Realistic Cell-Oriented Adaptive Admission Control for QoS Support in Wireless Multimedia Networks

Jae Young Lee, Jin-Ghoo Choi, Kihong Park, and Saewoong Bahk, *Member, IEEE*

***Abstract*—An important quality-of-service (QoS) issue in wireless multimedia networks is how to control handoff drops. In this paper, we propose admission-control algorithms that adaptively control the admission threshold in each cell in order to keep the handoff-dropping probability below a predefined level. The admission threshold is dynamically adjusted based on handoff-dropping events. We first present a simple admission-control scheme that brings out an important performance evaluation criterion—intercell fairness—and serves as a reference point. We then investigate the intercell unfairness problem and develop two enhanced schemes to overcome this problem. The performance of these protocols is benchmarked and compared with other competitive schemes. The results indicate that our schemes perform very well while, in addition, achieving significantly reduced complexity and signaling load.**

***Index Terms*—Adaptive control, handoff, intercell unfairness, probabilistic quality of service (QoS) guarantee.**

I. Introduction

THE MOBILE USER population has been growing at a rapid rate. More recently, the demand for multimedia applications requiring high bandwidth, such as video, image, and interactive Web information, has increased. The current trend in wireless networks is to decrease cell sizes—microcells or picocells—to provide higher capacity and accommodate more users in a given area. Small cell sizes, however, cause more frequent handoffs, resulting in increased variability and burstiness of network load and traffic conditions. This, in turn, has amplified the difficulties associated with quality-of-service (QoS) provisioning in wireless networks [1].

An important QoS issue in wireless networks is how to control handoff drops. When a mobile moves into an adjacent cell during a session, a handoff occurs and the mobile can communicate continuously through the new base station (BS). The handoff could fail, however, if available bandwidth in the new cell is insufficient, which leads to handoff drops. Handoff drops are generally considered to be more detrimental to network performance than new call blocks. Thus, strategies for prioritizing handoff calls vis-à-vis new calls are needed, for instance, by maintaining bandwidth reserves for future handoffs. The concept of bandwidth reservation for handoffs was first introduced in the mid-1980s [2]. Since then, various strategies that assign priority to handoffs have been studied [3]–[19].

Ideally, no handoff drops are desirable. This, however, requires that the network reserve bandwidth in all cells that a mobile might pass through, resulting in potentially lower utilization and/or higher new call blocking probability (P_b). To achieve increased efficiency, several approaches have advocated providing *probabilistic* QoS guarantees by keeping the handoff-dropping probability (P_d) below a certain level [8], [11], [13], [14], [17], [18].

In [8], the admission threshold needed to satisfy a QoS constraint is calculated based on the number of users in the current cell and adjacent cells, given the probability that a mobile would handoff within some time interval. A drawback of this scheme is that it does not specify how to predict user mobility, which plays an essential role in the proposed method. Moreover, the model assumes exponential distribution of the cell-residence time. In [11], a technique to compute the reserved bandwidth to maintain P_d within a specified level is proposed. However, as with [8], it is assumed that the cell residence time is exponential. This weakens the advanced conclusions since it has been shown that, in practice, cell residence time may not be exponentially distributed [21].

In [13], the shadow cluster concept has been used to estimate future resource requirements and perform admission control in order to limit P_d. In this method, mobiles inform neighboring BSs of their bandwidth requirements and movement patterns at the call setup time. Based on this information, BSs predict future demands and admit only the mobiles that can be supported adequately. The drawbacks of this scheme are that precise user mobility needs to be known *a priori*—an impractical assumption—and requires the exchange of a large number of messages among BSs, which can exert a significant overhead cost in wireless networks.

A method for predicting user mobility has been presented recently in [14], which uses a predictive bandwidth-reservation scheme to provide probabilistic QoS guarantees. This method is based on the observed history of mobility information, which is used to calculate the reserved bandwidth. Although this method does not rely on potentially unrealistic assumptions, it suffers the drawback of high complexity and implementation overhead [20]. Another scheme, in [17], also uses user mobility prediction to guarantee the handoff-dropping probability below a target value, where the prediction algorithm is derived by data-compression techniques. Bandwidth is reserved based on the mobility prediction and, in turn, the reservation level is adaptively controlled by monitoring handoff-dropping events.

Manuscript received March 14, 2002; revised September 4, 2002 and November 21, 2002. This work was supported by the University IT Research Supporting Program under the Ministry of Information & Communication of Korea.

J. Y. Lee, J.-G. Choi, and S. Bahk are with the School of Electrical Engineering and INMC **AU: PLEASE SPELL OUT INMCAU**, Seoul National University, Seoul, Korea (e-mail: jylee@netlab.snu.ac.kr, cjk@netlab.snu.ac.kr, sbahk@netlab.snu.ac.kr).

K. Park is with the Department of Computer Sciences, Purdue University, West Lafayette, IN 47907 USA (e-mail: park@cs.purdue.edu).

Digital Object Identifier 10.1109/TVT.2003.810975

0018-9545/03$17.00 © 2003 IEEE

IEEE Proof

Prediction-based schemes, such as in [14] and [17], use complex mobility prediction techniques, but the actual reservation level is indirectly adjusted according to dropped handoff events. As they do not rely on prediction for their final decision, we are motivated to exclude the mobility prediction part from the algorithms. This means that we try to obtain the optimal reserved bandwidth by directly controlling the reservation level without access to user mobility information.

We emphasize that several previously proposed schemes are based on user mobility information. We call these *mobile-oriented* reservation schemes. If the design goal is to minimize handoff drops, user mobility information must be used to reserve bandwidth by predicting mobiles' handoff times and next cell movements. However, if the goal is to keep P_d below a certain target level, this may be effectively achieved without access to user mobility information since a handoff drop is, to a large extent, a *cell-oriented* event: a handoff drop occurs when a cell is overloaded, which can be controlled for a range of $P_d > 0$ by dynamic control of reserved bandwidth.

A practical cell-oriented scheme was introduced in [18]. This method determines the amount of reserved bandwidth by the largest of all the requested bandwidths from adjacent cells. After some bandwidth is reserved, its value is dynamically adjusted at each cell to keep P_d below a target value. This, however, gives rise to a potentially serious intercell unfairness problem, which can significantly impede system utilization and performance.

In this paper, we consider intercell fairness with respect to its role as a relevant performance evaluation criterion for adaptive admission control and handoff in wireless mobile environments. In tandem, we propose new algorithms that effectively address the intercell unfairness problem in the context of cell-oriented adaptive admission control, which does not require user mobility information. We combine a simple admission test with an adaptive algorithm to adjust the admission threshold in each cell. The proposed scheme is able to provide probabilistic QoS guarantees while achieving high channel utilization at the same time. Since our protocol is simply based on handoff-dropping events at each cell—and not on individual calls' mobility—it has significantly lower complexity than mobile-oriented methods.

The rest of this paper is organized as follows. Section II describes the system model and presents a simple admission-control scheme, which introduces the intercell fairness issue and serves as a reference point. In Section III, we present an analysis of the intercell unfairness problem. In Section IV, we consider two enhanced protocols to overcome the unfairness problem. In Section V, we give a detailed discussion of existing practical methods advanced in [14] and [18], putting our schemes in a comparative perspective. Section VI presents simulation results of our three proposed schemes and compares our best one with existing protocols. We conclude with a discussion of our approach and results.

II. System Model and Simple Admission Control

We consider a mobile network with a cellular infrastructure. We assume that the system uses a fixed channel allocation (FCA) scheme and a cell, i, has capacity $C(i)$. Also, the service model accommodates multiple classes of traffic (e.g., voice and video). Let BU denote the bandwidth unit and assume that 1 BU is required by a voice call.

```
S_P = ⌈1/P_QoS⌉; L_P = S_P;
S_H = 0; S_HD = 0; L_H = 0; L_HD = 0; T = T_init;
WHILE (time increases)
    IF(a mobile handoffs into the current cell) THEN
        S_H = S_H + 1; L_H = L_H + 1;
        IF(it is dropped) THEN
            S_HD = S_HD + 1; L_HD = L_HD + 1;
            IF(L_HD > 1) THEN
                L_P = L_P + S_P;
                T = max(T − d, T_min);
        IF(S_H == S_P) THEN
            IF(S_HD < 1) THEN
                T = min(T + d, T_max);
            S_H = 0; S_HD = 0;
            IF(L_H == L_P) THEN
                L_H = 0; L_HD = 0; L_P = S_P;
```

Fig. 1. Adaptive-control algorithm (A1).

A. Admission-Control Test

A new call setup request is accepted into cell i through the following admission test named T1:

$$C_a(i) + B_{\text{new}} \leq T(i) \tag{1}$$

where $C_a(i)$ is the allocated bandwidth of cell i, B_{new} is the required bandwidth of the new call ,and $T(i)$ is the admission threshold of cell i. The latter satisfies $T(i) \leq C(i)$. This indicates that a new call request is accepted if the allocated bandwidth plus the new call bandwidth is less than or equal to the admission threshold. In the case of a handoff call, it is accepted if there is bandwidth available to accept the handoff call with bandwidth requirement B_{handoff}. That is

$$C_a(i) + B_{\text{handoff}} \leq C(i). \tag{2}$$

This admission test gives priority to handoff calls over new calls and $C(i) - T(i)$ can be interpreted as the reserved bandwidth for handoff calls at cell i.

B. Adaptive-Control Algorithm to Adjust the Admission Threshold

There might exist an optimal steady-state admission threshold $T_{\text{opt}}(i)$ at cell i for a specific traffic load and user mobility.[1] Here we use the term "optimal" in the sense of maximizing (minimizing) utilization (P_b) while keeping P_d below a target value P_{QoS}. If the admission threshold T is below T_{opt}, utilization can be improved by increasing T. On the other hand, if T is above T_{opt}, T must be decreased to keep P_d below P_{QoS}. The problem is how to adjust T as close as possible to, but not over, T_{opt}. First, we describe an adaptive algorithm to adjust the admission threshold based on monitored handoff drops at each cell. Fig. 1 shows algorithm **A1**, executed by the BS of each cell in a distributed manner. Here, we use $T_{\min}$ and $T_{\max}$ to represent the range of T, which is given by $0 \leq T_{\min} < T_{\max} \leq C$.

