A Call Admission Control Scheme for Packet Data in CDMA Cellular Communications

Lei Wang, *Student Member, IEEE,* and Weihua Zhuang, *Senior Member, IEEE*

***Abstract*— In wireless cellular communication systems, call admission control (CAC) is to ensure satisfactory services for mobile users and maximize the utilization of the limited radio spectrum. In this paper, we propose a new CAC scheme for a code division multiple access (CDMA) wireless cellular network supporting heterogeneous self-similar data traffic. In addition to ensuring transmission accuracy at the bit level, the CAC scheme guarantees service requirements at both the call level and the packet level. The grade of service (GoS) at the call level and the quality of service (QoS) at the packet level are evaluated using the handoff call dropping probability and the packet transmission delay, respectively. The effective bandwidth approach for data traffic is applied to guarantee QoS requirements. Handoff probability and cell overload probability are derived via the traffic aggregation method. The two probabilities are used to determine the handoff call dropping probability, and the GoS requirement can be guaranteed on a per call basis. Numerical analysis and computer simulation results demonstrate that the proposed CAC scheme can meet both QoS and GoS requirements and achieve efficient resource utilization.**

***Index Terms*— Call admission control (CAC), code division multiple access (CDMA), grade of service (GoS), handoff, packet data, quality of service (QoS), self-similar traffic.**

I. INTRODUCTION

WITH the significant growth of the Internet and wireless communication systems, it is expected that the next generation of wireless networks will use packet-switching technology to provide multimedia services to mobile users. However, the rapidly increasing number of mobile subscribers and the diversity of multimedia services pose significant technical challenges because of the limited radio spectrum. Therefore, improving spectrum utilization is a major design objective of any wireless network. Call admission control (CAC) is increasingly becoming important for achieving a good compromise between high resource utilization and satisfactory service provisioning.

Various CAC schemes are proposed in the literature. For example, in [1], an effective bandwidth based method is presented for traffic flows with the QoS requirement on bandwidth overload probability. In [2], an algorithm to estimate channel occupancy based on user and system history information is discussed. The resource reservation scheme in [3] is extended in [4] to include multiple classes of constant-bandwidth traffic flows. Different CAC schemes for a code-division multiple access (CDMA) cellular system that contains intercell interference are analyzed in [5]. In [6], CAC schemes are examined considering more general assumptions about the distribution of channel holding time. In [7], two prediction algorithms for multi-classes of constant bandwidth traffic are studied. However, because of complex issues caused by user mobility, most of the previous works has focused on the circuit-switched voice (or constant rate) traffic. The QoS parameters are often limited to the handoff call dropping probability or the cell overload probability. The traffic load is usually estimated based on history information, and the QoS provisioning is often enforced on a per cell basis. For the third generation (and beyond) wireless networks based on packet switching technology, the packet level performance (e.g., packet delay bound, or packet delay probability) should also be taken into account in CAC. It has been found that the data traffic can be described statistically by self-similarity models [8]–[10]. Preliminary measurements and simulation results [11] show that the data traffic in a wireless data network also exhibits self-similarity characteristics. Resource allocation based on self-similar traffic models in wireless networks has started to gain the attention from researchers [12].

In this paper we propose a new effective bandwidth based CAC scheme for the uplink of a CDMA cellular system that supports heterogeneous data traffic with self similarity. The traffic model is more accurate for data traffic in a hybrid wireless/Internet environment as compared with previously used models. The service quality parameters under consideration are transmission accuracy at the bit level (QoS at the physical layer), packet transmission delay at the packet level (QoS at the link layer), and the handoff call dropping probability at the call level (GoS at the network layer), respectively. Because of the dynamics of data traffic flows and user mobility, development of a CAC scheme to guarantee the performance simultaneously at all the three layers for self-similar traffic is technically very challenging. To overcome the complexity, we use the effective bandwidth approach to decouple the QoS provisioning and GoS provisioning. The requirements at the bit and packet levels and the impact of traffic statistics on the capacity requirement are represented by the effective bandwidth of each traffic flow. As a result, we are able to focus on user mobility issues in the GoS provisioning, where each call is characterized by its effective bandwidth. Different from previous works, the effective bandwidth derived in this paper for each call changes from time to time in order to accurately

Manuscript received October 2003; revised October 2004 and January 2005; accepted March 7, 2005. The editor coordinating the review of this paper and approving it for publication is L. Wang. This work was supported by research grants from Bell University Labs (BUL) at the University of Waterloo, Communications Information Technology Ontario (CITO), and the Natural Science and Engineering Research Council (NSERC) of Canada.

The authors are with the Centre for Wireless Communications (CWC), Department of Electrical and Computer Engineering, University of Waterloo, Waterloo, ON N2L 3G1, Canada (e-mail: {l29wang, wzhuang}@bbcr.uwaterloo.ca).

Digital Object Identifier 10.1109/TWC.2006.02020.

capture the dynamics of statistical multiplexing and the impact of intercell interference on cell capacity. Based on the traffic aggregation instead of the estimation from history information, a more precise and rigorous presentation of the cell overload probability is derived. The CAC procedure is then designed to ensure both QoS and GoS provisioning on a per call basis. In comparison with previous studies, the novelty of the CAC scheme proposed here lies in that: 1) the system model is based on the packet-switching mode with statistical multiplexing at the packet level; 2) the system supports heterogeneous data traffic with self-similarity characteristics, which is a more accurate traffic model for wireless Internet access; 3) the CAC scheme is designed to guarantee both GoS at the call level and QoS at the packet and bit levels; and 4) the GoS and QoS provisioning can be implemented either on a per cell basis or on a per call basis.

The reminder of this paper is organized as follows. Section II describes the system model that supports heterogeneous data traffic, and specifies the service requirements. Section III outlines the admission control procedures used in our CAC scheme. Section IV presents the QoS provisioning and Section V studies the GoS provisioning under the QoS constraint. Section VI presents numerical results to demonstrate the performance of the proposed CAC scheme, followed by the conclusions in Section VII.

II. System Model

Consider a cellular CDMA system having K hexagonal cells, providing multimedia services to mobile users with perfect power control. Under the assumption that the complete partitioning mechanism is used for resource allocation among the services, the admission control of each service can be considered separately. In this paper, we focus on CAC for data service only.

A. *Traffic and Mobility Models*

The packet data traffic with self-similarity [10] is considered in this paper. Given a stationary time series $X = (X_t; t = 1, 2, 3, ...)$ with zero mean, $X^{(j)} = (X_k^{(j)}; k = 1, 2, 3, ...)$ is defined by summing the original X over non-overlapping blocks of size j. If for all positive j, $X^{(j)}$ has the same distribution as X rescaled by j^H, X is self-similar with Hurst parameter H. Self-similar processes also show long range dependence (LRD). Among various self-similar process models, the fractional Brownian motion model [13], [14] is more mathematically tractable and is, therefore, used here to model the packet arrival process. A fractional Brownian motion process $Z(t)$ is defined by following properties [14]: $Z(t)$ is Gaussian with stationary increment and continuous path, and its initial value, mean and variance are given by $Z(0) = 0$, $E[Z(t)] = 0$ and $Var[Z(t)] = \sigma_b^2 t^{2H}$, respectively.

Let M denote the number of data traffic classes. All the classes of traffic have self-similar characteristic. The packet traffic of a class m ($m \in \{1, 2, ..., M\}$) source is a fractional Brownian traffic characterized by parameters λ_{bm}, σ_{bm}^2, and H, where λ_{bm} is the mean packet arrival rate, H is the Hurst parameter, and $\sigma_{bm}^2 t^{2H}$ is the variance of total packets arrived by time t. Previous studies [9], [10], [15] show that for the Internet data traffic, the Hurst parameter H usually lies between $(0.8, 0.9)$. The number of packets sent by a class m traffic source during $[0, t]$ is given by

$$Y_m(t) = \lambda_{bm} t + Z_m(t) \tag{1}$$

where $Z_m(t)$ is a fractional Brownian motion process with variance $Var[Z_m(t)] = \sigma_{bm}^2 t^{2H}$.

It has been found that, for the Internet data traffic, although the packet arrival process is not Poisson, the arrival process of the aggregate connections (e.g., TCP connections) at a network node can be well modeled by a Poisson process [8]. Similarly, the call arrival process at the base station (BS) is considered to be Poisson. Assuming that the packet generating processes of all the users in a cell are independent, the call arrival process from every user is also Poisson [16] with a mean call arrival rate λ.

