^* \qquad (16)$$

when $\tilde{m}_{ij}>0$. In order to allow more simultaneous packet transmissions, equality in (16) should hold. After some manipulation, we have

$$\left(\frac{3G}{2\gamma_i^*}+\tilde{m}_{ij}\right)\tilde{S}_{ij}-\sum_{l=1}^{K+1}\sum_{n=1}^{k_l}\tilde{m}_{ln}\tilde{S}_{ln}=Y+\frac{3GN_0B_{\rm b}}{2}, \qquad (17)$$

when $\tilde{m}_{ij}>0$, for $i=1,2,\ldots,K+1$ and $j=1,2,\ldots,k_i$. When $\tilde{m}_{ij}>0$, $\tilde{S}_{ij}$ can be found by solving the linear equation system in (17); when $\tilde{m}_{ij}=0$, $\tilde{S}_{ij}=0$.

Let $\tilde{P}_{ij}$ be the transmission power from the MS carrying the *j*th connection in the *i*th class. We have

$$\tilde{P}_{ij}=\tilde{S}_{ij}\tilde{d}_{ij}^{\alpha}e^{\beta\tilde{X}_0}, \qquad (18)$$

for $i=1,2,\ldots,K+1$ and $j=1,2,\ldots,k_i$, where $\tilde{d}_{ij}$ is the distance between the MS's current location and its serving BS, and $e^{\beta\tilde{X}_0}$ represents the current channel shadowing effect. The values of $\tilde{P}_{ij}$ should be less than the MS maximum transmission power.

The packet scheduling scheme is to decide the transmission rate and power for each connection in each time slot based on (17) and (18). All packet transmission requests from the Type I data connections wait in a first-come-first-served (FCFS) queue, indexed by $q=1,\ldots,x^{\rm a}$, and all packet transmission requests from the Type II data connections wait in another FCFS queue, indexed by $q=x^{\rm a}+1,\ldots,x^{\rm a}+x^{\rm u}$, where $x^{\rm a}$ and $x^{\rm u}$ are the total number of packet transmission requests from the Type I and Type II data connections, respectively. Let the *q*th transmission request be from the j_qth connection in the i_qth class. For the *j*th connection in the *i*th class, let $m_{ij}^{\rm r}R_{\rm b}$ be the requested rate, $m_{ij}^{\rm s}R_{\rm b}$ be the scheduled transmission rate, $S_{ij}^{\rm s}$ be the scheduled target receive power at the BS, and $P_{ij}^{\rm s}$ be the scheduled transmission power from the MS. The values of $m_{ij}^{\rm s}$, $S_{ij}^{\rm s}$, and $P_{ij}^{\rm s}$ can be found using Pseudocode 1, where $q'(i,j)$ represents the q'th transmission request which comes from the *j*th connection in the *i*th class.

Pseudocode 1: packet transmission scheduling

```
% Rate allocation for voice connections
for i = 1 : K_1
    for j = 1 : k_i
        if connection is active
            % m̃_ij: current transmission rate
            m̃_ij = m_i;
            m^s_ij = m_i;
        else
            m̃_ij = 0;
            m^s_ij = 0;
        end
    end
end
% Rate allocation for video connections
for i = K_1 + 1 : K_1 + K_2
    for j = 1 : k_i
        m̃_ij = m_i;
        m^s_ij = m_i;
    end
end
% Rate allocation for data connection requests
for q = 1 : x^a + x^u
    m̃_{i_q j_q} = m^r_{i_q j_q}; % Temporary rate for qth request
    Temporary rate allocations for other
      connections;
    GOON=1;
    while(GOON=1)
      TEMP=1;
      Check S̃_ij;
      if TEMP=1
       Check P̃_ij;
      end
      if TEMP=1
       m^s_{i_q j_q} = m^r_{i_q j_q};
       GOON=0;
      else
       m̃_{i_q j_q} = m̃_{i_q j_q} - 1;
         if m̃_{i_q j_q} <= m_{i_q};
           m^s_{i_q j_q} = m_{i_q};      % m_{i_q} is guaranteed by CAC
           GOON=0;
         end
      end
    end
end
% Power allocation for all connections
for i = 1 : K + 1
    for j = 1 : k_i
      if m^s_ij > 0
         S^s_ij = S̃_ij;
         P^s_ij = S̃_ij d^α_ij e^{βX̃_0};
      else
         S^s_ij = 0;
         P^s_ij = 0;
      end
    end
end
% Temporary rate allocations for other
%   connections;
for i = K_1 + K_2 + 1 : K + 1  % Type I and Type II data
    for j = 1 : k_i
      if q'(i,j) < q
      % Connections already scheduled
         m̃_ij = m^s_ij;
      elseif q'(i,j) > q
      % Connections not scheduled
         m̃_ij = m_i;
      end
    end
end
% Check S̃_ij;
for i = 1 : K + 1
    for j = 1 : k_i
      if S̃_ij < 0        % Using (17) to calculate S̃_ij
         TEMP=0;
         break;
      end
```

```
    end
end
% Check P̃_ij;
for i = 1 : K+1
    for j = 1 : k_i
      if P̃_ij > P^m          % Using (18) to calculate P̃_ij
        TEMP=0;
        break;
      end
    end
end
```

7 GOS PERFORMANCE

The GOS performance, in terms of NCBP, HCDP, and resource utilization, is derived in this section. Resource utilization is defined as the average number of connections in each traffic class that the system can accept for given offered traffic load. If the system traffic load and all the mobile users are uniformly distributed in all the cells, and the handoff probabilities of transiting into and out of every cell for each traffic class are the same, then both the traffic characteristics and the GOS performance are the same for different cells. It is then only necessary to consider a typical cell. Let the connection arrival process follow a Poisson distribution with mean rate λ_i, and the connection duration be exponentially distributed with mean $1/\mu_i$ for the ith traffic class. Let $f_{T_i^{\mathrm{I}}}(t)$ be the pdf of the interhandoff time, T_i^{I}, which is the time between two successive handoff events for an ith class connection assuming that the connection duration is infinitely large. To facilitate the analysis, we approximate $f_{T_i^{\mathrm{I}}}(t)$ by an exponential distribution with parameter ν_i, which can be obtained by solving the following equation [33]:

$$\int_0^{\infty}\left[f_{T_i^{\mathrm{I}}}(t)-\nu_i e^{-\nu_i t}\right]dt=0. \tag{19}$$

Let p_i^{H} be the probability that an ith class connection will be handed off before the connection ends, then

$$p_i^{\mathrm{H}}=\int_0^{\infty}\nu_i e^{-\nu_i t}dt\int_t^{\infty}\mu_i e^{-\mu_i t'}dt'=\frac{\nu_i}{\mu_i+\nu_i}. \tag{20}$$