[1]We will drop the index i for notational simplicity when the reference is clear

IEEE Proof

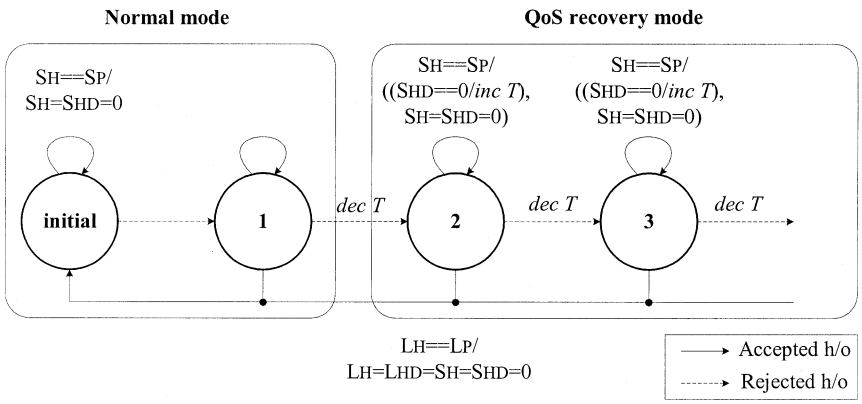


Fig. 2. Illustration of the behavior of **A1**.

The main idea in the adaptation is to monitor handoff-dropping events over both the short- and long-term. The objective of long-term monitoring is to keep P_d below P_{QoS}, whereas short-term monitoring is used to maximize utilization. The short-term period[2] S_P is given by the number of handoff attempts, $\lceil 1/P_{\mathrm{QoS}} \rceil$. The counts for the short-term handoff attempts, S_H, and handoff drops, S_{HD}, are reset to 0 at the start of each period. The long-term period, L_P, is determined by handoff attempts as $S_P \times \max(L_{HD}, 1)$, where L_{HD} is the count of the long-term handoff drops. The counts for the long-term handoff attempts, L_H, and handoff drops, L_{HD}, are reset to 0 at the start of each long-term period. At initialization, L_P is set to S_P. The system is always in one of two modes: *normal mode* or *QoS recovery mode*.

We show the behavior of **A1** by the state diagram in Fig. 2. The state transition occurs whenever a handoff is requested. The solid (or dotted) arrow represents state transition when a handoff request is accepted (or rejected). In the diagram, "**IF** *condition* **THEN** *action*" is briefly denoted as "*condition*/*action*." An accepted (or rejected) handoff request increases S_H and L_H (or S_H, L_H, S_{HD}, and L_{HD}) by one, respectively. In the *initial* state, $S_H = S_{HD} = L_H = L_{HD} = 0$ and L_P is equal to S_P. In the state i, L_{HD} is i and L_P is i times S_P. When the system is in the normal mode, the handoff drop probability is less than or equal to the required value P_{QoS}. In the QoS recovery mode, on the contrary, since the drop probability temporarily exceeds P_{QoS}, we try to satisfy the required QoS by decreasing T. The following gives detailed explanations for the two modes.

1) Normal Mode: When one or no handoff drop occurs for the first L_P ($= S_P$) handoff attempts, the system is in normal mode. In normal mode, the dropping probability is

$$P_d = \frac{L_{HD}}{L_H} \leq \frac{1}{L_P} = \frac{1}{S_P} \approx P_{\mathrm{QoS}}. \tag{3}$$

Hence, P_d during this period is kept below P_{QoS}. If no handoff drop has occurred, it is likely that $T < T_{\mathrm{opt}}$. So T is increased by a predetermined step size d. In normal mode, the long-term period ends with the short-term period and the system state immediately goes to the initial state.

2) QoS Recovery Mode: When more than one handoff drop occurs for the first L_P handoff attempts, the system enters the QoS recovery mode. In this mode, L_P is increased by S_P and T is decreased by d whenever a handoff drop occurs. QoS recovery mode goes to the initial state when $L_H = L_P$. While the first period shows higher short-term dropping probability than the target value, the overall long-term dropping probability is maintained at the target value

$$P_d = \frac{L_{HD}}{L_H} = \frac{L_{HD}}{L_P} = \frac{L_{HD}}{S_P L_{HD}} = \frac{1}{S_P} \approx P_{\mathrm{QoS}}. \tag{4}$$

This is made possible by decreasing T whenever a handoff drop occurs. By doing so, T will approach T_{opt} within some bounded time. By adopting this conservative policy, QoS recovery mode ends within a bounded time.

Another policy to be considered in QoS recovery mode is regarding when to increase T. It is possible for a conservative policy to increase T only after a long-term period ends. We do not, however, adopt this policy because T can be decreased too much for the following reasons: (1) the effect of the decreased threshold is not shown immediately, because it takes some time for existing calls to depart the residing cell either by handoff or by completion and (2) handoffs may occur in a burst even when $T \approx T_{\mathrm{opt}}$. An alternative, more aggressive policy in A1 is to increase T when no handoff drop occurs during a short-term period, even if a long-term period has not ended. Combining the conservative threshold-decreasing policy with this aggressive increasing policy results in a slightly less conservative policy. Note that T is still decreased when one handoff drop occurs during a short-term period in QoS recovery mode. By doing so, we are trying to adjust T as close to, but not exceeding, T_{opt}.

Now let us consider the increment/decrement step size d. If it is too large, it may result in an over-reaction, i.e., oscillation between over- and underreservation. If it is too small, on the other hand, it may result in an under-reaction. Hence, the magnitude of d must be carefully chosen. It is possible to have a different step size for each class. We, however, apply an equal step size to all classes, considering that we should reserve bandwidth for future handoff calls but do not know which class of calls will

[2]The period is determined not by the length of time, but by the number of *events* that the system monitors.

IEEE Proof

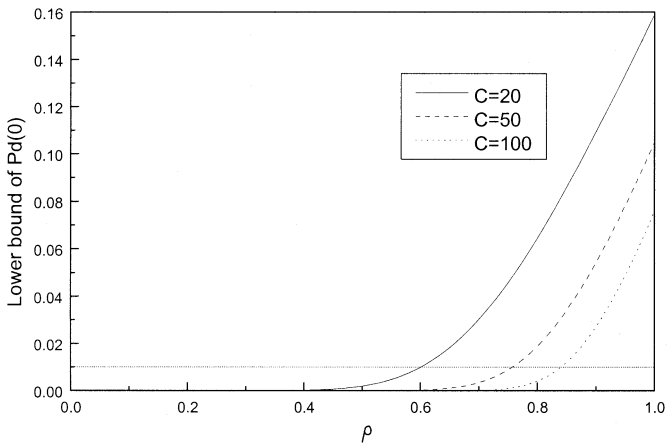


Fig. 3. Lower bound of the dropping probability.

arrive next. In general, if there are M classes of traffic, the step size

$$d = \sum_{i=1}^{M} F_i B_i \tag{5}$$

is found to be a reasonable choice through extensive simulations where F_i is the fraction of class i call requests and B_i is the required bandwidth of a class i call. F_i may be determined from the traffic history available at a BS. If a particular traffic class k is dominant, d is almost the same as B_k. The admission-control scheme, which uses the admission test T1 with the adaptive algorithm A1, will henceforth be referred to as **AC1**.

AC1 can be, with minor modifications, applied for non-real-time (or data) traffic, though it is mainly designed for real-time traffic. A data call can also require a certain amount of bandwidth, say the required minimum bandwidth of $\min_{bw}$ and the (wishing) peak rate of $\max_{bw}$, to achieve its performance. It must receive at least the bandwidth of $\min_{bw}$ once accepted and generated packets over this rate will be buffered at either the BS or mobile terminal. If there exists more bandwidth available, the user can transmit data at the rate of up to $\max_{bw}$. At present, data traffic usually uses the best effort service, which does not require any specific bandwidth, i.e., $\min_{bw} = 0$ and $\max_{bw} =$ unknown. In this case, when there is no free bandwidth, the network can accept the call instead of dropping it. The network just keeps the connection information and allows the data call to increase its rate whenever the channel is available.

When real-time and data traffic are competing, the network allocates $\min_{bw}$ for data traffic. A new data call will be accepted when condition (1) is met, where B_{new} is $\min_{bw}$ of the new data call. $C_a(i)$ is the sum of allocated bandwidth for real-time traffic and $\min_{bw}$ for data traffic. Note that $C_a(i)$ does not reflect the actual link utilization, because it just counts $\min_{bw}$ of data calls. A handoff data call is accepted when condition (2) is met, where B_{handoff} is $\min_{bw}$ of the handoff data call. $C_a(i)$ is the same as the above. If $\min_{bw}$ for the handoff call is not available, the call will be dropped and the handoff drop counts, S_{HD} and L_{HD}, will be increased by one, respectively, since the reserved bandwidth is insufficient. This indicates that $\min_{bw}$ for data traffic has the same meaning as the required bandwidth for real-time traffic. When extra bandwidth is available, it will be used by currently active data calls and later occupied by newly arriving real-time or data traffic of $\min_{bw}$.

As there exist many possible mechanisms of scheduling and resource management for combining data and real-time traffic, we do not investigate them further in this paper.

```
                    ⋮
3. WHILE (time increases)
*      IF(receive decrease_T message) THEN
*          IF(T̃ > avg. T̃ of adjacent cells) THEN
*              T = max(T − d, T_min);
*      IF(receive increase_T message) THEN
*          IF(T̃ < avg. T̃ of adjacent cells and QoS_state==IN)
THEN
*              T = min(T + d, T_max);
                    ⋮
8.             IF(L_HD > 1) THEN
9.                 L_P = L_P + S_P;
10.                T = max(T − d, T_min);
*                  send decrease_T messages to adjacent BSs;
*                  QoS_state=OUT;
11.        IF(S_H == S_P) THEN
12.            IF(S_HD < 1) THEN
13.                T = min(T + d, T_max);
*                  send increase_T messages to adjacent BSs;
14.            S_H = 0; S_HD = 0;
15.            IF(L_H == L_P) THEN
16.                L_H = 0; L_HD = 0; L_P = S_P;
*                  QoS_state=IN;
```

Fig. 4. Enhanced adaptive-control algorithm (A2).

```
IF P_d ≠ 0 THEN
   IF P_d ≥ thres_up_1 × P_QoS THEN
      R = min(up_1 × R, R_max)
      /* increase the size of reserved bandwidth by up_1(> 1) */
   ELSE
      IF P_d ≤ thres_down_1 × P_QoS THEN
         R = down_1 × R
            /* decrease the size of reserved bandwidth by down_1
(0 < down_1 < 1) */
ELSE /* P_d = 0 */
   IF U_R ≥ thres_up_2 THEN
      R = min(up_2 × R, R_max)
      /* increase the size of reserved bandwidth by up_2(> 1) */
   ELSE
      IF U_R ≤ thres_down_2 THEN
         R = down_2 × R
            /* decrease the size of reserved bandwidth by down_2
(0 < down_2 < 1) */
```

Fig. 5. Adaptive-control algorithm of OKS98.