Assume that the initial location of mobile users follows a uniform distribution in the service area. The movement of each user is modeled by random walk and is a stationary process. User's cell residence time t_r follows a negative exponential distribution with mean $1/\mu_r$, and the call duration t_c follows a negative exponential distribution with mean $1/\mu_c$. As a result, the channel holding time of each call, $t_h = \min\{t_r, t_c\}$, follows a negative exponential distribution with mean $1/h$, where $h = \mu_r + \mu_c$.

B. *Effective Capacity*

Based on the traffic and mobility models, we consider the uplink packet transmission of the CDMA system. Synchronous packet transmission is assumed for simplicity in the physical layer analysis. However, based on a statistical throughtput analysis similar to that for carrier sensing multiple access with collision detection (CSMA/CD) in computer networks, the proposed packet level QoS provisioning and the call level GoS provisioning can be extended to asynchronous or quasi-synchronous packet transmission which is more realistic for the uplink. Assume that all packets have a fixed size of L_p bits, and are transmitted with bit rate r_b. Let W denote the total available uplink bandwidth in each cell, γ_b the bit energy-to-interference-plus-noise density ratio (SINR) averaged over the time duration of each packet, P_j the average received power of a packet from jth mobile user, I_c the power of intercell interference, and P_n the power of background noise. Because we focus on the packet level and the call level analysis with a longer time-scale, with perfect power control at the receiver, we assume the channel to be relatively static. Given channel fading statistics, modulation and coding schemes, diversity, and receiver structure, there is a one-to-one mapping relation between the required bit error rate and the received SINR. To achieve a target transmission bit error rate in the CDMA system, it is required that γ_b should not be lower than a given threshold $(E_b/I_0)_d$ [17]. With a constant processing gain, r_b is same for all data flows. For the simplicity of presentation, $(E_b/I_0)_d$ values are assumed to be same for all the users. Let Γ denote the required minimum SINR [1], where $\Gamma = r_b \cdot (E_b/I_0)_d$. Without loss of generality, the reception of a packet from user 1 is considered. The average received signal power levels from all the users in the cell are the same. We have

$$\begin{aligned}\gamma_b \geq (\frac{E_b}{I_0})_d \quad &\Longleftrightarrow \quad \frac{\gamma_b \cdot r_b}{W} \geq \frac{\Gamma}{W} \\ &\Longleftrightarrow \quad \frac{P_1}{\sum_{j>1} P_j + I_c + P_n} \geq \frac{\Gamma}{W}. \quad (2)\end{aligned}$$

Therefore, the number of simultaneous packets at any time under the SINR constraint, N_p, should satisfy

$$N_p = 1 + \frac{\sum_{j>1} P_j}{P_1} \leq 1 + \frac{W}{\Gamma} - \frac{I_c}{P_1} - \frac{P_n}{P_1} \triangleq N_{\max}. \quad (3)$$

Let C_i^e denote the effective capacity of the target cell i, defined as the service rate (maximum number of packets successfully transmitted per second) in the uplink. We have

$$\begin{aligned}C_i^e &= \frac{N_{\max} \cdot r_b}{L_p} = \frac{r_b}{L_p} + \frac{W \cdot r_b}{\Gamma \cdot L_p} - \frac{P_n \cdot r_b}{P_1 \cdot L_p} - \frac{I_c \cdot r_b}{P_1 \cdot L_p} \\ &= C_s - C_i^I \quad (4)\end{aligned}$$

where $C_s = \frac{r_b}{L_p} + \frac{W \cdot r_b}{\Gamma \cdot L_p} - \frac{P_n \cdot r_b}{P_1 \cdot L_p}$ denotes the effective capacity in the absence of intercell interference (such as in a single-cell system), and $C_i^I = \frac{I_c \cdot r_b}{P_1 \cdot L_p}$ is the capacity loss caused by intercell interference (CLI) due to frequency reuse in the CDMA system.

Assuming that each mobile terminal has a buffer of very large size, a packet scheduling scheme can be designed such that a packet will be transmitted if and only if the SINR requirement is satisfied, otherwise the packet will be put into the buffer. The packets are transmitted in the order of their arrival times. If the cell effective capacity does not change much over a short observation period, the uplink packet transmission of the target cell can be approximately modeled by a $G/D/1/\infty$ queueing system with a constant service rate C_i^e, where the arrival process is an aggregate process of multiple fractional Brownian motion processes.

C. QoS and GoS Requirements

Let t_d denote the transmission delay of a packet. Given a packet delay threshold T_d, the QoS requirement on the transmission delay is given by parameter P_d where $P_r\{t_d > T_d\} \leq P_d$. Let p_d denote the handoff call dropping probability, the GoS requirement is given by parameter P_h where $p_d \leq P_h$. The resource utilization of the system is represented by $\frac{m_{si}}{C_i^e}$, where m_{si} denotes the average packet transmitting rate in the target cell i.

In a CDMA system, the transmission accuracy is guaranteed by power (resource) allocation to achieve the target $(E_b/I_0)_d$, the required packet delay bound is to be met by resource allocation in terms of the effective bandwidth, and the required handoff call dropping probability upper bound is to be ensured by resource reservation in neighboring cells. The objective of the CAC scheme is to provide users with guaranteed QoS at the bit and packet levels and guaranteed GoS at the call level, while maximizing the resource utilization by admitting a maximum number of mobile users for a given λ.

III. The CAC Procedure

As illustrated in Fig. 1, the proposed CAC scheme consists of two routines that are executed in each base station: the system monitoring routine and the incoming call admission routine. The system monitoring routine is invoked periodically and can also be triggered by system status changes. It continuously measures the intercell interference, estimates CLI, and then updates C_i^e accordingly. The invocation period is referred to as monitoring period. During each monitoring period, C_i^e is assumed to remain constant. The incoming call admission routine is invoked whenever a handoff or new call arrives. For each incoming handoff call, if the QoS requirements of all the ongoing calls and the incoming call can be satisfied in the current cell, the call is admitted, as discussed in Section IV. For each incoming new call, the call is admitted only if it first passes the QoS checking process (as for a handoff call) and then passes a GoS checking process under the QoS constraint. The latter is to ensure that the call will be provided guaranteed GoS and QoS in all possible subsequent cells (if handoff happens) during a control period T, as discussed in Section V. The control period T is introduced to reduce the complexity of the algorithm while maintaining satisfactory QoS and GoS provisioning, and is assumed to be much shorter than the monitoring period. It is a sliding time window for each call, as shown in Fig. 2.

The QoS checking process in both routines is implemented based on the effective bandwidth approach. The effective bandwidth depends on the packet delay bound requirement. A linear approximation of the effective bandwidth is applied to reduce the scheme complexity.

The GoS checking process is implemented based on the estimation of the traffic load in every cell. By aggregating the traffic load from all possible incoming mobile users, the future traffic load in each cell can be precisely estimated. Then, with derived cell overload probability and mobile user's handoff probability, the overall handoff call dropping probability can be estimated. Finally, a GoS guarantee subject to QoS constraint is implemented on both per call and per cell bases. The CAC procedure should be simple for online implementation. The complexity in designing the CAC scheme lies in how to obtain CAC parameters for QoS and GoS satisfaction, as discussed in Section IV and Section V.