Define p_i^{τ} as the probability that an ith class connection will be handed off to a new cell in the next time interval of length τ. Then, p_i^{τ} is equal to the probability that the connection will be handed off to a new cell in a time interval of length τ assuming it is of infinite duration, times the probability that the remaining time of the connection is larger than τ. According to the memoryless property of an exponential distribution, p_i^{τ} can be calculated as

$$p_i^{\tau}=\int_0^{\tau}\nu_i e^{-\nu_i t}dt\int_{\tau}^{\infty}\mu_i e^{-\mu_i t}dt=(1-e^{-\nu_i\tau})e^{-\mu_i\tau}. \tag{21}$$

The cell dwelling time for the ith class connection is the time that the connection stays in the cell until it is handed off to a new cell or terminated, whichever comes first. Therefore, the cell dwelling time is also exponentially distributed with average $\frac{1}{\mu_i+\nu_i}$. At the connection level, the service system in each cell can be considered as a connection blocking system. The NCBP for the ith traffic class, p_i^{n}, is the probability that the system does not have enough resources to accommodate any additional ith class connection after serving all the existing connections and reserving resources for potential handoff connections. The HCDP for the ith traffic class, p_i^{h}, is the probability that the system does not have enough resources to accommodate any additional ith class connection after serving all the existing connections. The HCDP is the fraction of handoff attempts which are unsuccessful. There is also forced connection termination probability (FCTP) defined as the probability that a connection, after it is admitted into the system, is eventually forced to terminate due to insufficient resources when it is handed off to a new cell. The FCTP, p_i^{f}, for the ith traffic class can be calculated by

$$p_i^{\mathrm{f}}=\sum_{J=1}^{\infty}(p_i^{\mathrm{H}})^J(1-p_i^{\mathrm{h}})^{J-1}p_i^{\mathrm{h}}=\frac{p_i^{\mathrm{H}}p_i^{\mathrm{h}}}{1-p_i^{\mathrm{H}}(1-p_i^{\mathrm{h}})}, \tag{22}$$

where J represents the number of handoffs during the connection's lifetime. The system supporting K traffic classes can be modeled as a K-dimensional Markov chain with transition probabilities given by

$$\begin{cases}\Pr\{\mathbf{k}^{\mathrm{a}}|\mathbf{k}^{\mathrm{a}}-\mathbf{e}_i\}=\lambda_i+\lambda_i^{\mathrm{h}}, & \text{when } \mathbf{k}^{\mathrm{a}}\in\mathbf{K}^{\mathrm{a}}\ \&\ \mathbf{k}^{\mathrm{a}}+\mathbf{k}^{\mathrm{H}}\in\mathbf{K}^{\mathrm{a}}\ \&\ k_i^{\mathrm{a}}>0\\ \Pr\{\mathbf{k}^{\mathrm{a}}|\mathbf{k}^{\mathrm{a}}-\mathbf{e}_i\}=\lambda_i^{\mathrm{h}}, & \text{when } \mathbf{k}^{\mathrm{a}}\in\mathbf{K}^{\mathrm{a}}\ \&\ \mathbf{k}^{\mathrm{a}}+\mathbf{k}^{\mathrm{H}}\notin\mathbf{K}^{\mathrm{a}}\ \&\ k_i^{\mathrm{a}}>0\\ \Pr\{\mathbf{k}^{\mathrm{a}}-\mathbf{e}_i|\mathbf{k}^{\mathrm{a}}\}=k_i^{\mathrm{a}}(\mu_i+\nu_i), & \text{when } \mathbf{k}^{\mathrm{a}}\in\mathbf{K}^{\mathrm{a}}\ \&\ k_i^{\mathrm{a}}>0\end{cases} \tag{23}$$

for all $i=1,2,\ldots,K$, where $\mathbf{e}_i$ is a vector with the ith component equal to 1 and all the other $(K-1)$ components equal to zero, λ_i^{h} is the average handoff connection arrival rate for the ith traffic class which can be calculated as follows according to [31]:

$$\lambda_i^{\mathrm{h}}=\frac{\nu_i(1-p_i^{\mathrm{n}})}{\mu_i+\nu_i p_i^{\mathrm{f}}}\lambda_i, \tag{24}$$

and $\mathbf{k}^{\mathrm{H}}$ is a vector representing the number of equivalent handoff connections in the next time interval of length τ and is given by

$$\begin{aligned}\mathbf{k}^{\mathrm{H}}&=(\phi_1 h_1^{\tau},\phi_2 h_2^{\tau},\ldots,\phi_K h_K^{\tau})\\&=(\phi_1 p_1^{\tau}k_1^{\mathrm{a}},\phi_2 p_2^{\tau}k_2^{\mathrm{a}},\ldots,\phi_K p_K^{\tau}k_K^{\mathrm{a}}).\end{aligned} \tag{25}$$

If $\pi(\mathbf{k}^{\mathrm{a}})$ is the equilibrium probability for state $\mathbf{k}^{\mathrm{a}}$, then we have $\sum_{\mathbf{k}^{\mathrm{a}}\in\mathbf{K}^{\mathrm{a}}}\pi(\mathbf{k}^{\mathrm{a}})=\mathbf{1}$. Together with (23), the equilibrium probability for the admissible state $\mathbf{k}^{\mathrm{a}}=(k_1^{\mathrm{a}},k_2^{\mathrm{a}},\ldots,k_K^{\mathrm{a}})$ is given by [30]:

$$\pi(\mathbf{k}^{\mathrm{a}})=\pi(\mathbf{0})\prod_{i=1}^{K}\frac{[\varrho_i(\mathbf{k}^{\mathrm{a}})]^{k_i^{\mathrm{a}}}}{k_i^{\mathrm{a}}!}, \tag{26}$$

where

$$\pi(\mathbf{0})=\left\{\sum_{\mathbf{k}^{\mathrm{a}}\in\mathbf{K}^{\mathrm{a}}}\prod_{i=1}^{K}\frac{[\varrho_i(\mathbf{k}^{\mathrm{a}})]^{k_i^{\mathrm{a}}}}{k_i^{\mathrm{a}}!}\right\}^{-1}, \tag{27}$$

$\varrho_i(\mathbf{k}^{\mathrm{a}})$ is given by

$$\varrho_i(\mathbf{k}^{\mathrm{a}})=\begin{cases}\frac{(\lambda_i+\lambda_i^{\mathrm{h}})}{\mu_i+\nu_i}, & \text{when } \mathbf{k}^{\mathrm{a}}\in\mathbf{K}^{\mathrm{a}}\ \&\ \mathbf{k}^{\mathrm{a}}+\mathbf{k}^{\mathrm{H}}\in\mathbf{K}^{\mathrm{a}}\\ \frac{\lambda_i^{\mathrm{h}}}{\mu_i+\nu_i}, & \text{when } \mathbf{k}^{\mathrm{a}}\in\mathbf{K}^{\mathrm{a}}\ \&\ \mathbf{k}^{\mathrm{a}}+\mathbf{k}^{\mathrm{H}}\notin\mathbf{K}^{\mathrm{a}}\end{cases} \tag{28}$$

for all $i=1,2,\ldots,K$.