III. Intercell Unfairness Problem

When the offered load is light or the user mobility low, AC1 works well. However, when the offered load is heavy and the user mobility is high, an undesirable situation can happen. When a BS dynamically adjusts its admission threshold regardless of the state of its adjacent base stations, as in AC1, an intercell unfairness problem arises, which can adversely impact performance [14]. In this section, we analyze this problem more closely. It is important to note that this is a universal

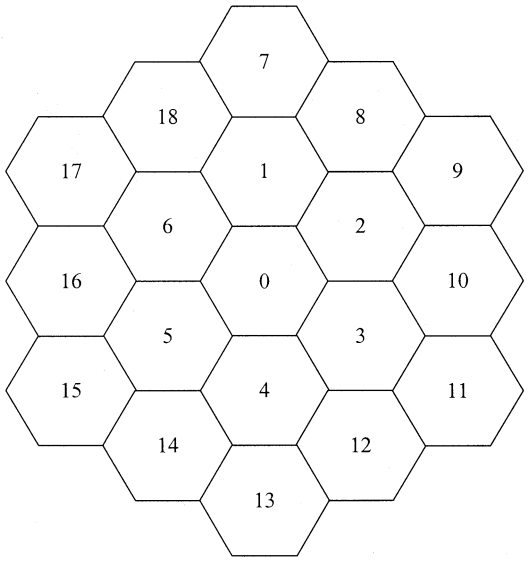
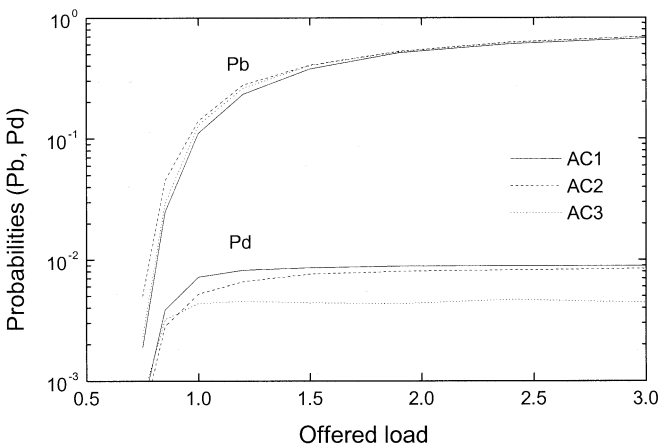
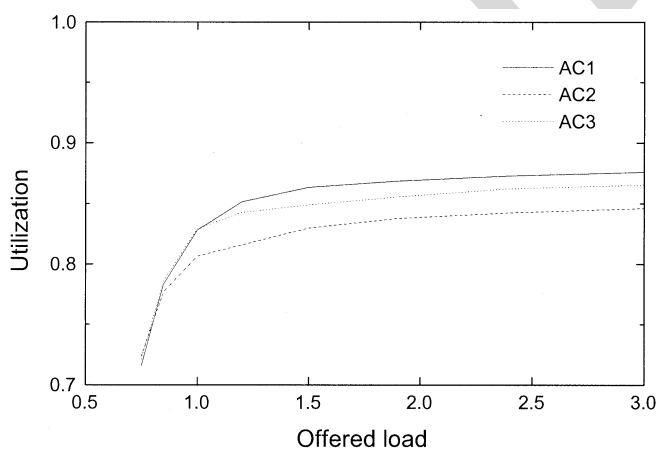


Fig. 6. Simulated cellular network topology.



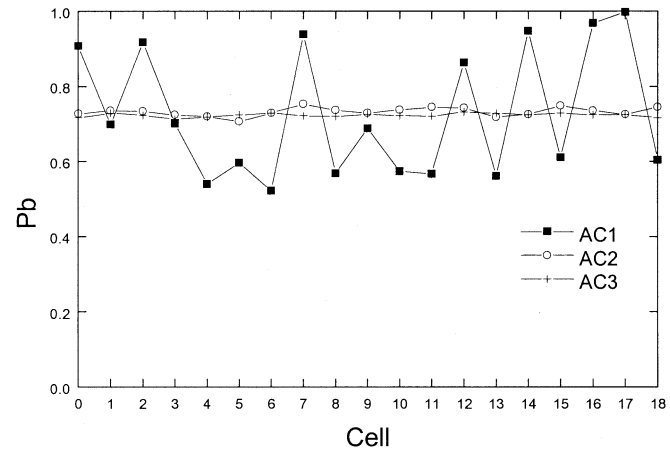
(a)



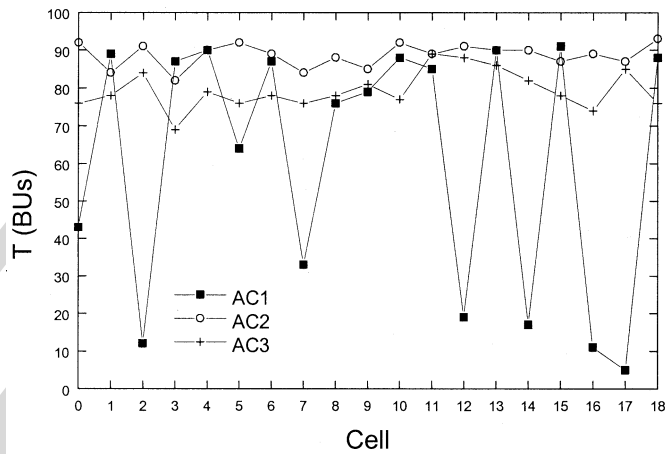
(b)

Fig. 7. Comparison of AC1, AC2, and AC3 with high mobility and $F_1 = 1.0$.

phenomenon that is shared by all schemes that dynamically adjust the admission threshold (i.e., reserved bandwidth) [8], [11], [14], [18].



(a)



(b)

Fig. 8. Status of each cell at the end of the simulation with offered load $= 3.0$, high mobility, $F_1 = 1.0$, and $P_{\mathrm{QoS}} = 0.001$.

Intercell unfairness is defined as the imbalance of admission thresholds among neighboring cells above and beyond those dictated by the optimal values $T_{\mathrm{opt}}(i)$. Specifically, when intercell unfairness occurs, the P_d values of some cells are not kept below P_{QoS}, even with extremely low T values, while the P_d values of other cells are kept below P_{QoS}, even with high T's. This is "unfair" to cells with low T's, because almost all new calls are blocked in those cells.

The fundamental reason for this unfairness is that a cell can be overloaded both by (1) new calls and (2) incoming handoffs calls. (1) is related to the T value of the current cell, whereas (2) is related to the T values of adjacent cells. We will show that admission-control schemes that consider only (1) are susceptible to the unfairness problem. We use the following assumptions[3] in the analysis of the intercell unfairness problem:

- a single class of traffic;
- the arrival process of new calls is Poisson with rate λ_n;
- call duration is exponentially distributed with mean $1/\mu$;
- the unencumbered cell-residence time[4] is exponentially distributed with mean $1/\gamma$;

[3]Our admission-control schemes do not make use of any of these assumptions.
[4]Cell-residence time for a call that will be handed over.

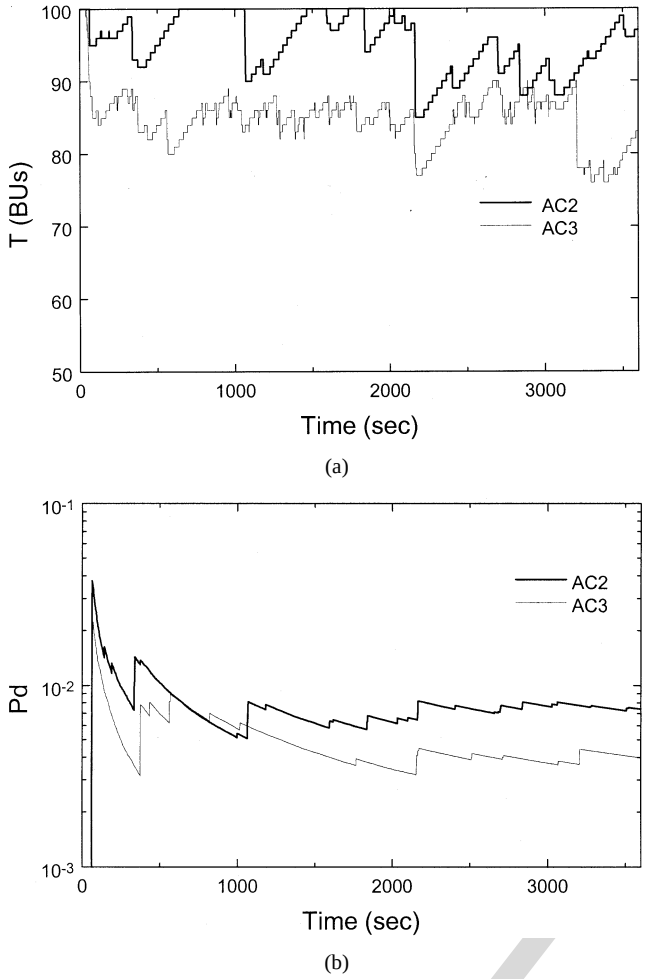


Fig. 9. Threshold and P_d versus time in cell 0 in AC2 and AC3 with offered load $= 3.0$, high mobility, and $F_1 = 1.0$.