IV. QoS Provisioning

For heterogeneous traffic, define $\{n_m : m \in \{1, ..., M\}\}$ as the state of the target cell i, where n_m is the number of ongoing class m calls in the cell. Let Q denote the packet queue length in the buffer of the $G/D/1/\infty$ queueing system. The packet-level QoS requirement, $P_r\{t_d > T_d\} \leq P_d$, is equivalent to

$$P_r\{Q > \frac{T_d \cdot r_b}{L_p}\} \leq P_d. \quad (5)$$

Note that the $G/D/1/\infty$ queueing model is valid for each monitoring period over which the service rate C_i^e is a constant. From one monitoring period to the next monitoring period, if C_i^e changes, the constant service rate is violated. On the other hand, the CAC routines adapt to the service rate variation from one period to another period. If the C_i^e change over

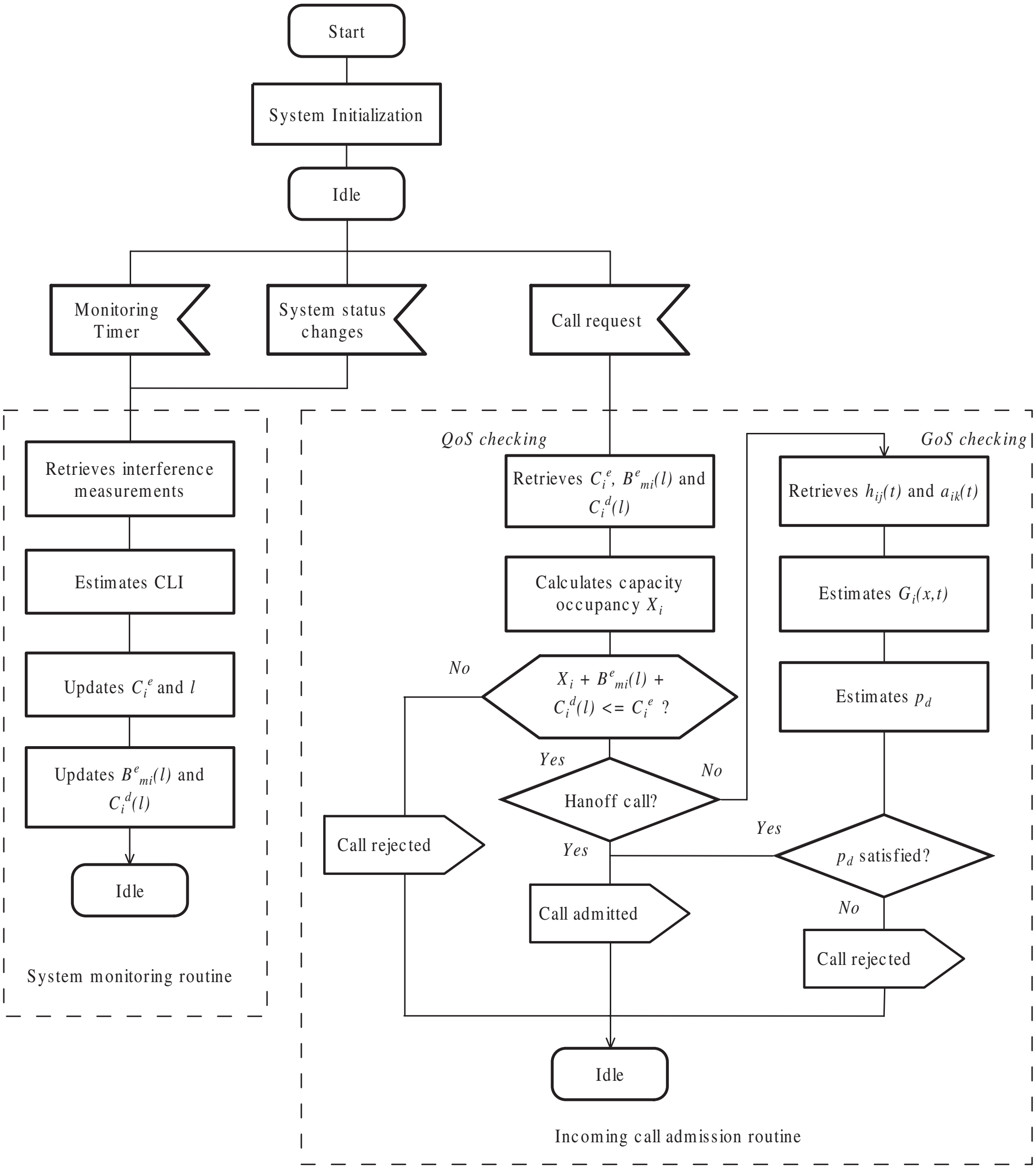

Start
System Initialization
Idle
Monitoring Timer
System status changes
Call request
Retrieves interference measurements
Estimates CLI
Updates C_i^e and l
Updates $B^e_{mi}(l)$ and $C_i^d(l)$
Idle
System monitoring routine
QoS checking
Retrieves C_i^e, $B^e_{mi}(l)$ and $C_i^d(l)$
Calculates capacity occupancy X_i
No
$X_i + B^e_{mi}(l) + C_i^d(l) <= C_i^e$?
Yes
Hanoff call?
No
Yes
Call rejected
Call admitted
GoS checking
Retrieves $h_{ij}(t)$ and $a_{ik}(t)$
Estimates $G_i(x,t)$
Estimates p_d
Yes
p_d satisfied?
No
Call rejected
Idle
Incoming call admission routine

Fig. 1. Flowchart of the proposed CAC scheme.

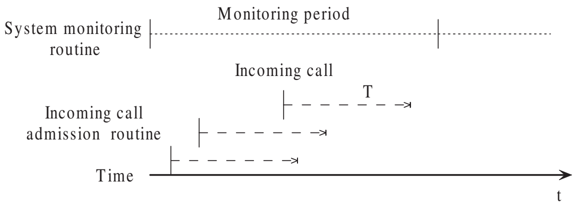


Fig. 2. Time frame for the system monitoring routine and incoming call admission routine.

adjacent monitoring periods is not significant, the queueing system should reach the steady state with the new service rate after a short transient period. In this section, to make the analysis tractable, we neglect the transient period in the queueing system.

A. Minimum Required Effective Capacity

In this section, we solve the QoS provisioning problem by deriving a relation between the number of ongoing calls in the cell and the required service rate to support these calls. Let $C_r(n_1, n_2, ..., n_M)$ denote the lower bound of the required effective capacity to serve all the n_1 class 1 calls, n_2 class 2 calls, ..., and n_M class M calls with guaranteed QoS. Extending the analysis of queueing performance in wireline networks given in Section 3.6.2 of [13], we can derive $C_r(n_1, n_2, ..., n_M)$ as

$$\begin{aligned} C_r(n_1,n_2,...,n_M) &= H(\frac{1-H}{\beta})^{\frac{1}{H}-1}(2\alpha\sum_{m=1}^{M} n_m\sigma_{bm}^2)^{\frac{1}{2H}} \\ &+ \sum_{m=1}^{M} n_m\lambda_{bm} \end{aligned} \tag{6}$$

where $\beta = \frac{T_d \cdot r_b}{L_p}$ and $\alpha = -\ln P_d$. From (6), the minimum required effective capacity depends on QoS parameters and traffic statistics, and is a nonlinear function of $n_1, n_2, \ldots, n_M$. Given the QoS parameters and traffic statistics, the effective bandwidth of each traffic flow is a decreasing function of the numbers of other traffic flows sharing the overall system resources. This is because an increased number of traffic flows results in better statistical multiplexing among the flows, leading to a reduced effective bandwidth for each of the flows. Given the cell effective capacity C_i^e determined by (4) and the cell state $\{n_1, n_2, ..., n_M\}$ with $C_r(n_1, ..., n_M) \leq C_i^e$, the QoS checking routine can be described as follows: When an incoming call of class m arrives, if

$$C_r(n_1, ..., n_{m-1}, n_m + 1, ..., n_M) \leq C_i^e \tag{7}$$

then the incoming call is admissible. To reduce the algorithm complexity in the heterogeneous traffic environment, the runtime QoS checking algorithm (7) can be approximately replaced by a table lookup, which is computed offline. This approximation is also necessary in order to simplify and execute the GoS provisioning step in the CAC scheme.

B. Linear Approximation for Homogeneous Environment

First consider the homogeneous traffic ($M = 1$) case. The maximum number of active calls in the target cell i, $n_{\max}$, can be determined by the constraint $C_r(n_{\max}) \leq C_i^e$. From (7),

$$\kappa \cdot n_{\max}^{\frac{1}{2H}} + n_{\max}\lambda_b = C_i^e \tag{8}$$

where $\kappa = H(\frac{1-H}{\beta})^{\frac{1}{H}-1}(2\alpha\sigma_b^2)^{\frac{1}{2H}}$. As the amount of resources required by each traffic flow depends on the degree of multiplexing in the cell, $C_r(n)$ is a nonlinear function of n (here for notation simplicity we use n to denote n_1). To simplify the CAC decision process, we linearize the function $C_r(n)$ locally over an interval of cell capacity containing C_i^e. As C_i^e changes with the intercell interference, we cannot predict its value. As a result, we need to obtain a linear approximation of $C_r(n)$ over all possible values of C_i^e. Partition the cell capacity interval, $[0, C_s]$, to L sections, and use piecewise linear approximation to represent $C_r(n)$. For each section $[C_{l-1}, C_l]$, where $l \in [1, L]$, and $0 \leq C_{l-1} < C_l \leq C_s$, apply a local linear approximation to $C_r(n)$, as shown in Fig. 3. The line $\lambda_b n$ represents the minimum capacity required to support n ongoing calls, determined by the mean packet arrival rate. The line $\lambda_p n$ represents the maximum capacity required to support n ongoing calls, determined by the peak packet arrival rate λ_p. The convex curve $C_r(n)$, which is the lower bound of capacity required to guarantee the QoS requirements of n ongoing calls, should lie between lines $\lambda_p n$ and $\lambda_b n$.