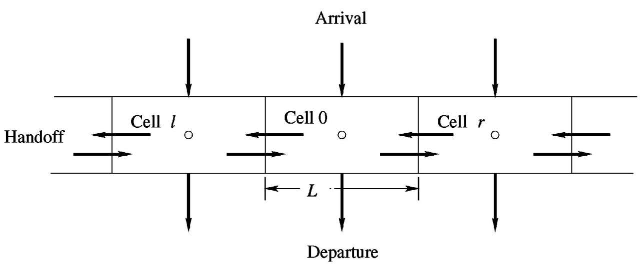


Fig. 3. 1D cell array.

The NCBP for the ith class is given by

$$p_i^{\mathrm{n}} = 1 - \sum_{\mathbf{k}^{\mathrm{a}} \in \mathbf{g}_i^{\mathrm{n}}} \pi(\mathbf{k}^{\mathrm{a}}) = 1 - \pi(\mathbf{0}) \sum_{\mathbf{k}^{\mathrm{a}} \in \mathbf{g}_i^{\mathrm{n}}} \prod_{i=1}^{K} \frac{[\varrho_i(\mathbf{k}^{\mathrm{a}})]^{k_i^{\mathrm{a}}}}{k_i^{\mathrm{a}}!}, \tag{29}$$

where $\mathbf{g}_i^{\mathrm{n}} = \{\mathbf{k}^{\mathrm{a}} | \mathbf{k}^{\mathrm{a}} \in \mathbf{K}^{\mathrm{a}}, \mathbf{k}^{\mathrm{a}} + \mathbf{k}^{\mathrm{H}} + \mathbf{e}_i \in \mathbf{K}^{\mathrm{a}}\}$. The HCDP for the ith traffic class is given by

$$p_i^{\mathrm{h}} = 1 - \sum_{\mathbf{k}^{\mathrm{a}} \in \mathbf{g}_i^{\mathrm{h}}} \pi(\mathbf{k}^{\mathrm{a}}) = 1 - \pi(\mathbf{0}) \sum_{\mathbf{k}^{\mathrm{a}} \in \mathbf{g}_i^{\mathrm{h}}} \prod_{i=1}^{K} \frac{[\varrho_i(\mathbf{k}^{\mathrm{a}})]^{k_i^{\mathrm{a}}}}{k_i^{\mathrm{a}}!}, \tag{30}$$

where $\mathbf{g}_i^{\mathrm{h}} = \{\mathbf{k}^{\mathrm{a}} | \mathbf{k}^{\mathrm{a}} \in \mathbf{K}^{\mathrm{a}}, \mathbf{k}^{\mathrm{a}} + \mathbf{e}_i \in \mathbf{K}^{\mathrm{a}}\}$. The resource utilization for the ith traffic class is given by

$$C_i = \sum_{k_i} k_i \pi_i(k_i), \tag{31}$$

where $\pi_i(k_i)$ is the probability that the number of the ith class connections in the cell is k_i and is equal to the ith marginal probability of the admissible state equilibrium probability when $k_i^{\mathrm{a}} = k_i$, i.e.,

$$\pi_i(k_i) = \sum_{\mathbf{k}^{\mathrm{a}} \in \mathbf{K}^{\mathrm{a}}, k_i^{\mathrm{a}} = k_i} \pi(\mathbf{k}^{\mathrm{a}}). \tag{32}$$

In general, accurate calculation of p_i^{n}, p_i^{h}, C_i, and $\pi(\mathbf{k}^{\mathrm{a}})$ is difficult because of the coupling between λ_i^{h}, p_i^{h}, p_i^{n}, and $\pi(\mathbf{k}^{\mathrm{a}})$. Numerical methods, such as the bisection method [35], can be used to find p_i^{n} from (26), (27), (28), (29), (30), and (31). In a practical system, p_i^{h} should be very small, e.g., $p_i^{\mathrm{h}} < 10^{-3}$. When this is true, the FCTP, p_i^{f}, is also very small and its effect on λ_i^{h} can be neglected. Then, (24) can be approximated as $\lambda_i^{\mathrm{h}} \approx \frac{\nu_i}{\mu_i}(1 - p_i^{\mathrm{n}})\lambda_i$, which can simplify the calculation.

8 Numerical Results

Without loss of generality, we consider a one-dimensional (1D) cellular array as shown in Fig. 3. All the cells in the network are arranged as a circle to avoid the boundary effect. At the time when an MS originates a new connection request, it can be in any one of the cells with equal probability. The initial location of an MS is uniformly distributed in the 1D region of its cell. The mobile velocity v is uniformly distributed between a minimum rate, $v_{\min}$, and a maximum rate, $v_{\max}$. The movement direction can be either one of the two directions with equal probability. When an MS reaches the boundary of two cells, it randomly chooses a new direction and velocity. The probability distribution function, $F_{T_i^{\mathrm{I}}}(t)$, of the connection interhandoff time can be found as follows:

$$F_{T_i^{\mathrm{I}}}(t) = \Pr\{T_i^{\mathrm{I}} \le t\} = \Pr\{v \ge L/t\}$$
$$= \begin{cases} 1, & \text{when } t \ge L/v_{\min} \\ \frac{1}{v_{\max} - v_{\min}}(v_{\max} - \frac{L}{t}), & \text{when } L/v_{\max} \le t < L/v_{\min} \\ 0, & \text{when } t < L/v_{\max}, \end{cases} \tag{33}$$

where L is the cell size. Then, the pdf of T_i^{I} is given by

$$f_{T_i^{\mathrm{I}}}(t) = \begin{cases} \frac{L}{(v_{\max} - v_{\min})t^2}, & \text{when } L/v_{\max} < t \le L/v_{\min} \\ 0, & \text{otherwise.} \end{cases} \tag{34}$$

The handoff probabilities of MS u from its current cell 0 to its two neighboring cells, cell r and cell l, are defined as $p_{0ru} = L'/L$ and $p_{0lu} = 1 - L'/L$, respectively, where L' is the distance between the current location of MS u and the left edge of cell 0. The parameters for the analysis and simulation are as follows: There are five cells in the system, each with cell size $L = 1{,}000$ m. The values of $v_{\min}$ and $v_{\max}$ are 15 km/h and 60 km/h, respectively. The spread spectrum bandwidth is $B_{\mathrm{ss}} = 5$ MHz. The basic MC-CDMA rate is $R_{\mathrm{b}} = 16$ kbps, and the spread spectrum processing gain is $G = 256$. The frame length is 10 ms. The one-sided AWGN PSD is $N_0 = 10^{-10}$ W/Hz. The maximum transmission power for each MS is 0.1 W. Both the communication outage probabilities are $p^{\mathrm{o}} = p^{\mathrm{y}} = 1\%$. The path loss exponent is $\alpha = 4.0$, and the standard deviation of X_b due to channel shadowing is $\sigma_x = 8$ dB.