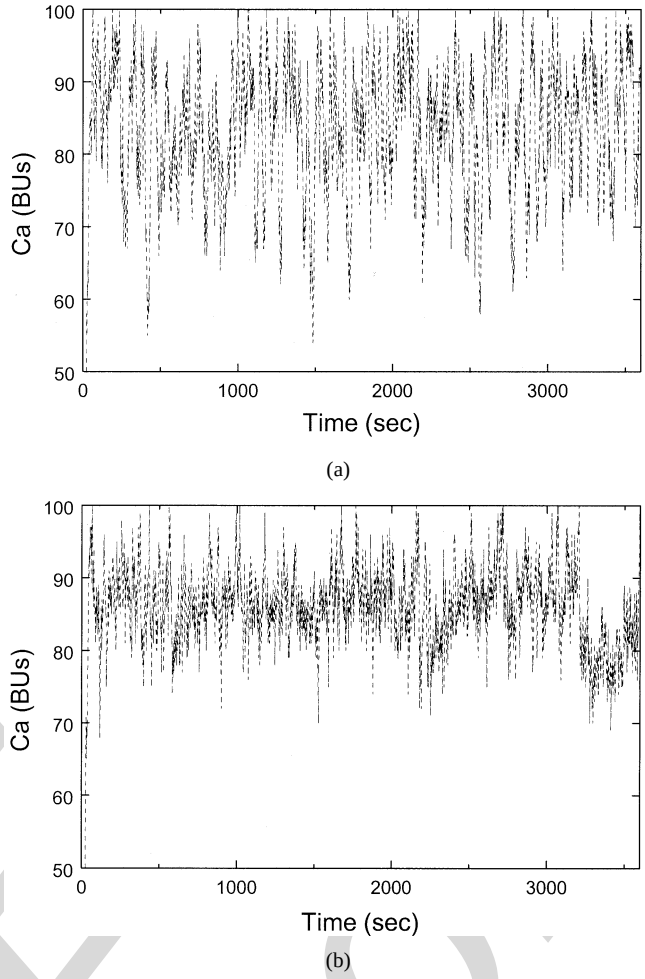


Fig. 10. C_a versus time in cell 0 in AC2 and AC3 with offered load $= 3.0$, high mobility, and $F_1 = 1.0$.

- the capacity of each cell is C.

Under these assumptions, the probability distribution function of the cell-residence time[5] is given by (see [19])

$$HC(\tau) = 1 - e^{-(\mu+\gamma)\tau}. \tag{6}$$

The probability that a mobile will handoff within time τ is

$$H(\tau) = \frac{\gamma}{\mu+\gamma}(1 - e^{-(\mu+\gamma)\tau}). \tag{7}$$

Let $P_{H_{j\to i}}$ be the probability that a mobile will handoff from an adjacent cell j into a test cell i given that a handoff occurs and let $M_j(t)$ be the number of mobiles in an adjacent cell j at times t. Then, the instantaneous handoff arrival rate in cell i at time t is

$$\begin{aligned}\lambda_h(t) &= \lim_{\tau\to 0} \frac{\sum_{j\in A_i} H(\tau) M_j(t) P_{H_{j\to i}}}{\tau}. \\ &= \sum_{j\in A_i} \gamma M_j(t) P_{H_{j\to i}}\end{aligned} \tag{8}$$

[5]The probability that a mobile will either handoff or complete in the residing cell within time τ.

Let M_j and λ_h be the time averages of $M_j(t)$ and $\lambda_h(t)$, respectively. Then

$$\lambda_h = \sum_{j\in A_i} \gamma M_j P_{H_{j\to i}}. \tag{9}$$

By approximating the handoff call arrival by a Poisson process with rate λ_h, we can model the number of calls in cell i as a continuous time Markov chain. It is straightforward to derive the dropping probability when the admission threshold is T

$$P_d(T) = \frac{a^T b^{C-T}}{C!} P_0 \tag{10}$$

where $a = (\lambda_n + \lambda_h)/(\mu+\gamma)$, $b = \lambda_h/(\mu+\gamma)$, and $P_0 = (\sum_{k=0}^{T}(a^k/k!) + \sum_{k=T+1}^{C}(a^T b^{k-T}/k!))^{-1}$. If we set T at 0, P_d is lower-bounded by

$$P_d(0) = \frac{b^C}{C!}\left(\sum_{k=0}^{C}\frac{b^k}{k!}\right)^{-1} > \frac{b^C}{C!}\left(\sum_{k=0}^{\infty}\frac{b^k}{k!}\right)^{-1} = \frac{b^C}{C!}e^{-b}. \tag{11}$$

To evaluate the effects of the utilization of adjacent cells and user mobility on this lower bound, we assume that a mobile handoffs into adjacent cells with equal probability. Then, $\lambda_h = \gamma \overline{M}_{A_i}$ in (9) where $\overline{M}_{A_i}$ is the average number of mobiles in the adjacent cells of cell i. Let $\overline{U}_{A_i}$ be the average utilization of

IEEE Proof

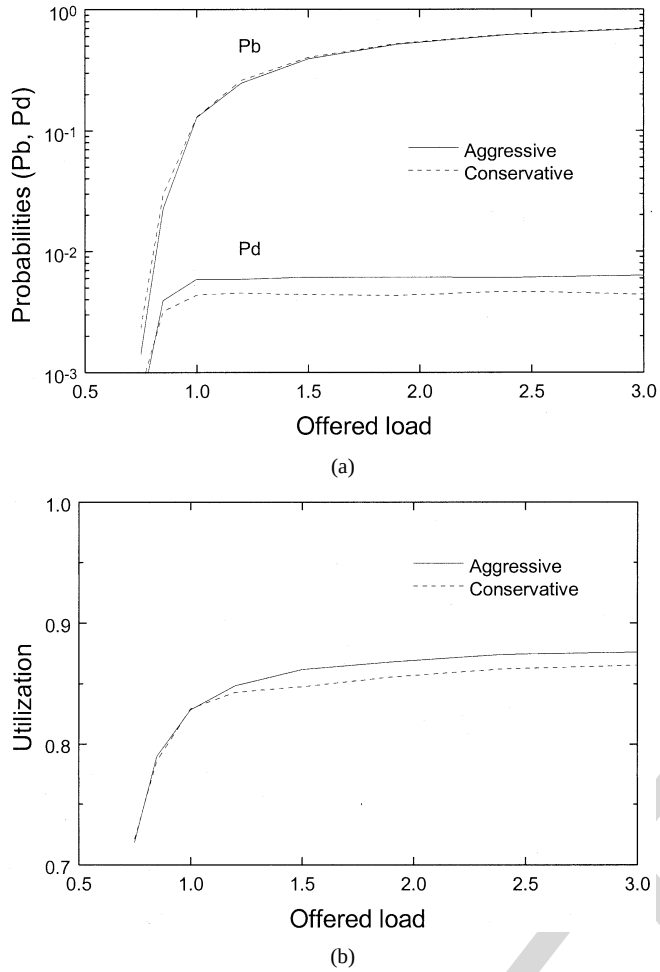


Fig. 11. Comparison of aggressive and conservative policies in AC3 with high mobility and $F_1 = 1.0$.

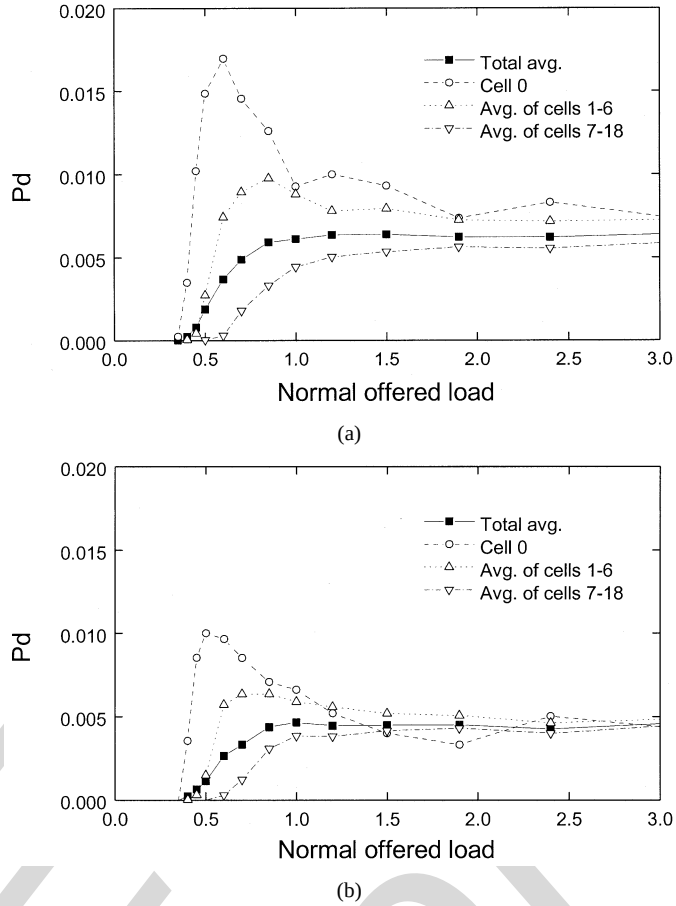


Fig. 12. Comparison of aggressive and conservative policies in AC3 with nonuniform loading, high mobility, and $F_1 = 1.0$.

adjacent cells of cell i (i.e., $\overline{U}_{A_i} = \overline{M}_{A_i}/C$). Also, let $\rho = b/C$ and $H = \gamma/\mu$ (H can be interpreted as the average number of handoffs). Then, we obtain

$$\rho = \frac{b}{C} = \frac{H\overline{U}_{A_i}}{H+1}. \tag{12}$$

The parameter ρ is related to both the user mobility and utilization of adjacent cells. The larger H (i.e., high mobility) and closer $\overline{U}_{A_i}$ approaches 1 (i.e., overloaded adjacent cells) the closer ρ approaches 1. Fig. 3 shows the lower bound of $P_d(0)$ versus ρ. When ρ exceeds a certain value, the lower bound of $P_d(0)$ exceeds some target value, say 0.01. In other words, no matter how the BS of cell i decreases its threshold, it cannot provide a probabilistic QoS guarantee.

An example scenario for the unfairness problem is given as follows. Assume high mobility and uniformly heavy load conditions for all cells. The system is at equilibrium if the threshold values of all cells are similar, while keeping P_d below a target value. In this situation, multiple handoff drops could occur in a cell, say cell i, for example, due to burstiness of handoff events. The BS of cell i will start to decrease $T(i)$. Until the decreased T becomes effective, incoming handoffs will be continuously dropped, triggering further decreases of $T(i)$. During this time, newly requested calls in cell i will be blocked, because of the overloaded cell condition and decreased threshold (i.e., $C_a(i) > T(i)$). Cell i may be still overloaded with incoming handoff calls rather than new calls. However, handoff calls have less chance to handoff than new calls, since they have already passed some cells and have limited call durations. That is, handoff calls have more chance to terminate in the residing cell than the new calls that originate from that cell. Thus, outgoing handoffs from cell i decrease, contributing to fewer handoff drops in adjacent cells A_i.[6]

Some BSs in A_i may increase their T's, thus admitting new calls. Some of these newly admitted calls will soon handoff into cell i, causing even more handoff drops in cell i and triggering further decreases of $T(i)$. Even if $T(i)$ is decreased down to 0, the system still may not keep P_d below a target value. However, in some cells of A_i, due to the decreased incoming handoffs, the P_d's may be below the target value even with high thresholds. The aforementioned qualitative sketch illustrates how the inter-cell unfairness problem can manifest itself.