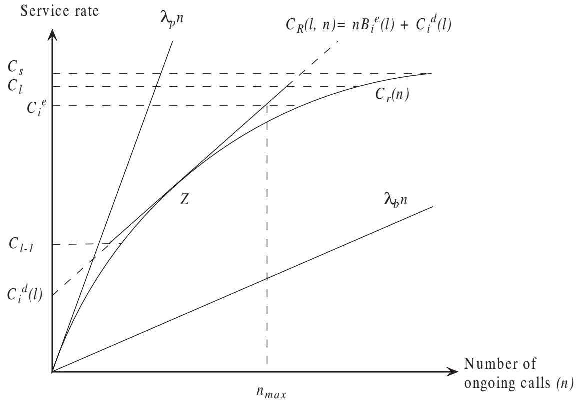


Fig. 3. The required capacity $C_r(n)$ and its local linear approximation $C_R(l, n)$ over the capacity interval $[C_{l-1}, C_l]$ for homogeneous traffic.

When $n \to \infty$, $C_r(n)$ should approach $\lambda_b n$ because of the perfect statistical multiplexing. Line $C_R(l, n)$ ($\geq C_r(n)$) is the approximation of $C_r(n)$ on the service rate interval $[C_{l-1}, C_l]$, and it touches the curve $C_r(n)$ at point Z. $C_R(l, n)$ can be represented by

$$C_R(l, n) = nB_i^e(l) + C_i^d(l) \tag{9}$$

where $B_i^e(l)$ is defined as the effective bandwidth of each user in cell i when the cell effective capacity C_i^e falls in $[C_{l-1}, C_l]$, and $C_i^d(l)$ is defined as the effective capacity adjustment. In general, $B_i^e(l)$ decreases as l increases, due to a higher statistical multiplexing gain. The maximum number of the active calls with guaranteed QoS can be determined by

$$n_{\max} = \frac{C_i^e - C_i^d(l)}{B_i^e(l)} \triangleq C_R^{-1}(l, C_i^e) \tag{10}$$

where $C_i^e \in [C_{l-1}, C_l]$. Note that $C_R^{-1}(l, C_i^e)$ is the inverse function of $C_R(l, n)$ with respect to the second variable.

Let $C_r^{-1}(C_x)$ denote the inverse function of $C_r(n)$, where C_x denote the service rate. If C_i^e is uniformly distributed on $[C_{l-1}, C_l]$, then the touch point Z and parameters $B_i^e(l)$ and $C_i^d(l)$ can be determined by

$$\min\{\int_{C_{l-1}}^{C_l} (C_r^{-1}(C_x) - C_R^{-1}(l, C_x))dC_x\}$$

in order to achieve the maximum utilization of capacity for $C_e^i \in [C_{l-1}, C_l]$. For each l, parameters $B_i^e(l)$ and $C_i^d(l)$ are computed offline and are downloaded to the BS during the system initialization. The admission criterion (7) is simplified to

$$(n+1) \cdot B_i^e(l) + C_i^d(l) \leq C_i^e. \tag{11}$$

C. Linear Approximation for Heterogeneous Environment

In the case of heterogeneous traffic ($M > 1$), the lower bound of the required effective capacity $C_r(n_1, n_2, ..., n_M)$ becomes a convex surface. The corresponding problem is to find a hyperplane $C_R(l, n_1, n_2, ..., n_M)$ on each partition $[C_{l-1}, C_l]$ that touches $C_r(n_1, n_2, ..., n_M)$ at a point while maximizing the capacity utilization, which can be represented in a form

$$C_R(l, n_1, n_2, ..., n_M) = \sum_{m=1}^{M} n_m B_{mi}^e(l) + C_i^d(l) \tag{12}$$

where $B_{mi}^e(l)$ is the effective bandwidth of each class m call in cell i. Define the capacity occupancy in cell i as

$$X_i = \sum_{m=1}^{M} n_m B_{mi}^e(l). \tag{13}$$

The admission criterion (7) for an incoming class-m call in cell i becomes

$$X_i + B^e_{mi}(l) + C^d_i(l) \le C^e_i \tag{14}$$

for $C^e_i \in [C_{l-1}, C_l]$.

The number of capacity partition intervals, L, is determined based on the precision requirement on the piece-wise linear approximation (9) of $C_r(n)$. A larger L value results in a more accurate approximation. As mentioned in Section III, the monitoring routine of the CAC scheme continuously measures the intercell interference and estimates C^I_i. During each monitoring period, C^e_i is determined and the corresponding capacity partition (denoted by l) is fixed. Given the l value, $B^e_{mi}(l)$ and $C^d_i(l)$ can be determined. That is, based on the measured intercell interference, the values of C^e_i, $B^e_{mi}(l)$, and $C^d_i(l)$ are updated and remain unchanged over each monitoring period.

V. GoS Provisioning Under QoS Constraint

For the call level GoS requirement, we consider the handoff call dropping probability. A handoff call is dropped if and only if the target new cell cannot guarantee the QoS requirements for the ongoing calls and for the incoming handoff call. For each handoff call, the admission criterion with GoS provisioning is to satisfy $p_d \le P_h$ in its call duration, after the QoS checking process. The handoff call dropping probability p_d can be derived from the probability of a call being handed off to a neighboring cell and the probability of the neighboring cell being overloaded at the moment that the handoff occurs.

Let $G_i(x,t)$ denote the traffic load probability distribution of cell i, defined as the probability that cell capacity occupancy is beyond x at time t. The traffic load probability of a cell at a future time t can be derived by estimating the contribution of every user in the system to the traffic load in the cell at time t. Given that a call is ongoing in cell j at time 0, let $h_{ij}(t)$ denote the conditional handoff probability, defined as the probability that at time t the call is ongoing in cell i and at least one handoff has occurred since time 0. The conditional handoff probability of each active user can be derived by enumerating all the possible paths from the user's current location to the target cell and by estimating the call duration. If both $G_i(x,t)$ and $h_{ij}(t)$ are known, the handoff call dropping probability p_d can be derived. To implement the CAC on a per cell basis, only $G_i(x,t)$ is needed; however, to implement the CAC on a per call basis, both $G_i(x,t)$ and $h_{ij}(t)$ are required.

With effective CAC ensuring a handoff call dropping probability of the order $10^{-4} \sim 10^{-2}$, we can assume essentially that all the handoffs are successful in deriving the handoff probability and traffic load probability. The resulting estimate of handoff call dropping probability will be slightly more conservative than that without the assumption, because the computed cell overload probability is larger than the actual value. In the following subsections, we first calculate $h_{ij}(t)$, then derive $G_i(x,t)$, and finally analyze the per-cell based and per-call based GoS provisioning.

A. Conditional Handoff Probability

Consider the coverage of a cellular system as illustrated in Fig. 4. The target cell j has several types of neighboring cells, depending on the relative position of each cell with respect to the target cell. There are 3 chains of hexagons centered at cell j. As all the tier-1 cells are on the chains, they all belong to the same type, say type 1 (denoted by T_1). For tier-2 cells, half of them are on the chains and are referred to as type 1 (denoted by T_{21}), and the other half are referred to as type 2 (denoted by T_{22}). Similarly, there are two types of tier-3 cells (denoted by T_{31} and T_{32}).

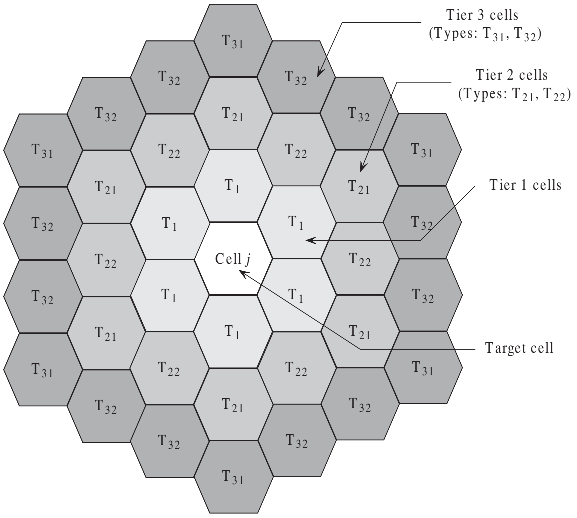


Fig. 4. Tiers and types of the neighboring cells.