There are three traffic classes: voice, video, and Type I data. The average connection duration is 200 s for each class. Convolutional coding with rate $\frac{1}{3}$ and constraint length 9 is used as the channel coding for voice traffic. Concatenated convolutional coding (inner coding) and Reed-Solomon (RS) coding (outer coding) are used as the channel coding for both video and data traffic, where the convolutional code has rate $\frac{1}{2}$ and constraint length 9, and the RS code is a $(16, 12, 5)$ code, i.e., each symbol contains 5 bits, each codeword includes 16 symbols and 12 of them are information symbols. The bit rate in the ON state from the voice encoder (G.723 standard encoder) is 5.3 kbps, which is then input to the channel encoder for voice traffic. The output bit rate from the channel encoder for a voice connection is 16 kbps. The activity factor for a voice connection is $\eta_i = \frac{3}{8}$. The video connections are low bit rate for video phone (H.263 encoded). Each video connection is modeled by 14 ON-OFF minisources, each minisource with parameters $A_i = 4.8$ kbps, $a_i = 0.8$, and $b_i = 0.2$. The effective transmission rate for each video connection is 60 kbps before channel coding, and 160 kbps after channel coding. The minimum required transmission rate for each data connection is 48 kbps before channel coding, and 128 kbps after channel coding. The maximum tolerable packet delays for a voice and video connection are 30 ms and 80 ms, respectively. The maximum tolerable BERs for a voice, video and data connection are 10^{-3}, 10^{-5}, and 10^{-7}, respectively. The loss rate due to buffer overflow, p_i^{b}, for a video connection is 2×10^{-6}. The resource reservation

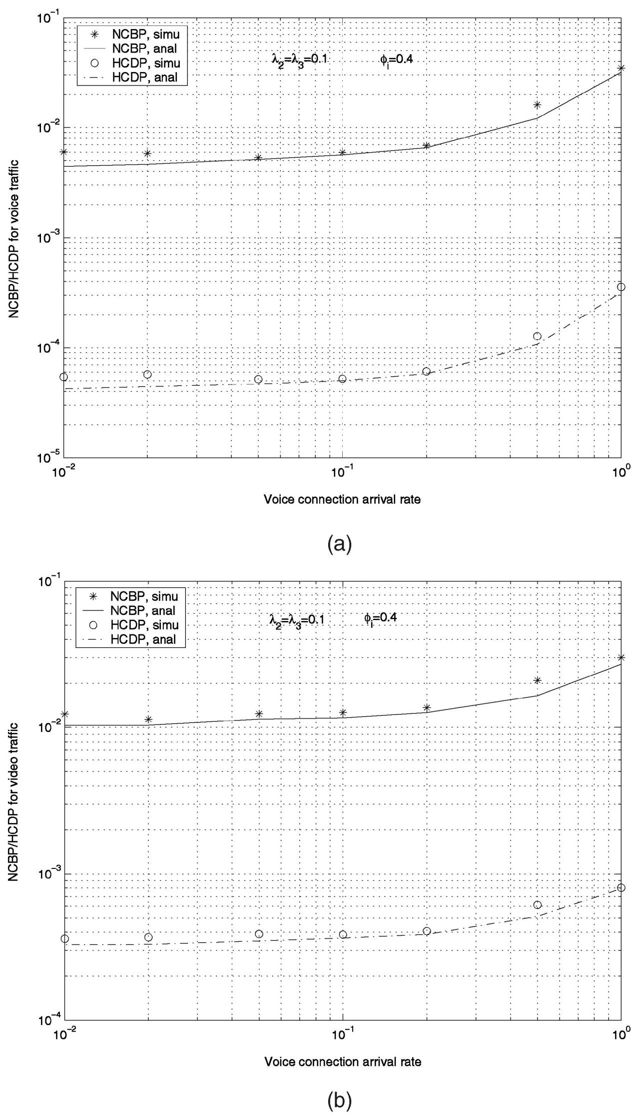


Fig. 4. NCBP/HCDP versus voice connection arrival rate: (a) voice traffic and (b) video traffic.

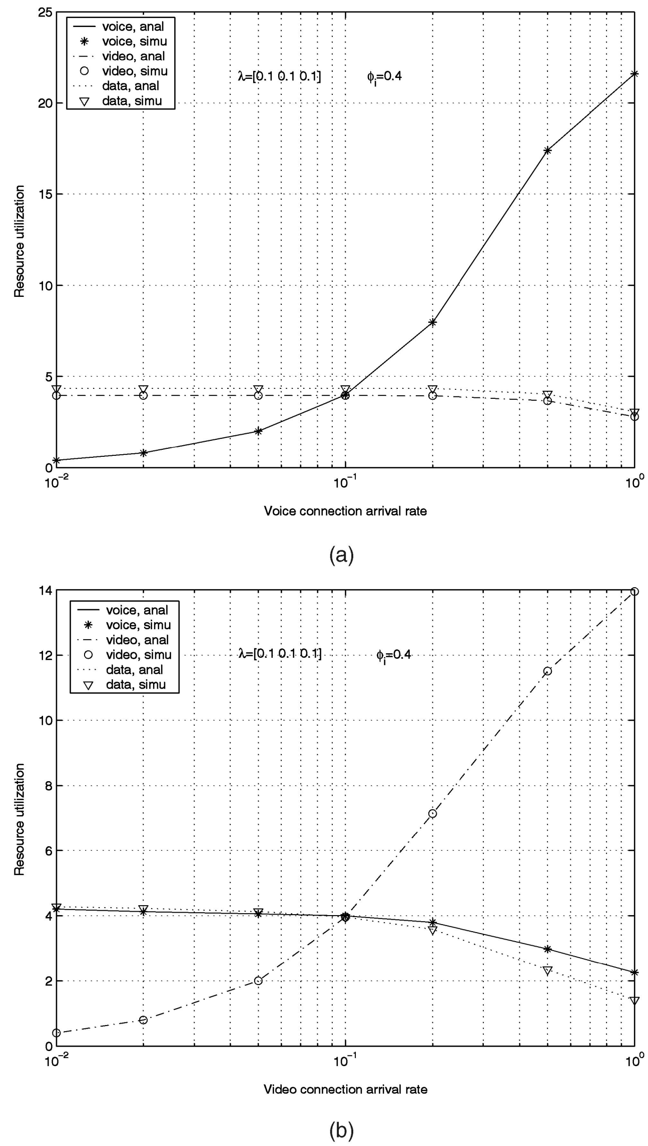
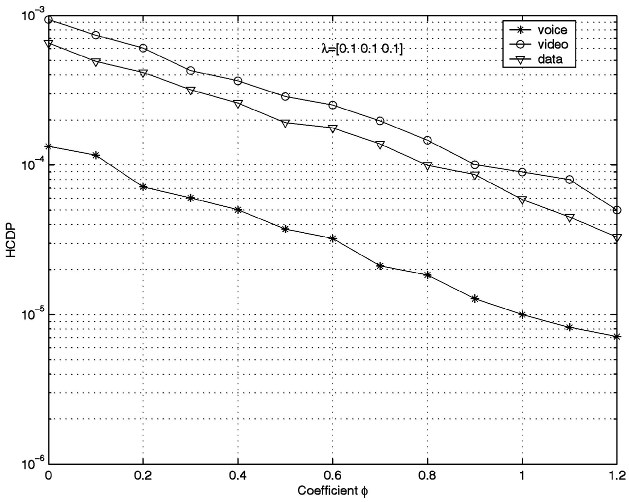