Let us see this in a quantitative manner in a simple two-cell network. Each cell is denoted by 1 and 2, respectively. Then the handoff arrival rates are given by

$$\begin{aligned} \lambda_{h1} &= \gamma M_2 = \gamma M(T_2, \lambda_{h2}, \lambda_{n2} C_2) \\ \lambda_{h2} &= \gamma M_1 = \gamma M(T_1, \lambda_{h1}, \lambda_{n1}, C_1). \end{aligned} \tag{13}$$

[6]This effect can be pronounced in a one-dimensional cellular structure where a cell is adjacent to only two other cells.

IEEE Proof

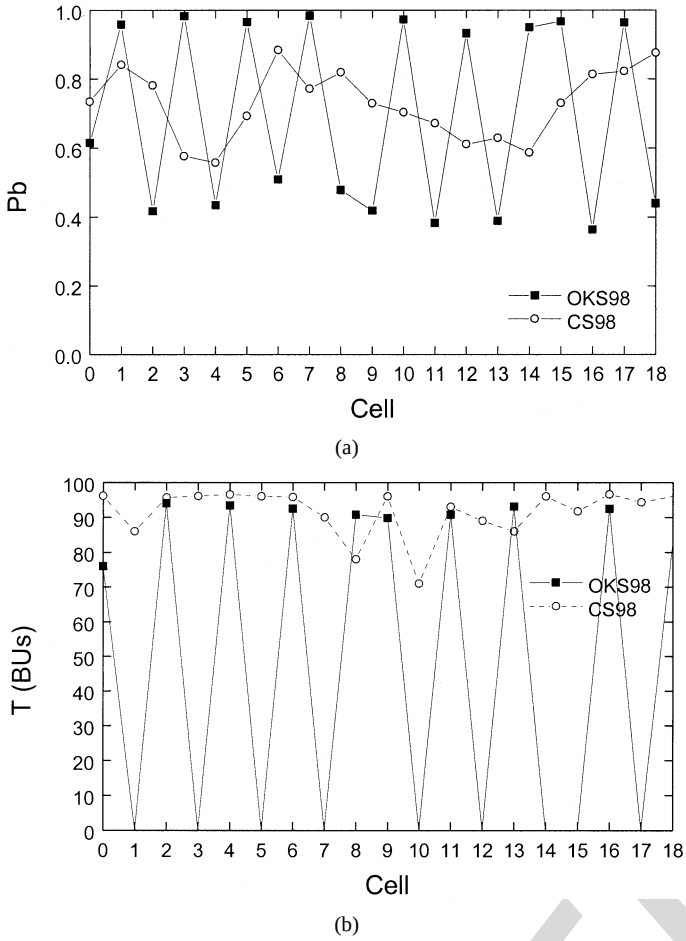


Fig. 13. Status of each cell at the end of simulations with offered load $= 3.0$, high mobility, and $F_1 = 1.0$.

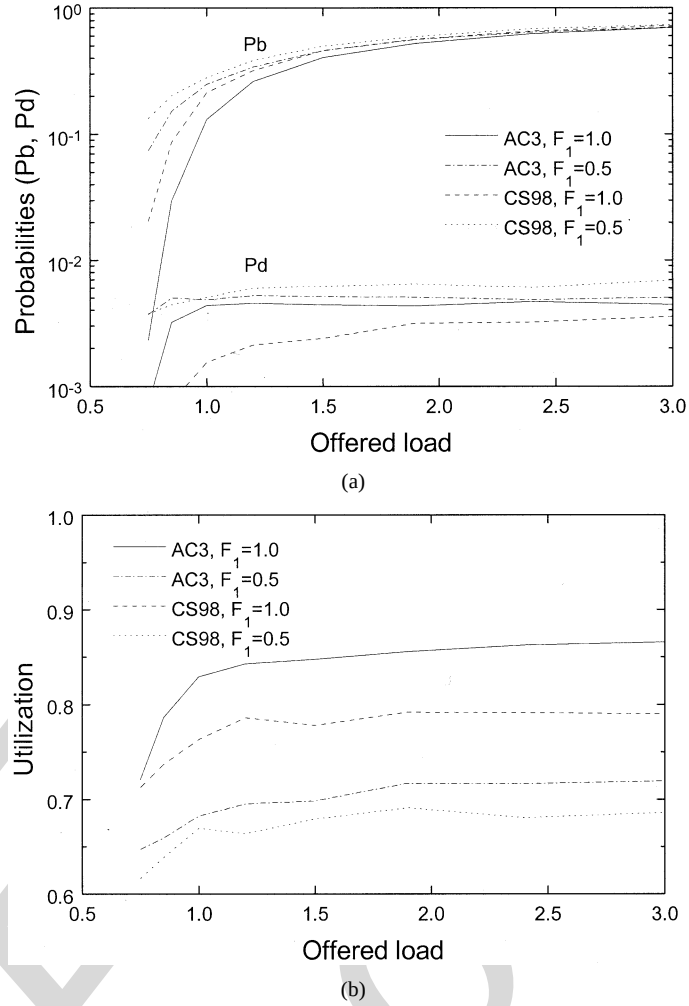


Fig. 14. Comparison of AC3 and CS98 for varying the offered load— high mobility and 4 BU of video bandwidth.

Bursty incoming handoffs will decrease T of a cell, say cell 1, which results in the decreased handoffs into cell 2 as

$$\frac{\partial \lambda_{h2}}{\partial T_1} = \gamma \frac{\partial M_1}{\partial T_1} = \gamma \frac{\partial M}{\partial T} \geq 0. \tag{14}$$

Most adaptation algorithms decrease T to maintain P_d below P_{QoS} for increased incoming handoffs. That is to say

$$\frac{\partial T_2}{\partial \lambda_{h2}} \leq 0 \tag{15}$$

holds. From (14) and (15) we obtain

$$\frac{\partial T_2}{\partial T_1} \leq 0 \tag{16}$$

which manifests the intercell unfairness issue. This means that two neighboring cells show the tendency of taking the opposite way in increasing or decreasing T values. Then a positive feedback loop can be closed by

$$\frac{\partial \lambda_{h1}}{\partial T_2} \geq 0. \tag{17}$$

IV. Enhanced Admission Control

We consider two complementary extensions to solve the unfairness problem. One is to modify the admission test and the other to modify the adaptive-control algorithm. Both take into account the current and adjacent cells together.

A. Enhanced Admission Test

We modify the admission test T1 as follows[7] and name it **T2**.

1) Check if $C_a(i) + B_{new} \leq T(i)$.
2) For all $j \in A_i$, check if $C_a(j) \leq T(j)$.
3) If both conditions are true, the new call is admitted.

In this test, if any of the adjacent cells are overloaded, the current cell blocks a new call request, even if it is not overloaded. In other words, when a cell is overloaded, new call requests are blocked in all adjacent cells. By doing so, the continuous handoffs into the overloaded cell can be reduced. Here, we assume that user movement information is not available.[8] The admission-control scheme that uses T2 and A1 is named **AC2**.

B. Enhanced Adaptive-Control Algorithm

Another method to solve the unfairness problem is to modify the algorithm A1. As was explained in Section III, if a cell is overloaded and multiple handoff drops occur, it is not sufficient

[7]A similar method can be found in [8] and [14].

[8]If the next cell into which the newly requested call will handoff can be known *a priori*, only the next cell needs to be checked.

IEE Proc

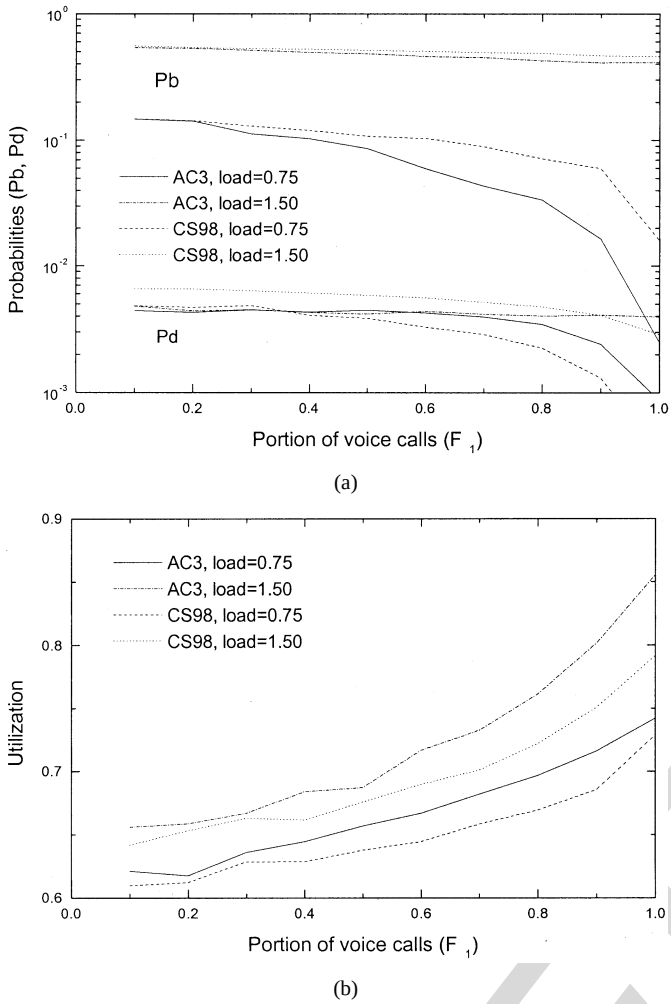


Fig. 15. Comparison of AC3 and CS98 for varying the portion of voice calls—high mobility and 4 BU of video bandwidth.