Given that a call is ongoing at time 0, let $q_n(t)$ denote call path probability, defined as the conditional probability of a call having n handoffs by time t. We have

$$\begin{aligned} q_0(t) &= P_r\{t_c > t, t_r > t\} = e^{-\mu_r t}e^{-\mu_c t} = e^{-(\mu_r+\mu_c)t} \\ q_1(t) &= P_r\{t_c > t, 1 \text{ handoff in } [0,t], \text{ handoff target} \\ &\qquad \text{cell is a Tier 1 cell}\} \\ &= \frac{1}{1!}(\frac{\mu_r t}{6})^1 e^{-(\mu_r+\mu_c)t}. \end{aligned}$$

In general, it can be derived that (also given in [2])

$$q_n(t) = \frac{1}{n!}(\frac{\mu_r t}{6})^n e^{-(\mu_r+\mu_c)t}. \tag{15}$$

Using an approach similar to that in [2], and based on the assumption of user random walk, for each $\{i,j\}$ pair, $h_{ij}(t)$ can be calculated by enumerating every possible path from cell j to cell i with one handoff, two handoffs, three handoffs, and so on. For example, for $i = j$, there are 6 possible paths with 2 handoffs (one handoff from cell j to a T_1 cell and one handoff from the T_1 cell back to cell j), and 12 paths with 3 handoffs (each path travels via 2 neighbouring T_1 cells and then back to cell j). Then, for the first 3 tiers of neighboring cells, we have

$$h_{ij}(t) = \begin{cases} 6q_2(t) + 12 \cdot q_3(t) + \cdots & i = j \\ q_1(t) + 2q_2(t) + 15 \cdot q_3(t) + \cdots & i \in T_1 \\ q_2(t) + 3 \cdot 2 \cdot q_3(t) + \cdots & i \in T_{21} \\ 2q_2(t) + 6 \cdot q_3(t) + \cdots & i \in T_{22} \\ q_3(t) + \cdots & i \in T_{31} \\ 3q_3(t) + \cdots & i \in T_{32}. \end{cases} \tag{16}$$

B. Cell Overload Probability

To derive the cell overload probability, let $p_n(t)$ denote user path probability, defined as the conditional probability that a user has crossed cell borders n times during period $[0, t]$, for a movement path. Given that a user is in cell k at time 0, define user transition probability $a_{ik}(t)$ as the conditional probability that the user is in cell i at time t. From the definition, it can be derived that

$$p_n(t) = \frac{1}{n!}(\frac{\mu_r t}{6})^n e^{-\mu_r t}. \tag{17}$$

In Fig. 4, if we replace the target cell j by cell k and let cell i denote a neighboring cell of cell k, it can be derived that

$$a_{ik}(t) = \begin{cases} p_0(t) + 6p_2(t) + 6 \cdot 2 \cdot p_3(t) + \cdots & i = k \\ p_1(t) + 2p_2(t) + 15 \cdot p_3(t) + \cdots & i \in T_1 \\ p_2(t) + 3 \cdot 2 \cdot p_3(t) + \cdots & i \in T_{21} \\ 2p_2(t) + 6 \cdot p_3(t) + \cdots & i \in T_{22} \\ p_3(t) + \cdots & i \in T_{31} \\ 3p_3(t) + \cdots & i \in T_{32}. \end{cases} \tag{18}$$

In practice, the average channel holding time and average cell residence time should be much shorter than the average call inter-arrival time from a single user. As a result, we assume $\mu_c \ll \frac{1}{\lambda}$, and $\mu_r \ll \frac{1}{\lambda}$. For the purpose of CAC, we only need to consider the possibility of at most one new call from one user during a period $[0, t]$.

Assuming that a call is ongoing at time 0, the probability that the call is still ongoing at time t is $e^{-\mu_c t}$. If a user is idle at time 0, the probability that one call is generated by the user during period $[0, t]$ is $\lambda t e^{-\lambda t}$. Given that there is one call generated by the user during period $[0, t]$, the call arrival moment is uniformly distributed on $[0, t]$. Therefore, if a user is idle at time 0, the probability that at time t the user has a call ongoing is given by

$$\begin{aligned} & P_r\{\text{User made a call in } [0, t]\} \\ & \cdot P_r\{\text{Call is still on at time } t \mid \text{user made a call in } [0, t]\} \\ & = \lambda t(1 - P_b)e^{-\lambda t} \int_0^t \frac{1}{t} e^{-\mu_c(t-s)} ds = \frac{\lambda(1 - P_b)}{\mu_c} e^{-\lambda t} \end{aligned}$$

where P_b is the average new call blocking probability in the system estimated by runtime measurements. During each control period, C_i^e, C_i^d, and B_{mi}^e are approximately considered to be constant (here we omit the variable l for notation simplicity). The cell overload probability can be estimated from the distribution of the cell capacity occupancy. Let $X_i(t)$ $(\leq C_i^e - C_i^d)$ denote the capacity occupancy in cell i at time t, where $i \in \{1, .., K\}$. At time 0 in cell k, among all the subscribers, there are n_{mk0}^a class-m users having an active connection, and the rest n_{mk0}^i class-m users are in an idle state, $m \in \{1, ..., M\}$, and $k \in \{1, ..., K\}$. $X_i(t)$ can be viewed as the sum of $2MK$ independent Bernoulli random variables: for $\forall k \in \{1, ..., K\}$, and $\forall m \in \{1, .., M\}$, among which n_{mk0}^a i.i.d. random variables have a probability of $e^{-\mu_c t} a_{ik}(t)$ to take value B_{mi}^e, and the rest n_{mk0}^i i.i.d. random variables have a probability of $\frac{\lambda(1-P_b)}{\mu_c} e^{-\lambda t} a_{ik}(t)$ to take value B_{mi}^e. These n_{mk0}^a and n_{mk0}^i random variables characterize the handoff traffic contributions from neighboring cell k to target cell i. Each call will have different effective bandwidth contributions to different cells at time t (> 0). No matter what the initial effective bandwidth that a call has in cell k at time 0, its effective bandwidth contributing to cell i depends on the status of cell i at time t, i.e., B_{mi}^e can be calculated based on the traffic load and interference level in cell i. When there are a large number of users, based on the central limit theorem, $X_i(t)$ at any t is approximately a Gaussian random variable, i.e., $X_i(t) \sim N(\mu_{fi}(t), \sigma_{fi}^2(t))$ with

$$\mu_{fi}(t) = \textstyle\sum_{k=1}^{K} \sum_{m=1}^{M} \{n_{mk0}^a \mu_{mik}^c(t) + n_{mk0}^i \mu_{mik}^i(t)\}$$

$$\sigma_{fi}^2(t) = \textstyle\sum_{k=1}^{K} \sum_{m=1}^{M} \{n_{mk0}^a (\sigma_{mik}^c(t))^2 + n_{mk0}^i (\sigma_{mik}^i(t))^2\}$$

where

$$\begin{aligned} \mu_{mik}^c(t) \quad & = e^{-\mu_c t} a_{ik}(t) \cdot B_{mi}^e \\ \mu_{mik}^i(t) \quad & = \tfrac{\lambda(1-P_b)}{\mu_c} e^{-\lambda t} a_{ik}(t) \cdot B_{mi}^e \\ (\sigma_{mik}^c(t))^2 \quad & = (e^{-\mu_c t} a_{ik}(t) - (e^{-\mu_c t} a_{ik}(t))^2) \cdot (B_{mi}^e)^2 \\ (\sigma_{mik}^i(t))^2 \quad & = (\tfrac{\lambda(1-P_b)}{\mu_c} e^{-\lambda t} a_{ik}(t) \\ & \quad -(\tfrac{\lambda(1-P_b)}{\mu_c} e^{-\lambda t} a_{ik}(t))^2) \cdot (B_{mi}^e)^2. \end{aligned}$$

As a result, the traffic load probability of cell i at time t is approximately given by

$$G_i(x, t) = \frac{1}{2} \text{erfc} \left[\frac{x - \mu_{fi}(t)}{\sqrt{2}\sigma_{fi}(t)} \right]. \tag{19}$$

Given effective capacity $C_i^e = C_s - C_i^I$, the overload probability of cell i is $G_i(C_i^e - C_i^d, t)$.