Fig. 5. Resource utilization versus (a) voice connection arrival rate and (b) video connection arrival rate.

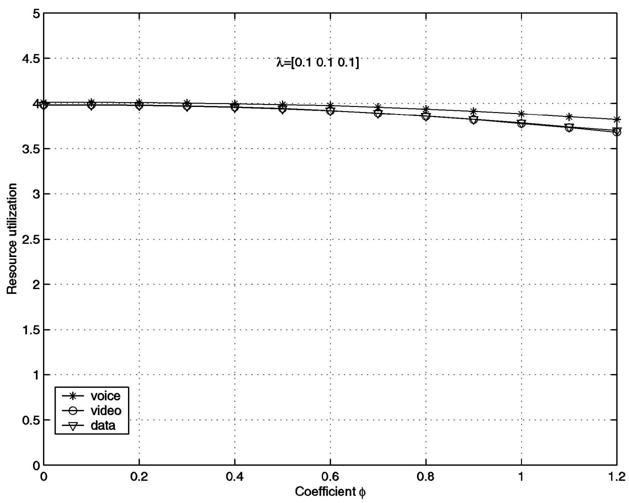
coefficients are the same for different traffic classes, i.e., $\phi_1 = \phi_2 = \phi_3 = \phi$.

Simulation and analytical results of NCBP/HCDP's for voice and video traffic as a function of voice connection arrival rate are shown in Fig. 4. Throughout the numerical results, the unit of the average connection arrival rate is the number of connections per second. It can be seen that the analytical results match the simulation results quite well. Fig. 4 shows that NCBP/HCDPs for both voice and video traffic classes increase as the voice connection arrival rate increases. This demonstrates the resource sharing effect in CDMA systems: increasing traffic load in one class increases NCBP/HCDP of other traffic classes in the system. From Fig. 4, it can be seen that voice traffic receives better NCBP and HCDP performance than video traffic. This is due to the complete resource sharing policy used and the fact that a voice connection requires much less resources than a video connection. The unfairness can be overcome by using a partition-based approach, i.e., a certain amount of system resources is dedicated to serve a particular traffic class. However, the partition-based approach may reduce resource utilization compared with the complete sharing approach [32].

Resource utilization for each traffic class is shown in Fig. 5. A very good match is achieved between the simulation and analytical results. Fig. 5a shows that, as the voice connection arrival rate increases, resource utilization for voice traffic increases significantly, while that for video and data traffic decreases very slightly. Fig. 5b shows that, as the video connection arrival rate increases, resource utilization for video traffic increases significantly, and that for voice and data traffic decreases slightly. Fig. 5 shows that video traffic has a greater impact on the performance of each traffic class than voice traffic. The reason is that a video connection requires much more system resources, including transmission rate and power, than a voice connection. Similarly, data traffic also has a greater impact on the GOS

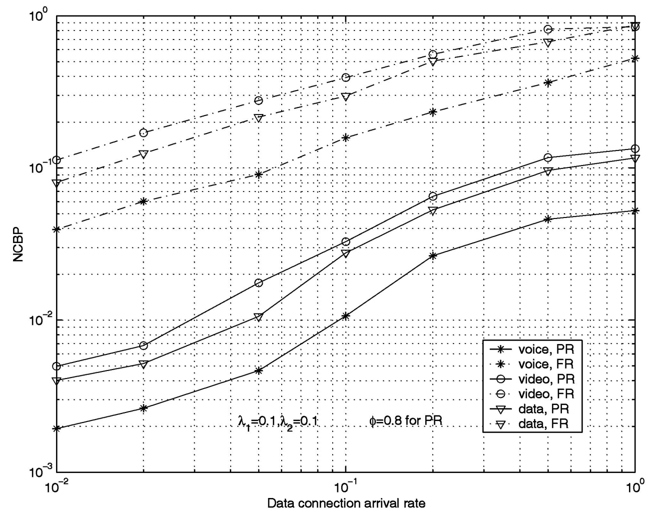


(a)

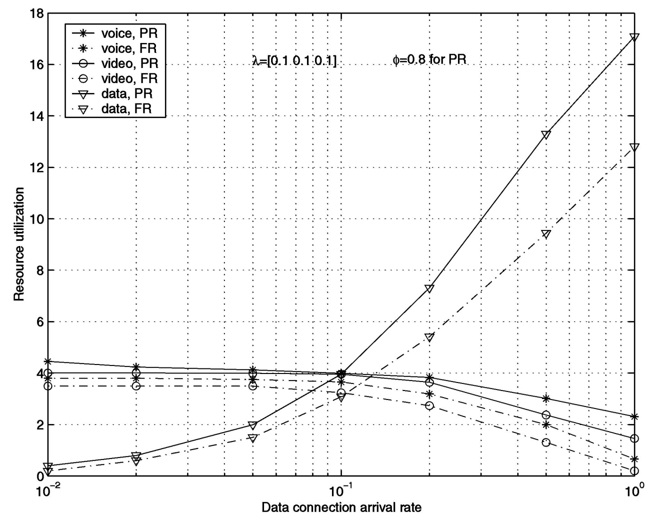


(b)

Fig. 6. Effect of coefficient ϕ_i: (a) HCDP and (b) resource utilization.



(a)



(b)

Fig. 7. Comparison between proposed RRM and FR RRM: (a) NCBP and (b) resource utilization.

performance than voice traffic because of the higher transmission rate and low BER requirements of a data connection.