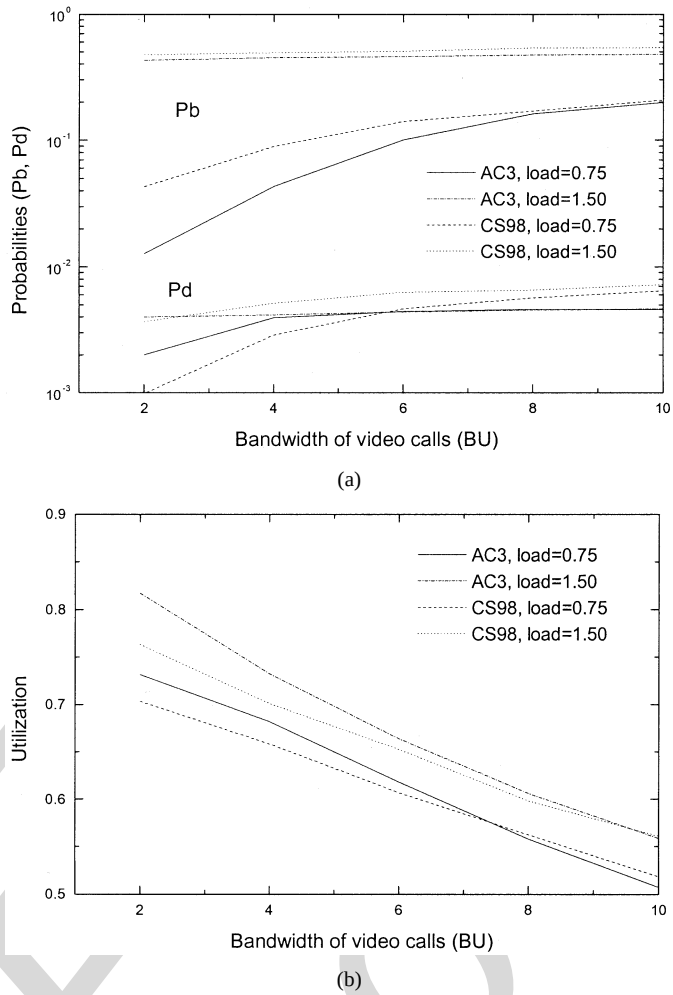


Fig. 16. Comparison of AC3 and CS98 for varying the bandwidth of the video call—high mobility and $F_1 = 0.7$.

to decrease only the threshold of the current cell—the thresholds of the adjacent cells must be decreased to reduce incoming handoffs. In order to compensate for too much threshold reduction and maximize the utilization, the thresholds of the adjacent cells should be properly increased when the threshold of the current cell is increased. So the basic idea is to decrease, or increase, the thresholds of adjacent cells along with that of the current cell. Fig. 4 shows the enhanced algorithm named **A2**.

The * indicates the newly inserted lines. When the BS of cell i decreases its threshold, it sends *decrease_T* messages to the BSs of A_i. When the BS of a cell j ($\in A_i$) receives this message, it decreases T if the normalized threshold[9] is higher than the average normalized threshold of adjacent cells. Thus, the thresholds of some adjacent cells that "appear" to have higher thresholds are decreased. Likewise, when the BS of cell i increases T, it sends *increase_T* messages to the BSs of A_i. When the BS of cell j receives this message, it increases T if the normalized threshold is lower than the average normalized threshold of adjacent cells and if its *QoS_state* is *IN*, which indicates that the long-term QoS is satisfied, i.e., P_d for the long-term is below the target value. Thus, the thresholds of some adjacent cells that "appear" to have lower thresholds are likely to be increased. These increases and decreases will result in the soft balancing of thresholds among neighboring cells, alleviating the intercell unfairness.

However, in some cases such as nonuniform loading conditions, it would be better for cells to have different thresholds. A2 consider these cases as well. Assume a cell i is heavily loaded and adjacent cells A_i are not. Then, cell i is more likely to be overloaded and have more handoff drops than A_i, causing the decrease of $T(i)$. In fact, this decreased $T(i)$ is close to $T_{\text{opt}}(i)$. The averaging effect, however, will not increase $T(i)$ above $T_{\text{opt}}(i)$, since $T(i)$ is increased only if the long-term QoS is satisfied. In addition, if it is above $T_{\text{opt}}(i)$, $P_d(i)$ will be higher than P_{QoS}, making the adaptive algorithm decrease $T(i)$ properly.

The above definition of *QoS_state* is a conservative policy. By defining it differently, an aggressive policy can be created. Let us define the *QoS_state* to be in *IN* when no handoff drop occurs during a short-term period, even if the long-term QoS is not satisfied.[10] We will compare the aggressive policy with the conservative policy in Section VI-B. Unless stated otherwise, the algorithm A2 obeys the conservative policy. We will refer to the admission-control scheme that uses T1 and A2 as **AC3**.

[9]The threshold normalized by the cell capacity, i.e., $\widetilde{T}(j) = T(j)/C(j)$.

[10]"$QoS_state = IN$" is included between lines 12 and 13 in the code.

V. Existing Admission-Control Schemes

In this section, we give a brief overview of existing practical admission-control schemes advanced in [14] and [18] for performance-comparison purposes. We will call the scheme in [14] **CS98** and that in [18] **OKS98**. Both methods use adaptive bandwidth reservation to keep P_d below a target value.

A. CS98 Method

CS98 is considered as a representative mobile-oriented scheme. Its authors compared their scheme with other existing mobile-oriented schemes and concluded that CS98 performs better, with reasonable complexity, in [20]. In CS98, a predictive and adaptive bandwidth-reservation scheme is employed. First, user mobility is estimated, based on an aggregate history of handoffs observed in each cell. This user mobility information is then used to (probabilistically) to predict mobiles' moving directions and handoff times by Bayes' estimation. Each cell calculates the bandwidth to reserve by estimating the total sum of fractional bandwidths of the expected handoffs within an estimation time window $T_{\rm est}$. $T_{\rm est}$ is adaptively controlled for the efficient use of bandwidth and effective response to time-varying traffic/mobility and inaccuracy of mobility estimation. When compared to our proposed schemes, CS98 has the following differences.

- CS98 uses a complex history-based method to calculate the target reserved bandwidth. Handoff events must be cached and the handoff probability of every call in adjacent cells must be calculated whenever a new call is tested for admission. Our schemes do not require such procedures.
- CS98 controls $T_{\rm est}$ adaptively, to satisfy the QoS constraint and maximize utilization by using an adaptive algorithm similar to A1. The target reserved bandwidth is an increasing function of $T_{\rm est}$. Thus, an indirect method to adjust the reserved bandwidth is used. Our methods, however, directly adjust the admission threshold to control the reservation level, based on handoff-dropping events.

The admission test in CS98 is almost the same as T2, which solves the unfairness problem. The difference is that, in CS98, only some of the adjacent cells that "appear" to be overloaded will participate in the admission test to reduce the complexity.

B. OKS98 Method

In this admission-control scheme, two classes of traffic are assumed: class I real-time traffic (such as voice and video) and class II nonreal-time traffic (such as data). The amount of bandwidth to reserve is determined by the largest of all the requested bandwidths from adjacent cells. As network conditions change after reservation, the reserved bandwidth needs to be dynamically adjusted. It is assumed that each BS continuously monitors the dropping probability P_d and the reserved bandwidth utilization U_R (i.e., the percentage of reserved bandwidth that is actually being used). The reserved bandwidth R is dynamically adjusted in each cell by the adaptive algorithm in Fig. 5.

OKS98 has some underspecified features and parameters. First, there is no specification of the monitored period of P_d. In this paper, we set this period to the long-term period, i.e., $P_d = L_{HD}/L_H$ when performing comparative evaluations. Second, it is not clear when to increase or decrease R, i.e., repeatedly or only once after the conditions are met. In this paper, we increase R at every handoff drop if the conditions are met. Likewise, R is decreased at the end of each short-term period if the conditions are met. The admission test is T1.

VI. Comparative Performance Evaluation

This section evaluates our three proposed admission-control schemes and compares our best one with competitive adaptive schemes, CS98 and OKS98 in particular. We first describe the simulation environment and parameter settings.

A. Simulation Environment and Parameters

We consider a two-dimensional cellular system. The topology of a wireless network is shown in Fig. 6. The cells are wrapped around to alleviate the *finite size effect*. The assumptions for our simulation study are as follows.

- The arrival process of new call requests is Poisson with rate λ (calls/s/cell), which is uniform to all cells unless stated otherwise.
- A new call is either for voice (1 BU) or video (4 BU), with the probability of F_1 and $1 - F_1$, respectively.
- The velocity of a mobile is randomly selected from $[V_{\min}, V_{\max}]$ (km/h) and the moving direction is also randomly selected. Once determined, its values are fixed until the call completes.
- The duration of a call is exponentially distributed with mean μ^{-1} (= 120 s).
- The capacity of each cell is C (= 100 BUs) and the cell's diameter is 1 km.

The other simulation parameters are $T_{\rm init} = T_{\max} = 100$ (BU's), $T_{\min} = 0$ (BU), and $P_{\rm QoS} = 0.01$, if not stated otherwise. The offered load per cell, L, is calculated as follows

$$L = (1\,F_1 + 4\,(1 - F_1))\frac{\lambda\,\mu^{-1}}{C}. \tag{18}$$

The numerator represents the average total bandwidth required to support all existing calls in a cell. The range of offered load was from 0.7 to 3.0. We consider two cases of user mobility, high mobility with range [80, 120], and low mobility with [40, 60].

B. Comparison of the Three Proposed Schemes

First, we simulated the three proposed admission-control schemes: AC1, AC2, and AC3. Fig. 7 plots (a) P_b and P_d and (b) utilization versus offered load for high mobility and $F_1 = 1.0$. Before we compare the three algorithms, let us focus on AC1. Although AC1 performs very well in $P_{\rm QoS} = 0.01$, it was observed to suffer from an intercell unfairness problem in $P_{\rm QoS} = 0.001$. This is due to the fact that the threshold is very hard to restore to its optimal value once it is decreased by bursty handoffs in $P_{\rm QoS} = 0.001$. Fig. 8 shows the status of each cell at the end of simulations with the offered load 3.0 and $P_{\rm QoS} = 0.001$. In AC1, P_b and T oscillate severely. In some cells, such as cells 0,2,7,..., the T values are extremely low and the P_b values are near 1.0. Thus, almost all new calls

IEEE Proof

are blocked and the P_d values are not kept below 0.001 in these cells. In the other cells, however, the P_d values were below 0.001 even at high T values. Once the intercell unfairness arises, the extremely decreased T does not return to its optimal value easily. In this case, the recovery is very difficult without traffic being greatly reduced, since the positive feedback loop is created between the cells of large and small T's, as explained in Section III. AC2 and AC3 resolve this unfairness problem effectively.