C. GoS Provisioning Subject to QoS Constraint

With the conditional handoff probability given by (16) and the cell overload probability given by (19), the handoff call dropping probability can be derived for the GoS provisioning. To enforce GoS provisioning in a control period $[0, T]$, we have two options:

(1) GoS provisioning on a per cell basis — When the system capacity occupancy is ergodic, by enforcing the same admission criteria in every cell, the limiting handoff call dropping probability of a single user equals to the average cell overload probability, given by

$$\tilde{p}_d = \frac{1}{T} \int_o^T G_i(C_i^e - C_i^d, t) dt. \tag{20}$$

If a new call in cell $i, i \in \{1, ..., K\}$, is admitted with a probability r_i (i.e., the new call blocking probability is $1 - r_i$), the effective user call arrival rate to the system is $r_i \lambda$. Then $\mu_{mik}^i(t)$, $(\sigma_{mik}^i(t))^2$, $\mu_{fi}(t)$, $\sigma_{fi}^2(t)$, and $G_i(C_i^e - C_i^d, t)$ are updated accordingly. The probability r_i is selected to ensure $\tilde{p_d} \leq P_h$. In this way, GoS/QoS provisioning is implemented for every cell;

(2) GoS provisioning on a per call basis — By computing the handoff call dropping probability for each new call, we can have more accurate GoS/QoS provisioning. Given an ongoing call in cell j with conditional handoff probability $h_{ij}(t)$ for handoffs from cell j to i, the moment of its last handoff (to cell

i) is uniformly distributed on $[0, t]$. The probability that the last handoff occurs during $[t - dt, t]$ is $h_{ij}(t)\frac{dt}{t}$. The probability that, during $[t-dt, t]$, the call handoffs to cell i, but is dropped because the QoS cannot be guaranteed (due to cell overload) in cell i is $G_i(C_i^e - C_i^d, t)h_{ij}(t)\frac{dt}{t}$. Therefore, the overall handoff call dropping probability for the new call generated in cell j during its lifetime can be estimated by

$$p_d = \sum_{i=1}^{K} \int_0^{\infty} \frac{1}{t} G_i(C_i^e - C_i^d, t)h_{ij}(t)dt. \tag{21}$$

The admission criterion for GoS provisioning is $p_d \leq P_h$. If the condition is satisfied, then the new call is admissible. In this way, the GoS/QoS provisioning is implemented for every call.

If the mobile user location is uniformly distributed in the system coverage area and the system is stationary, the two CAC options should have the same performance. However, in a practical system where the traffic distribution fluctuates and where hot spots exist from time to time, the second CAC option is more appropriate because it responses more promptly to the system traffic load changes. Also, the second option achieves better fairness among users of the same traffic class because the QoS/GoS provisioning is implemented on a per call basis.

To summarize, similar to many previous CAC schemes, the QoS/GoS provisioning scheme proposed here is essentially based on traffic prediction and resource reservation. However, our approach in resource allocation/reservation is adaptive to traffic load dynamics and takes into account various degrees of statistical multiplexing among all the users sharing the total available resources of a cell. Given the current system traffic load status, when calculating the cell overload probability, the future traffic load is predicted based on the system model with random user walk, Poisson call arrival and negative exponential call duration. To guarantee GoS and QoS requirements, radio spectrum resources are implicitly reserved for calls already in service by comparing the total effective bandwidth with the cell effective capacity.

VI. Numerical Results

Two scenarios are considered for numerical analysis. First, we study the effect of traffic parameters on the effective bandwidth and QoS provisioning; then, we study the GoS provisioning as a function of the traffic load. We focus on the packet level QoS and call level GoS, and consider homogeneous traffic. The system parameters used in the analysis and simulation are: $C_i^e \in [0.098 \times 10^6, 0.187 \times 10^6]$ packets per second, $\beta = 200$ packets, $P_d = 0.002$, $H = 0.86$, $\lambda_b = 10,240$ packets per second, $\sigma_b^2 = 17.3 \times 10^6$, $L = 20$. The fractional Brownian motion traffic flow is generated by a modified random midpoint displacement (RMD) algorithm [18], [19].

A. Effective Bandwidth and QoS Provisioning

Figs. 5-6 illustrate how the effective bandwidth (in packets per second) changes with the QoS requirements and traffic parameters, where the maximum number of active mobile users in each cell is $N = 10$. The simulation results are presented to verify the linear approximation (9). The solid line is the interpolation of the simulation points denoted by symbols x, $\star$, $\circ$, and $\triangle$. The effective bandwidth decreases as the required delay bound increases, because the cell with the same capacity can support more calls with less stringent QoS requirements. The effective bandwidth increases when the packet arrival rate increases. Because the increased packet volume results in an increase of packet transmission delay, the cell has to reduce the number of simultaneous calls to limit the aggregate traffic. When σ_b^2 increases, the packet arrival process from each user becomes more bursty, which reduces the statistical multiplexing gain and results in an increase of the packet transmission delay. As a result, the number of simultaneous calls should be reduced to maintain the same QoS.

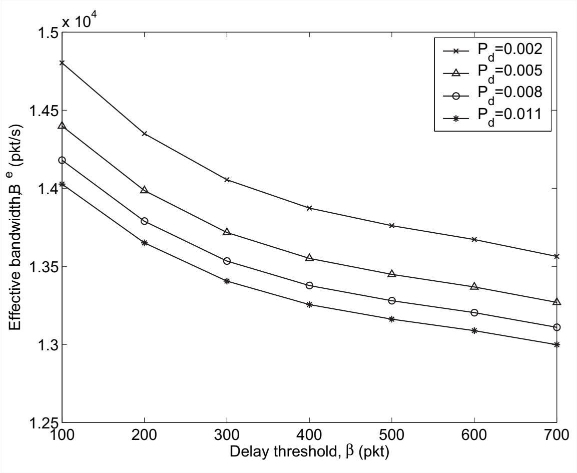


Fig. 5. Effect of the QoS requirements on the effective bandwidth.

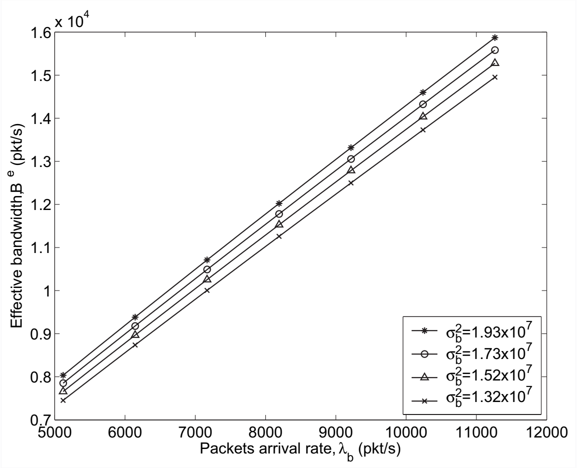
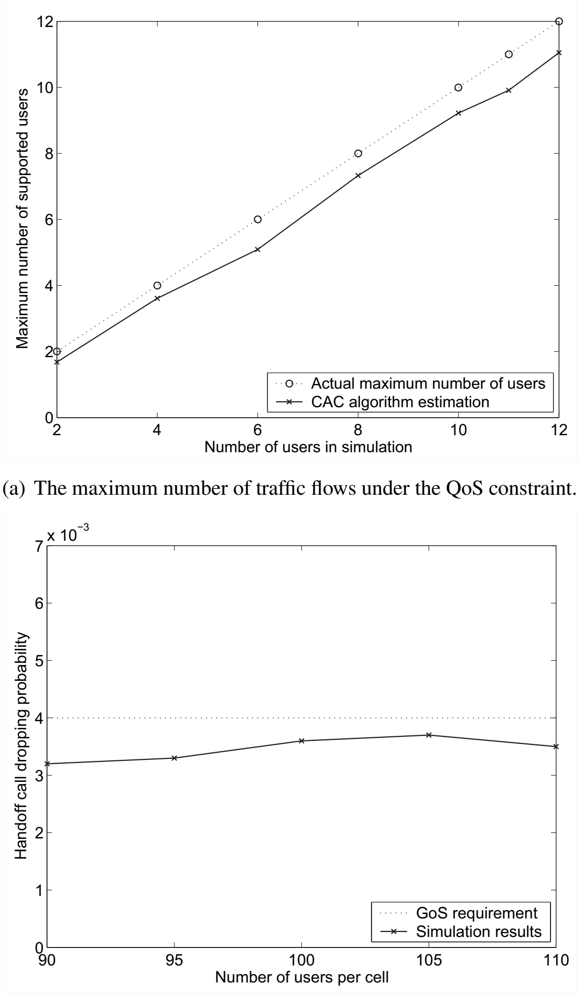


Fig. 6. Effect of the traffic parameters on the effective bandwidth.