Fig. 6 shows the effect of coefficient ϕ_i on the performance of the proposed RRM. When $\phi = 0$, no resource is reserved and the HCDP equals the NCBP. Since the amount of reserved resources is based on the handoff probabilities, and the actual handoff events occur randomly, the reserved resources cannot absolutely guarantee zero HCDP. From Fig. 6a, it can be seen that increasing the coefficient can significantly decrease the HCDP for each traffic class since more resources are reserved for handoff connections. It can also be seen from Fig. 6b that the resource utilization only decreases slightly by reserving resources for handoff connections. This demonstrates the effectiveness and the efficiency of the proposed RRM scheme.

Figs. 7 and 8 show the advantage of resource reservation based on user mobility information in the proposed joint RRM scheme. Performance comparison is given between the proposed RRM (PR) scheme and two nonjoint RRM schemes, respectively, which do not make use of the user mobility information. The two nonjoint RRM schemes are full reservation (FR) and no reservation (NR). The difference between the two nonjoint RRM schemes and the PR scheme is that the FR scheme reserves 100 percent of resources for each connection in each of its neighboring cells, while the NR scheme does not reserve resources for handoff connections. From Fig. 7, it can be seen that, compared with the FR scheme, the PR scheme achieves both lower NCBP and higher resource utilization. Although the FR scheme can achieve zero HCDP, it unnecessarily underutilizes the system resources. In contrast, the PR scheme can achieve both low HCDP (shown in Fig. 8a) and relatively high resource utilization. The NR RRM treats handoff connections the same as the new connections and results in the same HCDP as NCBP, which is undesirable, especially at high traffic load. Fig. 8a shows that the PR scheme achieves much lower HCDP compared to the

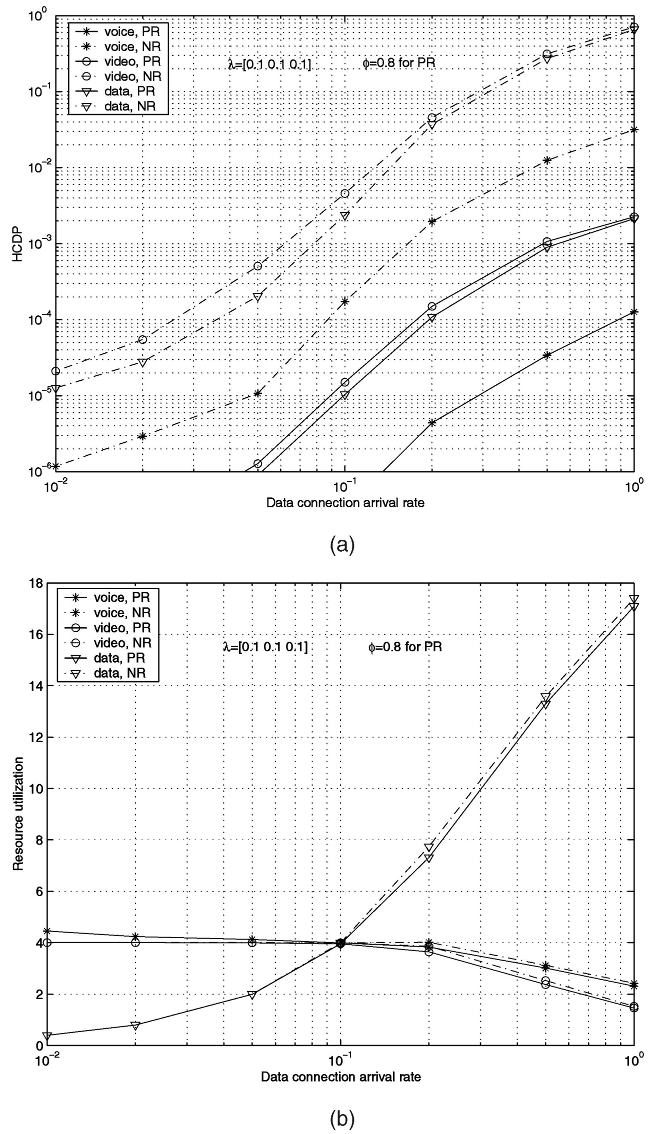


Fig. 8. Comparison between proposed RRM and NR RRM: (a) HCDP and (b) resource Utilization.

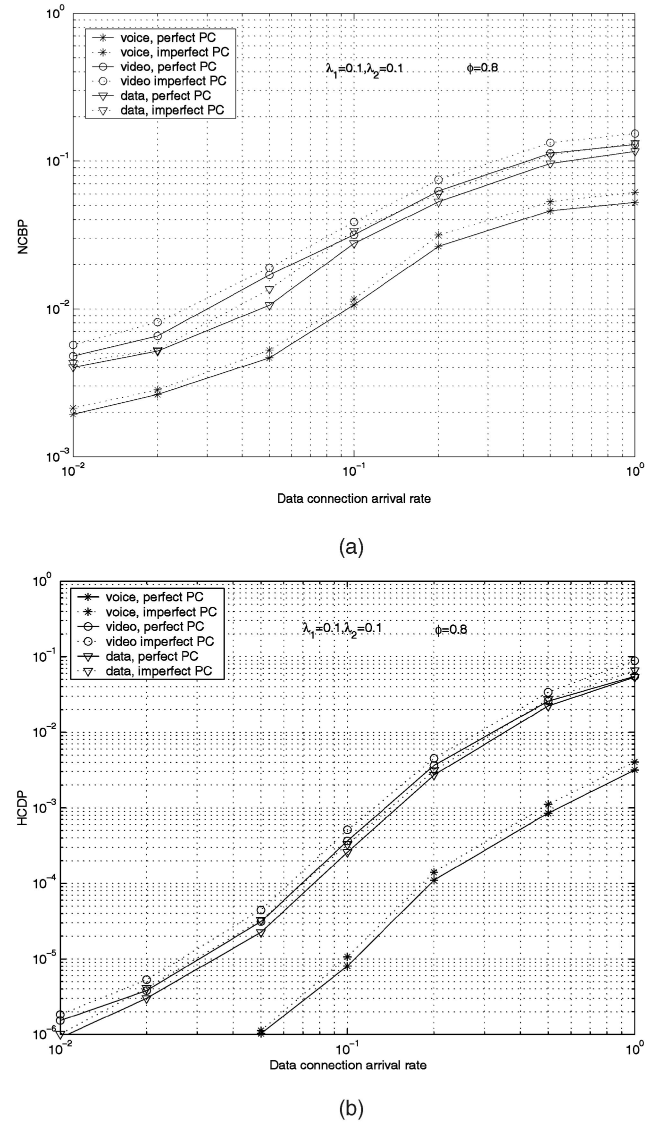


Fig. 9. Effect of imperfect power control: (a) NCBP and (b) HCDP.