Now let us go back to Fig. 7 for the comparison of AC2 and AC3. In terms of P_d, both protocols meet the QoS constraint independent of the offered load. AC2 shows higher P_b and lower utilization than AC3.

Let us consider the detailed operations in a specific cell to further investigate the differences between the two schemes. Fig. 9 shows (a) T and (b) time average P_d, starting from the beginning of a simulation run for the offered load $= 3.0$, high mobility, and $F_1 = 1.0$ in cell 0. Other cells exhibit similar patterns. In Fig. 9(a), the values of T are observed to fluctuate as time progresses. In AC2, a decrease of T by one corresponds to a handoff drop, while an increase of T by one means that no handoff drop occurred during a short-term period. In AC3, however, the same correspondence does not apply since T can be decreased or increased by the corresponding signaling messages from adjacent cells.

We also observe that the decreasing instances of T exactly coincide with the increasing instances of P_d in Fig. 9(b), as expected. The P_d value near the starting point exceeds the target value 0.01 for both schemes, because the simulation starts with $T = T_{\text{init}} = 100$ (BUs).

However, as time goes on, P_d eventually goes below 0.01, affected by the adaptiveness of T. This shows how the proposed schemes can handle bursty handoff drops. Although our schemes cannot predict future handoff attempts, since it is purely based on handoff-drop events, it is almost impossible to predict bursty handoffs in real situations. In addition, if some reservation is made for future bursty handoffs, it would result in low channel utilization. Our schemes guarantee that the handoff drop probability for the long-term should be below a predefined level P_{QoS}, although the short-term handoff drop probability may not be below P_{QoS} due to bursty handoff attempts.

Another noteworthy point is that the T value in AC2 is higher than that of AC3 in Fig. 9(a). In AC2, a high T value in a cell does not necessarily mean that additional new calls will be admitted into that cell, because the T's of adjacent cells are also considered in T2. Fig. 10(a) shows the allocated bandwidth C_a obtained from the same simulation run as Fig. 9. We observe that C_a is low even at high values of T. The severe fluctuation between under- and over-utilization also reflects the characteristic of T2; when any of the six adjacent cells is overloaded, new calls will be blocked irrespective of the T value of the current cell. However, in AC3, only the current cell is considered when a new call is tested for admission. So the T value of the current cell directly affects the admission test. In Fig. 10(b), we can see that C_a in AC3 fluctuates less severely and the average pattern of C_a is similar to that of T in Fig. 9(a). As a whole, these differences make a noticeable difference in terms of average utilization in Fig. 7(b).

Now, we explain why the P_d value in AC3 is much lower than the target value in Fig. 7(a). In Section IV-B, we mentioned that A2, as used in AC3, affects a conservative policy. By taking the aggressive policy, P_d slightly increases within a range below the target value and utilization can be improved slightly as shown in Fig. 11. The reason that AC3 uses the conservative policy is to avoid the following situations. Suppose the central cell 0, its adjacent cells A_0, and the border cells have four-fold, two-fold, or one-fold the normal offered load, respectively (i.e., nonuniform loading condition). Fig. 12 shows the average P_d's of several regions as the normal offered load is increased. Although the total average P_d's are below 0.01 for both policies independent of the offered load, the P_d of cell 0 for the aggressive policy is above 0.01 when the offered load is near 0.6. At this load, only cell 0 is heavily loaded. Accordingly, the T value of cell 0 must be lower than any T values of A_0. In the aggressive policy, the BS of cell 0 will increase T when it receives the *increase_T* message from A_0 only if the short-term QoS is satisfied. This raises the P_d value in cell 0 above 0.01. In the conservative policy, however, the BS of cell 0 will not increase T in the same situation unless the long-term QoS is satisfied. This difference in the definition of *QoS_state* enables P_d to be kept below the target value even in a hot-spot cell. Thus, we choose the conservative policy[11] in AC3 as the best among our proposed schemes.

C. Comparison of AC3 With Existing Schemes

We now compare AC3 with existing adaptive admission-control schemes advanced in [14] and [18], denoted CS98 and OKS98, respectively. First, we simulated CS98 and OKS98 with the offered load 3.0, high mobility, and $F_1 = 1.0$, to check intercell fairness.[12] Fig. 13 shows the status of each cell at the end of simulation. OKS98 exhibits severe oscillations of P_b and T similar to AC1, because it adjusts the reserved bandwidth R $(= C - T)$ without considering the status of adjacent cells. CS98, however, solves this intercell unfairness problem because its admission test is similar to that of AC2. For this reason, we omit OKS98 henceforth and focus on the comparison of AC3 and CS98.

Figs. 14-16 compare the performance of AC3 and CS98 according to the offered load, the portion of voice calls, and the required bandwidth of the video call. Both schemes satisfy the QoS requirements, i.e., P_d's are kept below 0.01. In terms of P_b and utilization, both schemes show higher P_b and lower utilization for the smaller portion of voice calls and the larger bandwidth requirement of video calls. We can also see that the utilization in AC3 is higher than that in CS98 for a wide range of parameters.[13] However, in Fig. 16 the utilization of AC3 decreases faster than CS98 as the bandwidth of video calls increases. AC3 reserves the resource in unit of the averaged requested bandwidth, while CS98 does so in unit of the estimation

[11]In practice, the choice of policy is up to the service provider. The total average P_d in the aggressive policy is still below the target value, while at the same time achieving higher utilization than the conservative policy.

[12]The simulation parameters used in CS98 and OKS98 are the same as those in [14] and [18], respectively.

[13]We also simulated the low-mobility and time-varying traffic/mobility cases. They showed similar tendencies (i.e., both schemes guaranteed dropping probability and AC3 showed better utilization than CS98) and is eliminated due to space constraints.

IEEE Proof

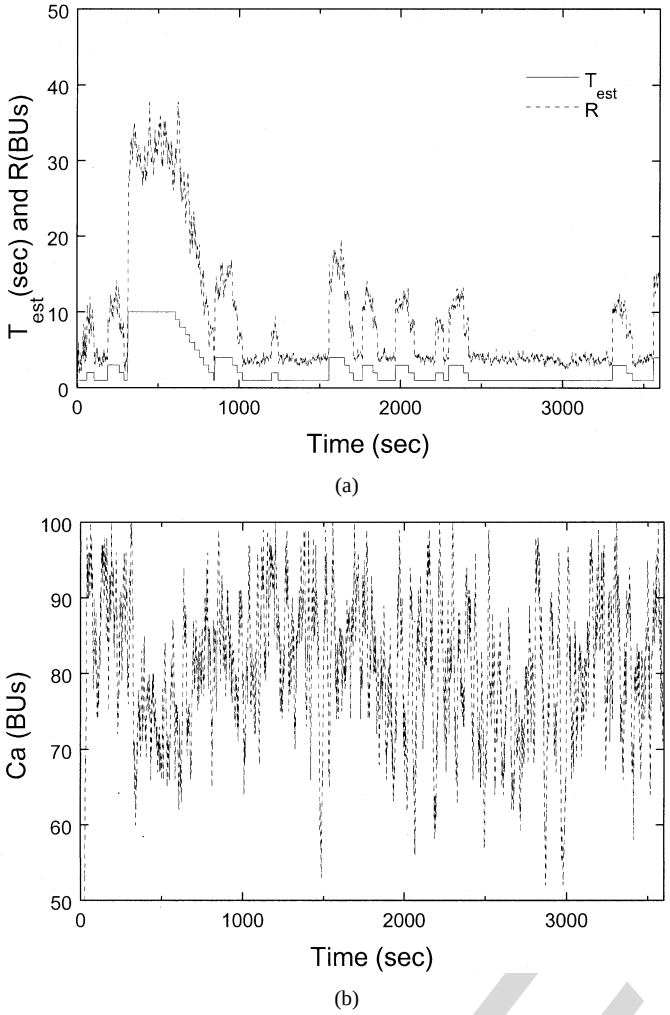


Fig. 17. $T_{\rm est}$, R and C_a versus time in cell 0 in CS98 with offered load $= 3.0$, high mobility, and $F_1 = 1.0$.

time. As the bandwidth of a requested video call increases, the reservation granularity becomes coarse in AC3, which degrades its performance.

To illustrate the notable differences between CS98 and AC3 in terms of utilization, we investigated the detailed operations of CS98 in a specific cell. Fig. 17 plots (a) time estimation window $T_{\rm est}$ and reserved bandwidth R and (b) C_a versus time in cell 0. It shows that the pattern of C_a is similar to that of AC2 in Fig. 10(a). This is because the admission test in CS98 is similar to T2. The other reasons come from the adaptive side of the algorithm. First, we observe that the increase or decrease of $T_{\rm est}$ is similar to the decrease or increase of the threshold T for AC2 in Fig. 9(a). The difference is the flat region of $T_{\rm est}$ $\big($e.g., during $t = [300{,}600]$ (sec)$\big)$, which corresponds to a long-term period. It results from the conservative $T_{\rm est}$ decrease policy, which corresponds to a conservative T increase policy in A1. We can also see that R is an increasing function of $T_{\rm est}$ and fluctuates even when $T_{\rm est}$ is constant. Hence, the coarse granularity associated with indirectly adjusting the reserved bandwidth is identified as another reason.

Finally, we compare the complexity of the two schemes. First, we compare the computational complexity for an admission decision. The complexity of CS98 with respect to an admission decision depends on $N_{\rm quad}$, which is the size of the cached history used for mobility estimation [20]. Fig. 18(a) shows the average numbers of numerical operations (i.e., summations and multiplications) and comparisons used by an admission decision. For CS98, the simplest case, $N_{\rm quad} = 1$ is used. While CS98 has a significant complexity overhead, AC3 requires only one operation and comparison in (1) for an admission decision.

Next, we compare the number of signaling messages among cells. In CS98, when the BS of cell i calculates the reserved bandwidth for an admission decision, it sends signaling messages to the BSs of adjacent cells A_i. The BS in cell j ($\in A_i$) then calculates the required bandwidth for the expected handoffs into cell i and informs this value back to cell i. So at least 12 messages[14] are required for an admission decision in a cell. On the other hand, in AC3, signaling messages are comprised of (1) *increase_T* messages, (2) *decrease_T* messages, and (3) $\widetilde{T}$ information messages. (3) is needed only when a BS increases or decreases T by receiving (1) or (2).[15] Fig. 18(b) shows the average number of messages sent at each cell per minute. The number of signaling messages in CS98 linearly increases according to the offered load, since accordingly more frequent admission tests are needed. However, in AC3, the number of signaling messages is "almost" independent of the offered load. As a whole, AC3 has a significantly smaller complexity overhead than CS98. In cellular wireless networks where both bandwidth and power consumption are at a premium, AC3 exerts an important advantage.

