Practical Cell-oriented Adaptive Admission Control Guaranteeing QoS in Wireless Multimedia Networks

Jae Young Lee†, Jin-Ghoo Choi‡, Saewoong Bahk‡ and Kihong Park†
†Department of Computer Sciences, Purdue University
E-mail: {jylee, park} @purdue.edu
‡School of Electrical Engineering & Computer Science, Seoul, Korea
E-mail: {cjk, sbahk} @netlab.snu.ac.kr

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 pre-defined level. The admission threshold is dynamically adjusted based on handoff dropping events. We first present a simple admission control scheme which brings out an important performance evaluation criterion—inter-cell fairness—and serves as a reference point. We then investigate the inter-cell 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.

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—micro-cells or pico-cells—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 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 [2]–[8].

In [2], 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. In [3], the shadow cluster concept has been used to estimate future resource requirements and perform admission control in order to limit the handoff dropping probability ($P_d$). In this method, mobiles inform neighboring BSs of their bandwidth requirements and movement patterns at the call set-up time. Based on this information, BSs predict future demands and admit only those 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 [4], 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.

We emphasize that several previously proposed schemes are based on the user mobility information. 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 by dynamic control of reserved bandwidth. A practical cell-oriented scheme was introduced in [6]. 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 inter-cell unfairness problem which can significantly impede system utilization and performance.

In this paper, we consider inter-cell 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 inter-cell unfairness problem in the context of cell-oriented adaptive admission control which does not require user mobility information. The proposed scheme is able to provide probabilistic QoS guarantees while at the same time achieving high channel utilization. Since our protocol is simply based on handoff dropping events at each cell—and not based 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 inter-cell fairness issue and serves as a reference point. In Section III, we present the inter-cell unfairness problem and consider two enhanced protocols to overcome the unfairness problem. Section IV presents simulation results of our three proposed schemes, and compares our best one with existing protocols. We conclude with a discussion.
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.

A. Admission Control Test

A new call setup request is accepted into cell $i$ through the following admission test named $T_1$:

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

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 case of a handoff call, it is accepted if there is available bandwidth to accept the handoff call with bandwidth requirement $B_{handoff}$. That is,

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

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

There might exist an optimal steady-state admission threshold $T_{opt}(i)$ at cell $i$ for a specific traffic load and user mobility. Here we use the term “optimal” in the sense of maximizing (minimizing) utilization (new call blocking probability, $P_b$) 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 $A_1$, 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 $0 \leq T_{min} < T_{max} \leq C$.

The main idea in the adaptation is to monitor handoff dropping events over both the short-term and long-term. The objective of long-term monitoring is to keep $P_d$ below $P_{QoS}$, whereas the short-term monitoring is used to maximize utilization. The short-term period $T_S$ is given by the number of handoff attempts $1/P_{QoS}$. 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 $T_L$ is determined by handoff attempts as $S_H \times \text{max}(L_{HD}, 1)$ 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 $S_P$. After the system observes the first $L_P$ handoff attempts, it comes to one of the following two states.

1) State 1

When one or no handoff drop occurs for the first $L_P$ (= $S_P$) handoff attempts, the system enters State 1. In State 1, the dropping probability is

$$P_d = \frac{L_{HD}}{L_H} \leq \frac{1}{L_P} = \frac{1}{S_P} \approx P_{QoS}. \quad (3)$$

Hence, $P_d$ during this period is kept below $P_{QoS}$. If no handoff drop has occurred, it is likely that $T < T_{opt}$. So $T$ is increased by a predetermined step size $d$. In State 1, the long-term period ends with the short-term period, and the system state immediately goes to the initial state.

2) State 2

When more than one handoff drop occurs for the first $L_P$ handoff attempts, the system enters State 2. In this state, $L_P$ is increased by $S_P$ and $T$ is decreased by $d$ whenever a handoff drop occurs. State 2 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:

$$P_d = \frac{L_{HD}}{L_H} = \frac{L_{HD}}{L_P} \leq \frac{L_{HD}}{S_P L_{HD}} = \frac{1}{S_P} \approx P_{QoS}. \quad (4)$$

This is made possible by decreasing $T$ whenever a handoff drop occurs. By doing so, $T$ will approach $T_{opt}$ within some bounded time.