NR scheme. The reduced HCDP is achieved by resource reservation. As is shown in Fig. 8b, there is only a slight decrease in the resource utilization of the PR scheme compared with that of the NR scheme. Therefore, the proposed RRM scheme can achieve better trade off between resource utilization and HCDP, and is more desirable than both the FR and NR schemes.

In practice, power control usually cannot completely compensate for the random and fast varying effect of channel fading. It is demonstrated in [34] that, when taking power control errors into consideration, the actual receive power approximately follows a log-normal distribution. Fig. 9 compares the NCBP and HCDP for different traffic classes, where the standard deviation of power control errors is 1 dB [34]. It can be seen that, power control imperfection increases the NCBP and HCDP slightly. The reason for the reduced GOS performance is that, power control imperfection increases the required target receive power in order to achieve the same SINR with the same required communication outage probability. On the other hand, because of the symmetrical distribution of power control errors in dB, the average value of the actual received power is also above the target receive power [24]. Therefore, the GOS performance degradation is not significant.

9 Conclusions

A novel RRM scheme, which jointly considers the effects of the network characteristics from physical layer to network layer, for supporting heterogeneous services has been proposed. The scheme achieves effective and efficient radio resource utilization, which can benefit both the mobile users and the network service providers. Although the proposed RRM is for FDD/CDMA systems, it is equally applicable to TDD/CDMA systems if perfect synchronization can be achieved in the entire system. The proposed RRM can also be extended to support multimedia traffic where different types of traffic may be carried by one MS. In this case, the MS transmission power would limit multiple connections carried by the MS.

ACKNOWLEDGMENTS

This work has been supported by a grant from the Communications and Information Technology Ontario (CITO), Ontario, Canada. The authors would like to thank the anonymous reviewers for their valuable comments which helped to improve the paper presentation.

REFERENCES

[1] S.V. Hanly, "An Algorithm for Combined Cell-Site Selection and Power Control to Maximize Cellular Spread Spectrum Capacity," *IEEE J. Selected Areas Comm.,* vol. 13, pp. 1332-1340, Sept. 1995.

[2] L.C. Yun and D.G. Messerschmitt, "Power Control for Variable QoS on a CDMA Channel," *Proc. IEEE Military Comm. Conf.,* pp. 178-182, Oct. 1994.

[3] A. Sampath, P.S. Kumar, and J.M. Holtman, "Power Control and Resource Management for a Multimedia CDMA Wireless Systems," *Proc. IEEE Int'l Symp. Personal, Indoor, and Mobile Radio Comm.,* pp. 21-25, 1995.

[4] S. Yao and E. Geraniotis, "Optimal Power Control Law for Multimedia Multirate CDMA Systems," *Proc. IEEE Vehicular Technology Conf.,* pp. 392-396, Apr. 1996.

[5] C.-L. I and K.K. Sabnani, "Variable Spreading Gain CDMA with Adaptive Control for True Packet Switching Wireless DS/CDMA Networks," *Proc. IEEE Int'l Conf. Comm.,* pp. 725-730, 1995.

[6] C-L. I and R.D. Gitlin, "Multi-Code CDMA Wireless Personal Communications Networks," *Proc. Int'l Conf. Comm.,* pp. 1060-1064, 1995.

[7] M.J. Karol, Z. Liu, and K.Y. Eng, "Distributed-Queueing Request Update Multiple Access (DQRUMA) for Wireless Packet (ATM) Networks," *Proc. Int'l Conf. Comm.,* pp. 1224-1231, 1995.

[8] J.I. Capetanakis, "Tree Algorithms for Packet Broadcast Channels," *IEEE Trans. Information Theory,* vol. 25, pp. 505-515, Sept. 1979.

[9] Z. Liu, M.J. Karol, M.E. Zarki, and K.Y. Eng, "Channel Access and Interference Issues in Multi-Code DS-CDMA Wireless Packet (ATM) Networks," *Wireless Networks,* vol. 2, pp. 173-193, 1996.

[10] S. Choi and K.G. Shin, "An Uplink CDMA System Architecture with Diverse QoS Guarantees for Heterogeneous Traffic," *IEEE/ACM Trans. Networking,* vol. 7, pp. 616-628, Oct. 1999.

[11] I.E. Akyildiz, D.A. Levine, and I. Joe, "A Slotted CDMA Protocol with BER Scheduling for Wireless Multimedia Networks," *IEEE/ACM Trans. Networking,* vol. 7, pp. 146-158, Apr. 1999.

[12] J.H. Kim, T.S. Kim, Y.W. Kim, and D.K. Sung, "Hybrid Channel Assignment Scheme for Accommodating Voice/Data Traffic in DS-CDMA Cellular Systems," *IEEE Trans. Vehicular Technology,* vol. 49, pp. 1566-1577, Sept. 2000.

[13] L. Song and N.B. Mandayam, "Hierarchical SIR and Rate Control on the Forward Link for CDMA Data Users under Delay and Error Constraints," *IEEE J. Selected Areas Comm.,* vol. 19, pp. 1871-1882, Oct. 2001.

[14] A.C. Kam, T. Minn, and K.-Y. Siu, "Supporting Rate Guarantee and Fair Access for Bursty Data Traffic in W-CDMA," *IEEE J. Selected Areas Comm.,* vol. 19, pp. 2121-2130, Oct. 2001.

[15] S.M. Shin, C.-H. Cho, and D.K. Sung, "Interference-Based Channel Assignment for DS-CDMA Cellular Systems," *IEEE Trans. Vehicular Technology,* vol. 48, pp. 233-239, Jan. 1999.

[16] Z. Liu and M.E. Zarki, "SIR-Based Call Admission Control for DS-CDMA Cellular Systems," *IEEE J. Selected Areas Comm.,* vol. 12, pp. 638-644, May 1994.

[17] Y. Ishikawa and N. Umeda, "Capacity Design and Performance of Call Admission Control in Cellular CDMA Systems," *IEEE J. Selected Areas Comm.,* vol. 15, pp. 1627-1635, Oct. 1997.

[18] M. Xiao, N.B. Shroff, and E.K.P. Chong, "Distributed Admission Control for Power-Controlled Cellular Wireless Systems," *IEEE/ACM Trans. Networking,* vol. 9, pp. 790-800, Dec. 2001.

[19] Y.B. Lin, A. Noerpel, and D. Harasty, "A Nonblocking Channel Assignment Strategy for Hand-Offs," *Proc. IEEE Int'l Conf. Universal Personal Comm.,* pp. 558-562, Sept. 1994.

[20] A.S. Acampora and M. Naghshineh, "An Architecture and Methodology for Mobile-Executed Handoff in Cellular ATM Networks," *IEEE J. Selected Areas Comm.,* vol. 12, pp. 1365-1374, Oct. 1994.