Fig. 7 shows the maximum number of users that can be supported in each cell, and the handoff call dropping probability under the constraint that the QoS of all the users are satisfied. In Fig. 7(a), the CAC algorithm estimation of the maximum number of active users, which is the analytical results of the CAC scheme, are represented by the solid line. The results are compared with the actual maximum number of

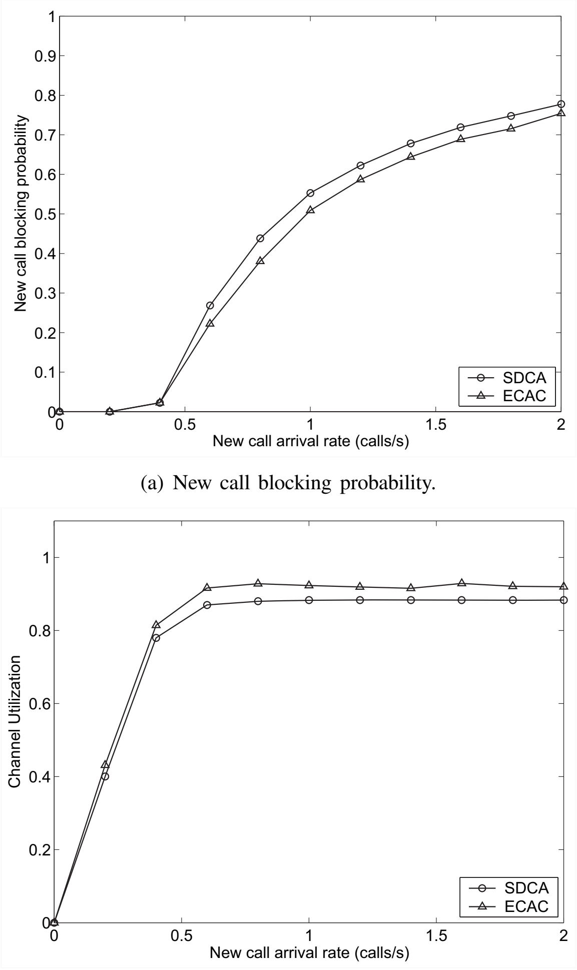


(a) The maximum number of traffic flows under the QoS constraint.

(b) The handoff call dropping probability.

Fig. 7. QoS and GoS provisioning.

active users in the simulation (represented by the dot line). The lines are the interpolations of the estimation points denoted by symbol x and the corresponding actual points denoted by symbol o, respectively. The analytical results are calculated based on (10) and the packet transmission delay probabilities obtained from simulations. It is observed that the analytical results agree with the simulation results, but are slightly more conservative. The reasons for the deviation are as follows: first, the capacity given in (6) is a lower bound and it is accurate when the number of traffic flows is large; second, there exists a linearization error in $C_r(n) \approx C_R(l,n)$; and last, there exits a simulation error in generating the self-similar traffic flows using the RMD algorithm. Fig. 7(b) shows the performance of GoS provisioning in terms of the handoff call dropping probability. The required upper bound is 4×10^{-3}. Simulations are taken on a 19-cell system where the cell diameter to user movement speed ratio is 200 s, and user call parameters are $\lambda = 0.001, \mu_c = 1/250$. The actual handoff call dropping probability meets the requirement when the traffic load in the system varies.



(a) New call blocking probability.

(b) Channel utilization.

Fig. 8. Comparison of the newly proposed ECAC with SDCA of [2].

B. GoS Provisioning Subject to QoS Constraint

Simulation results in Figs. 8-12 show the performance of our CAC scheme under different scenarios. The per call based CAC using (21) is simulated. To reduce the computation complexity, the integration interval (call duration) is approximated by $[0,T]$ and the segmented approximation is used to compute the integration. In fact, each thread of $h_{ij}(t)$ and $G_i(x,t)$ can be approximated by a lookup table. Therefore, the computation speed can be further improved at the cost of large computational memory. The basic system parameters are (unless otherwise specified): $K = 19$, $\mu_c = 1/200$ s^{-1}, $h = 1/100$ s^{-1}, $P_h = 0.01, T = 400$ s, initially the number of mobile users in each cell is $N = 100$. Simulations are executed with average new call arrival rate in each cell varying from 0.2 to 2 calls per second.

Fig. 8 compares the performance of our effective-bandwidth CAC (ECAC) scheme with that of the stable dynamic call admission control (SDCA) scheme proposed in [2] for uniformly distributed traffic, in terms of the new call blocking probability and channel utilization. The proposed ECAC outperforms SDCA in four aspects:

- The ECAC achieves a lower new call blocking probability while satisfying the QoS and GoS requirements;
- For a call arrival rate equal to or larger than 0.6 calls

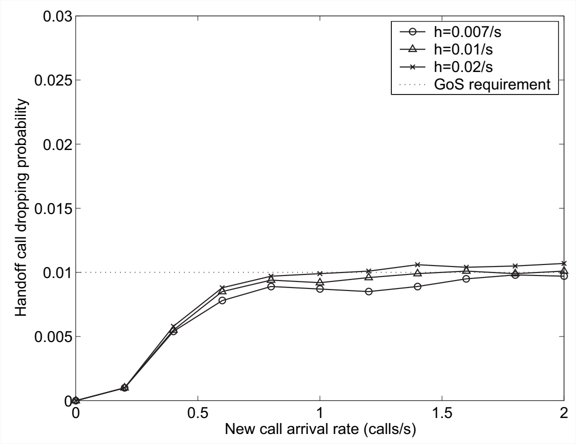


Fig. 9. Handoff call dropping probability for different user mobility (handoff rates).

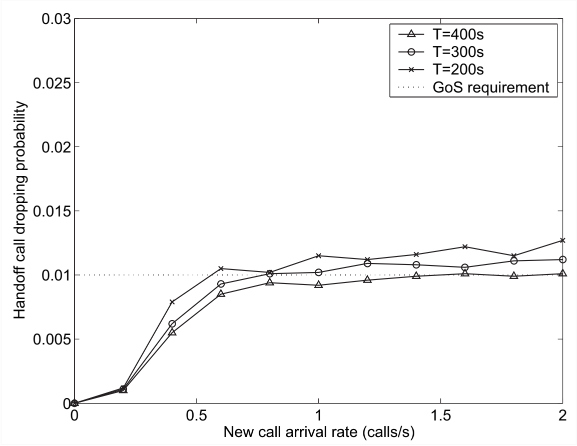


Fig. 11. Handoff call dropping probability for different control periods.

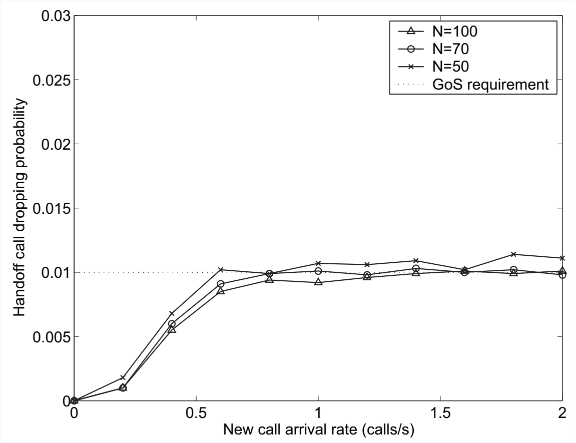


Fig. 10. Handoff call dropping probability for different initial numbers of users per cell.

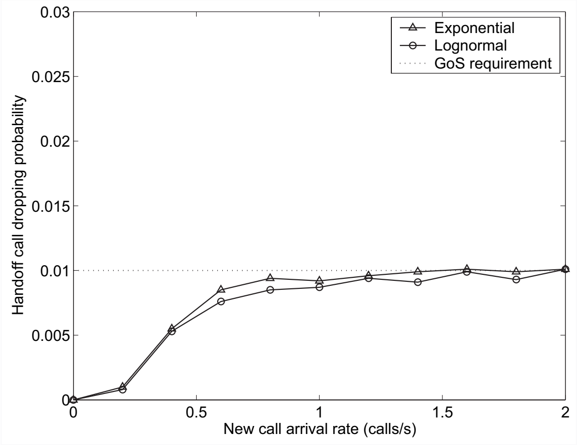


Fig. 12. Handoff call dropping probability for exponentially and lognormally distributed call duration.

per second, the ECAC scheme improves the channel utilization efficiency by approximately 4%. Taking into account that the channel utilization using the SDCA scheme is already very high, it is nontrivial to achieve the further improvement;

- The newly proposed CAC scheme guarantees both packet level and bit level QoS satisfaction, in addition to the call level GoS provisioning as proposed in [2];
- The ECAC scheme can be implemented on a per call basis in addition to that on a per cell basis.