VII. Conclusion

In this paper, we proposed and evaluated realistic adaptive admission-control algorithms to keep the handoff-dropping probability below a predefined level while maximizing utilization. We investigated the intercell unfairness problem as a new performance evaluation criterion. We classified our protocols into three types, according to the type of admission test and the adaptive algorithm used to control the admission threshold. Through performance comparisons, we showed that AC3, which combines the simple admission test and the enhanced adaptive algorithm, is superior to the others in terms of fairness and performance. We also compared our AC3 scheme with other existing competitive bandwidth-reservation methods, in particular CS98 and OKS98. Our proposed scheme solved the intercell unfairness problem and showed high utilization under a variety of traffic loads, call bandwidths, and mobility conditions. In addition, it has extremely low complexity overhead and signaling load, making it readily implementable in real wireless networks.

[14]In the case of heavily loaded networks, more signaling messages are required to consider adjacent cells together in the admission test T2 [14].

[15]In most cases, a BS need not send (3) separately. By piggybacking (3) on (1) or (2), the BS can inform its $\widetilde{T}$ to the BSs of adjacent cells.

IEEE Proof

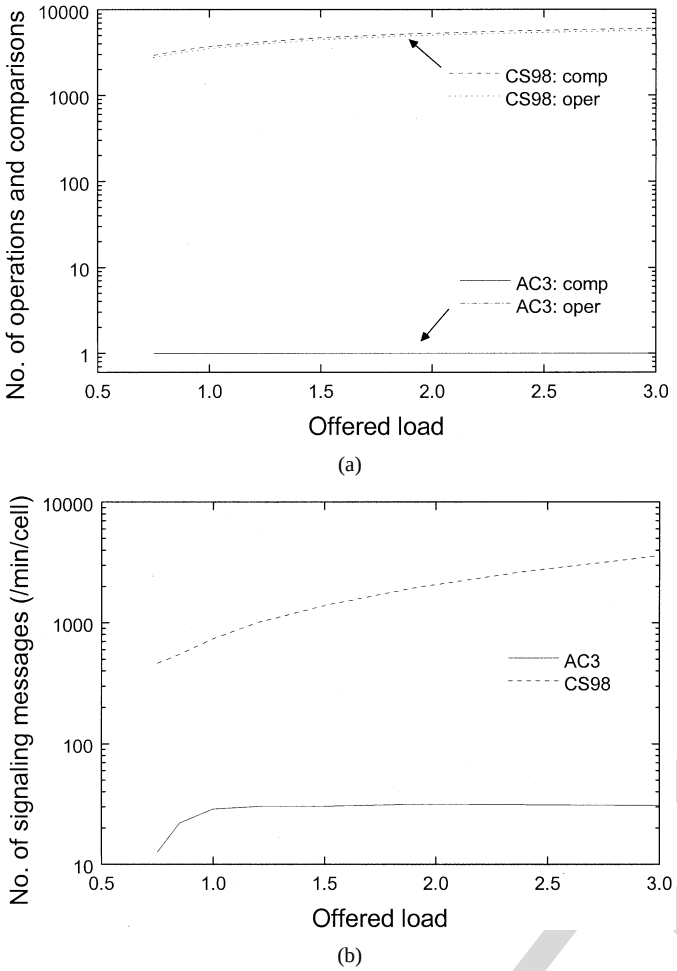


Fig. 18. Complexity comparison of AC3 and CS98.

REFERENCES

[1] M. Schwartz, "Network management and control issues in multimedia wireless networks," *IEEE Pers. Commun.*, vol. 33, pp. 8–16, June 1995.

[2] D. Hong and S. S. Rappaport, "Traffic model and performance analysis of cellular radio telephone systems with prioritized and nonprioritized handoff procedures," *IEEE Trans. Veh. Technol.*, vol. VT-35, Aug. 1986.

[3] S. Tekinay and B. Jabbari, "A measurement-based prioritization scheme for handovers in mobile cellular networks," *IEEE J. Select. Areas Commun.*, vol. 10, pp. 1343–1350, Oct. 1992.

[4] C. H. Yoon and C. K. Un, "Performance of personal portable radio telephone systems with and without guard channels," *IEEE J. Select. Areas Commun.*, vol. 11, pp. 911–917, Aug. 1993.

[5] K. Lee, "Supporting mobile multimedia in integrated service networks," *ACM Wireless Networks*, vol. 2, no. AU: PLEASE PROVIDE ISSUE NUMBER, pp. 205–217, 1996.

[6] P. Chong and C. Leung, "Capacity improvement in cellular systems with reuse partitioning," *J. Commun. Networks*, vol. 3, no. 3, pp. 280–287, Sept. 2001.

[7] R. Ramjee, R. Nagarajan, and D. Towsley, "On optimal call admission control in cellular networks," in *Proc. IEEE INFOCOM*, San Francisco, CA, Mar. 26–28, 1996, pp. 43–50.

[8] M. Naghshineh and M. Schwartz, "Distributed call admission control in mobile/wireless networks," *IEEE J. Select. Areas Commun.*, vol. 14, pp. 711–717, May 1996.

[9] S. Lu and V. Bharghavan, "Adaptive resource management algorithms for indoor mobile computing environments," in *Proc. ACM SIGCOMM AU: PLEASE SPELL OUT TITLE OF CONFERENCE.*, Stanford, CA, Aug. 28–30, 1996, pp. 231–242.

[10] A. Talukdar, B. Badrinath, and A. Acharya, "On accommodating mobile hosts in an integrated services packet network," in *Proc. IEEE INFOCOM*, Kobe, Japan, Apr. 9–11, 1997, pp. 1048–1055.

[11] O. Yu and V. Leung, "Adaptive resource allocation for prioritized call admission over an ATM-based wireless PCN," *IEEE J. Select. Areas Commun.*, vol. 15, no. 7, pp. 1942–1952, Sept. 1997.

[12] C. Chao and W. Chen, "Connection admission control for mobile multiple-class personal communication networks," *IEEE J. Select. Areas Commun.*, vol. 15, no. 8, pp. 1618–1626, Sept. 1997.

[13] D. Levine, I. 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.

[14] S. Choi and K. G. Shin, "Predictive and adaptive bandwidth reservation for handoffs in QoS-sensitive cellular networks," in *Proc. ACM SIGCOMM AU: PLEASE SPELL OUT TITLE OF CONFERENCE* , Vancouver, Canada, Sept. 2–4, 1998, pp. 155–166.

[15] A. Aljadhai and T. Znati, "A framework for call admission control and QoS support in wireless environments," in *Proc. IEEE INFOCOM*, Tel Aviv, Israel, Mar. 26–30, 1999, pp. 1019–1026.

[16] R. Jain and E. W. Knightly, "A framework for design and evaluation of admission control algorithms in multi-service mobile networks," in *Proc. INFOCOM*, Tel Aviv, Israel, Mar. 26–30, 1999, pp. 1027–1035.

[17] F. Yu and V. Leung, "Mobility-based predictive call admission control and bandwidth reservation in wireless cellular networks," in *Proc. IEEE INFOCOM*, Anchorage, AL, Apr. 22–26, 2001, pp. 518–526.

[18] C. Oliveira, J. Kim, and T. Suda, "An adaptive bandwidth reservation scheme for high-speed multimedia wireless networks," *IEEE J. Select. Areas Commun.*, vol. 16, pp. 858–874, Aug. 1998.

[19] P. Ramanathan, K. Sivalingam, P. Agrawal, and S. Kishore, "Dynamic resource allocation schemes during handoff for mobile multimedia wireless networks," *IEEE J. Select. Areas Commun.*, vol. 17, pp. 1270–1283, July 1999.

[20] S. Choi and K. G. Shin, "Comparison of connection admission-control schemes in the presence of handoffs in cellular networks," *Proc. ACM/IEEE Mobicom*, pp. 264–275, Oct. 1998.

[21] M. Zonoozi and P. Dassanayake, "User mobility modeling and characterization of mobility patterns," *IEEE J. Select. Areas Commun.*, vol. 15, pp. 1239–1252, Sept. 1997.

Jae Young Lee received the B.S. and M.S. degrees in electrical engineering from Seoul National University, Seoul, Korea, in 1997 and 2000, respectively. He is currently working toward the Ph.D. degree **AU: PLEASE PROVIDE AREA OF STUDY—Ed.** at the School of Electrical and Computer Engineering at Purdue University, West Lafayette, IN.

His areas of interests include optical communication and admission control at wireless networks.

Jin-Ghoo Choi received the B.S. and M.S. degrees in electrical engineering from Seoul National University, Seoul, Korea, in 1998 and 2000, respectively. He is currently working toward the Ph.D. degree **AU: PLEASE PROVIDE AREA OF STUDY—Ed.** at Seoul National University, Seoul, Korea.

His research interests include resource management and packet scheduling in wireless networks.

Kihong Park received the B.A. degree **AU: PLEASE PROVIDE AREA OF STUDY—Ed.** from Seoul National University, Seoul, Korea, and the Ph.D. degree in computer science from Boston University, Boston, MA.

He is an Associate Professor in the Department of Computer Sciences at Purdue University, West Lafayette, IN. His research interests include QoS provisioning, traffic modeling, network security, and fault-tolerance.

Dr. Park was a Presidential University Fellow at Boston University, is a recipient of the NSF CAREER Award, and is a Fellow-at-Large of the Santa Fe Institute. He serves on the editorial boards of *Computer Networks* and IEEE Communications Letters.

IEEE Proof

Saewoong Bahk (M'94) **AU: Please provide photos in tif, eps, or postscript formats with a resolution of at least 220 dpi—Ed.** received the B.S. and M.S. degrees in electrical engineering from Seoul National University, Seoul, Korea, in 1984 and 1986, respectively, and the Ph.D. degree **AU: PLEASE PROVIDE AREA OF STUDY** from the University of Pennsylvania, Philadelphia, in 1991.

He worked in the field of Network Management at AT&T Bell Laboratories as a Member of the Technical Staff from 1991 through 1994. Since 1994, he has been with the School of Electrical Engineering at Seoul National University and currently serves as an Associate Professor. He is an editor of *Journal of Communications and Networks*. His areas of interests include performance analysis of communication networks, network protocol design, and resource management at wireless and optical networks..

IEEE Proof