[21] D.A. Levine, I.F. Akyildiz, and M. Naghshineh, "A Resource Estimation and Call Admission Algorithm for Wireless Multimedia Networks Using the Shadow Cluster Concept," *IEEE/ACM Trans. Networking,* vol. 5, pp. 1-12, Feb. 1997.

[22] D. Zhao, X. Shen, and J.W. Mark, "Efficient Call Admission Control for Heterogeneous Services in Wireless Mobile ATM Networks," *IEEE Comm. Magazine,* vol. 38, pp. 72-78, Oct. 2000.

[23] A. Aljadhai and T.F. Znati, "Predictive Mobility Support for QoS Provisioning in Mobile Wireless Environments," *IEEE J. Selected Areas Comm.,* vol. 19, pp. 1915-1930, Oct. 2001.

[24] S. Ariyavisitakul and L.F. Chang, "Signal and Interference Statistics of a CDMA System with Feedback Power Control," *IEEE Trans. Comm.,* vol. 41, no. 11, pp. 1626-1634, Nov. 1993.

[25] J.Y. Kim and G.L. Stuber, "CDMA Soft Handoff Analysis in the Presence of Power Control Error and Shadowing Correlation," *IEEE Trans. Wireless Comm.,* vol. 1, pp. 245-255, Apr. 2002.

[26] A.J. Viterbi, *Principles of Spread Spectrum Communication.* Reading, Mass.: Addison-Wesley, 1995.

[27] M. Schwartz, *Broadband Integrated Networks.* Prentice-Hall, 1998.

[28] X. Shen and J.W. Mark, "Mobility Information for Resource Management in Wireless ATM Networks," *Computer Networks,* vol. 31, pp. 1049-1062, 1999.

[29] T. Liu, P. Bahl, and I. Chlamtac, "Mobility Modeling, Location Tracking, and Trajectory Prediction in Wireless ATM Networks," *IEEE J. Selected Areas Comm.,* vol. 16, pp. 922-936, Aug. 1998.

[30] F. Kelly, "Loss Networks," *The Annals of Applied Probability,* vol. 1, pp. 319-378, 1991.

[31] Y.-B. Lin, S. Mohan, and A. Noerpel, "Queueing Priority Channel Assignment Strategies for PCS Hand-Off and Initial Access," *IEEE Trans. Vehicular Technology,* vol. 43, pp. 704-712, Aug. 1994.

[32] B.M. Epstein and M. Schwartz, "Predictive QoS-Based Admission Control for Multiclass Traffic in Cellular Wireless Networks," *IEEE J. Selected Areas Comm.,* vol. 18, pp. 523-534, Mar. 2000.

[33] D. Hong and S.S. Rappaport, "Traffic Model and Performance Analysis for Cellular Mobile Radio Telephone Systems with Prioritized and Nonprioritized Handoff Procedures," *IEEE Trans. Vehicular Technology,* vol. 35, pp. 77-92, Aug. 1986.

[34] A.J. Viterbi, A.M. Viterbi, and E. Zehavi, "Performance of Power-Controlled Wideband Terrestrial Digital Communication," *IEEE Trans. Comm.,* vol. 41, no. 4, pp. 559-569, Apr. 1993.

[35] S.C. Chapra and R.P. Canale, *Numerical Methods for Engineers,* fourth ed. chapter 5, McGraw-Hill, 2002.

Dongmei Zhao received the PhD degree in electrical and computer engineering from University of Waterloo, Waterloo, Ontario, Canada, in June 2002. Since July 2002, she has been an assistant professor in the Department of Electrical and Computer Engineering at McMaster University, Hamilton, Ontario, Canada. Her research interests include quality of service provisioning and resource management in wireless cellular and ad hoc networks. She is a member of the IEEE.

Xuemin (Sherman) Shen received the BSc degree from Dalian Marine University, China, in 1982 and the MSc and PhD degrees from Rutgers University, New Jersey, in 1987 and 1990, respectively, all in electrical engineering. From September 1990 to September 1993, he was first with Howard University, Washington D.C., and then the University of Alberta, Edmonton, Canada. Since October 1993, he has been with the Department of Electrical and Computer Engineering, University of Waterloo, Canada, where he is a full professor. Dr. Shen's research focuses on mobility and resource management in interconnected wireless/wireline networks, stochastic process, and control. He is a coauthor of two books, and has publications in communications networks, control, and filtering. Dr. Shen received the Premier's Research Excellence Award (PREA) from the Province of Ontario for demonstrated excellence of scientific and academic contributions in 2003, and the Distinguished Performance Award from the Faculty of Engineering, University of Waterloo, for outstanding contribution in teaching, scholarship, and service in 2002. He is a senior member of the IEEE, and a registered professional engineer of Ontario, Canada.

Jon W. Mark (M'62-SM'80-F'88-LF'03) received the BASc degree from the University of Toronto in 1962, and the MEng. and PhD degrees from McMaster University in 1968 and 1970, respectively, all in electrical engineering. From 1962 to 1970, he was an engineer and then senior engineer with Canadian Westinghouse Co. Ltd., Hamilton, Ontario, Canada. During the period of October 1968 to August 1970, he was on leave of absence from Canadian Westinghouse to pursue PhD studies at McMaster University under the auspices of an NRC PIER Fellowship. In September 1970, he joined the Department of Electrical and Computer Engineering, University of Waterloo, Waterloo, Ontario, where he is currently a Distinguished Professor Emeritus. He served as department chairman during the period of July 1984 to June 1990. In 1996, he established the Centre for Wireless Communications (CWC) at the University of Waterloo and is currently serving as its founding director. Dr. Mark was on sabbatical leave at the IBM Thomas Watson Research Center, Yorktown Heights, New York, as a visiting research scientist (1976-1977); AT&T Bell Laboratories, Murray Hill, New Jersey, as a resident consultant (1982-1983); at the Laboratoire MASI, Universite Pierre et Marie Curie, Paris, France, as an invited professor (1990-1991); and at the Department of Electrical Engineering, National University of Singapore, as a visiting professor (1994-1995). He had previously worked in the areas of adaptive equalization, spread spectrum communications, and antijamming secure communication over satellites. His current research interests are in broadband and wireless communication networks, including network architecture, routing and control, and resource and mobility management in wireless and hybrid wireless/wireline communication networks. Dr. Mark is a Life Fellow of the IEEE. He was an editor of the *IEEE Transactions on Communications* for the period of 1983 to 1989. He served as the technical program chairman of INFOCOM '89. He was a member of the Inter-Society Steering Committee of the *IEEE/ACM Transactions on Networking* for the period of 1992-2003. He is currently an editor of the *ACM/Baltzer Wireless Networks* journal, and an associate editor of *Telecommunication Systems*.

▷ **For more information on this or any computing topic, please visit our Digital Library at** http://computer.org/publications/dlib.