For the GoS provisioning, the major difference between ECAC and SDCA is in the traffic estimation. The ECAC scheme fully utilizes the user mobility information and estimates the future traffic load based on the user location distribution, while the estimation in SDCA is developed from the diffusion equations. The improved performance of the ECAC scheme comes from the more accurate estimation of the dynamic traffic load.

Fig. 9 shows the handoff call dropping probability for different handoff rates. The QoS requirement is roughly satisfied, where a lower handoff rate results in a slightly lower handoff dropping probability. The result shows that the proposed CAC scheme works well under a medium to lower handoff traffic load, because in calculating $h_{ij}(t)$ and $a_{ik}(t)$ we neglect the possibility of more than 3 handoffs for each call in order to reduce the algorithm complexity. With a larger h value, each call is more likely to have many (more than 3) handoffs. Fig. 10 shows the handoff call dropping probability for different initial number of mobile users in each cell, N. The estimation of the cell overload probability is based on the central limit theorem, which is more accurate for a larger number of mobile users. As a result, the GoS provisioning is more accurate for a larger N value because of the better statistical multiplexing. The effect of the control period T on the handoff call dropping probability is shown in Fig. 11. In simulations, the integration interval in (21) is approximated by the control period T. Based on the mean call duration, T is chosen to balance the algorithm precision requirement and computation complexity. With a longer control period, more system status information and more history information are used in the computation, resulting in more accurate estimation of the traffic load and better performance.

Studies show that the lognormally distributed call duration is indeed a more accurate model for the data traffic than the conventional negative exponential distribution [20], [21].

Fig. 12 compares the handoff call dropping probabilities under different statistical models for the call duration: exponential distribution and lognormal distribution, both having the same mean and variance, respectively. It is observed that the proposed CAC scheme works well with the lognormally distributed call duration with a slightly more conservative result. With a lognormal distribution, there are a small number of calls having an extra-long call duration. These extra-long calls in general have more handoffs. However, when a new call blocking or handoff call dropping event happens to any of the extra-long calls, the actual handoff call arrival rate will be affected and reduced. Therefore, the system is more likely to have a lower handoff call dropping probability. The proposed CAC scheme works well in the lognormal call duration environment.

VII. Conclusions

In this paper, we have proposed a new CAC scheme for a CDMA cellular system supporting heterogeneous self-similar data traffic, which guarantees both the QoS requirement (on the bit error rate and packet transmission delay) and the GoS requirement (on the handoff call dropping probability). We use the effective bandwidth approach for the QoS provisioning, which takes into account the interference limited capacity of the CDMA system, dynamics of the traffic load, and statistical multiplexing. Based on the random walk user mobility and traffic models, we derive the handoff probability and cell overload probability. The GoS provisioning under the QoS constraint is supported via resource reservation based on traffic load prediction and user movement pattern. The GoS/QoS provisioning can be ensured either on a per cell basis or on a per call basis. Simulation results demonstrate that the proposed CAC scheme performs well under different system parameters and call duration models.

Acknowledgements

The authors would like to thank the anonymous reviewers for their constructive comments which improve the presentation of this paper.

References

[1] J. S. Evans and D. Everitt, “Effective bandwidth-based admission control for multiservice CDMA cellular networks,” *IEEE Trans. Veh. Technol.*, vol. 48, no. 1, pp. 36–46, Jan. 1999.

[2] S. Wu, K. Y. M. Wong, and B. Li, “A dynamic call admission policy with precision QoS guarantee using stochastic control for mobile wireless networks,” *IEEE/ACM Trans. Networking*, vol. 10, no. 2, pp. 257–271, 2002.

[3] O. T. W. Yu and V. C. M. Leung, “Adaptive resource allocation for prioritized call admission over an ATM-based wireless PCN,” *IEEE J. Select. Areas Commun.*, vol. 15, pp. 1208–1225, Sept. 1997.

[4] W. S. Jeon and D. G. Jeong, “Call admission for mobile multimedia communications with traffic asymmetry between uplink and downlink,” *IEEE Trans. Veh. Technol.*, vol. 50, no. 1, pp. 59–66, Jan. 2001.

[5] C.-J. Ho, J. A. Copeland, C.-T. Lea, and G. L. Stuber, “On call admission control in DS/CDMA cellular networks,” *IEEE Trans. Veh. Technol.*, vol. 50, no. 6, pp. 1328–1343, Nov. 2001.

[6] Y. Fang and Y. Zhang, “Call admission control schemes and performance analysis in wireless mobile networks,” *IEEE Trans. Veh. Technol.*, vol. 51, no. 2, pp. 371–382, Mar. 2002.

[7] B. M. Epstein and M. Schwartz, “Predictive QoS-based admission control for multiclass traffic in cellular wireless networks,” *IEEE J. Select. Areas Commun.*, vol. 18, no. 3, pp. 523–534, Mar. 2000.

[8] V. Paxson, and S. Floyd, “Wide area traffic: The failure of Poisson modeling,” *IEEE/ACM Trans. Networking*, vol. 3, no. 3, pp. 226–244, June 1995.

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

[10] M. E. Crovella, and A. Bestavros, “Self-similarity in world wide web traffic: evidence and possible causes,” *IEEE/ACM Trans. Networking*, vol. 5, no. 6, pp. 835–846, Dec. 1997.

[11] M. Jiang, M. Nikolic, S. Hardy, and L. Trajkovic, ”Impact of self-similarity on wireless data network performance,” in *Proc. IEEE ICC’01*, pp. 477–487.

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

[13] F. P. Kelly, “Notes on effective bandwidths,” in *Stochastic Networks: Theory and Applications*, Oxford University Press, 1996.

[14] I. Norros, “On the use of fractional Brownian motion in the theory of connectionless networks,” *IEEE J. Select. Areas Commun.*, vol. 13, no. 6, pp. 953–962, Aug. 1995.

[15] W. E. Leland, M. S. Taqqu, W. Willinger, and D. V. Wilson, “On the self-similar nature of Ethernet traffic (extended version),” *IEEE/ACM Trans. Networking*, vol. 2, pp. 1–15, 1994.

[16] S. M. Ross, *Introduction to Probability Models*, 6th ed., Academic Press, 1997, pp. 257–258.

[17] K. S. Gilhousen, I. M. Jacobs, R. Padovani, A. J. Viterbi, L. A. Weaver, and C. E. Wheatley, “On the capacity of a cellular CDMA system,” *IEEE Trans. Veh. Technol.*, vol. 40, pp. 303–312, May 1991.

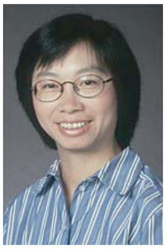
[18] W.-C. Lau, A. Erramilli, J. L. Wang, and W. Willinger, “Self-similar traffic generation: the random midpoint displacement algorithm and its properties,” in *Proc. IEEE ICC’95*, pp. 466–472.

[19] I. Norros, P. Mannersalo, and J. L. Wang, “Simulation of fractional Brownian motion with conditional random displacement,” *Advances in Performance Analysis*, vol. 2, no. 1, pp. 77–101, 1999.

[20] C. Jedrzycki and V. C. M. Leung, “Probability distributions of channel holding time in cellular telephony systems,” in *Proc. IEEE VTC’96*, pp. 247–251.

[21] J. Jordan and F. Barcelo, “Statistical modeling of transmission holding time in PAMR systems,” in *Proc. IEEE Globecom’97*, pp. 121–125.

Lei Wang (S’02) received the B.S.E. and M.S.E. degrees in electrical engineering from Huazhong University of Science and Technology, Wuhan, China, in 1994 and 1997, respectively. He is currently working toward the Ph.D. degree in electrical engineering at the University of Waterloo, Waterloo, ON, Canada. His research interests include admission control, resource management, and QoS provisioning in wireless networks.



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