The admission control scheme which uses the admission test $T_1$ with the adaptive algorithm $A_1$ will henceforth be referred to as $AC1$.

### III. ENHANCED ADMISSION CONTROL

We introduce the inter-cell unfairness problem and consider two complementary extensions to solve it. One is to modify

---

1. $S_P = \lceil 1/P_{QoS} \rceil$; $L_P = S_P$;
2. $S_H = 0$; $S_{HD} = 0$; $L_H = 0$; $L_{HD} = 0$; $T = T_{inst}$;
3. WHILE (time increases)
4. IF (a mobile handoffs into the current cell) THEN
5. $S_H = S_H + 1$; $L_H = L_H + 1$;
6. IF (it is dropped) THEN
7. $S_{HD} = S_{HD} + 1$; $L_{HD} = L_{HD} + 1$;
8. IF ($L_{HD} > 1$) THEN
9. $L_P = L_P + S_P$;
10. $T = \text{max}(T - d, T_{min})$;
11. IF ($S_H = S_P$) THEN
12. IF ($S_{HD} < 1$) THEN
13. $T = \text{min}(T + d, T_{max})$;
14. $S_H = 0$; $S_{HD} = 0$;
15. IF ($L_H = L_P$) THEN
16. $L_H = 0$; $L_{HD} = 0$; $L_P = S_P$;

---

Fig. 1. Adaptive control algorithm (A1).
the admission test, and the other to modify the adaptive control algorithm. Both take into account the current cell and the adjacent cells together.

A. Inter-cell 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 inter-cell unfairness problem arises which can adversely impact performance [4]. Inter-cell unfairness is defined as the imbalance of the state of its adjacent base stations as in AC1, an inter-cell unfairness problem arises which can adversely impact performance [4]. Inter-cell unfairness is defined as the imbalance of admission thresholds among neighboring cells above and beyond those dictated by the optimal values \(T_{opt}(i)\). Specifically, when inter-cell 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 \(T\). This is “unfair” to cells with low \(T\)’s because almost all new calls are blocked in those cells. It is important to note that this is a universal phenomenon that is shared by all schemes that dynamically adjust the admission threshold (i.e., reserved bandwidth) [2], [4], [6], [7].

An example scenario for the unfairness problem is given as follows. Assume high mobility and uniform 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., \(C_d(i) > T(i)\)). Cell \(i\) may be still overloaded with incoming handoff calls instead of new calls. However, handoff calls have less chances to handoff than new calls since they have already passed some cells and have limited call durations. That is, handoff calls have more chances to terminate in the residing cell than the new calls that originate from that cell. Thus, outgoing handoffs from cell \(i\) decrease, contributing to less handoff drops in adjacent cells \(A_i\).

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 may still not keep \(P_d\) below a target value. On the contrary, 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.

B. Enhanced Admission Test

We modify the admission test T1 as follows and name it T2:

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

In this test, if any of the adjacent cells is 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. The admission control scheme which uses T2 and A1 is named AC2.

C. Enhanced Adaptive Control Algorithm

Another method to solve the unfairness problem is to modify the algorithm A1. As was explained in Section 3, if a cell is overloaded and multiple handoff drops occur, it is not sufficient 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. 2 shows the enhanced algorithm named A2.

```
3. WHILE (time increases)
   * IF(receive decrease_T message) THEN
     * IF(T > avg. T of adjacent cells) THEN
       * T = max(T - d, T_{min});
     * IF(receive increase_T message) THEN
       * IF(T < avg. T of adjacent cells and QoS_state==IN) THEN
         * T = min(T + d, T_{max});
     * IF(S_{H D} > 1) THEN
       * L_P = L_P + S_P;
     * IF(S_{H D} < 1) THEN
       * T = min(T + d, T_{max});
     * IF(L_{H} == L_{P}) THEN
       * L_H = 0; L_{H D} = 0;
     * IF(L_{H} == L_{P}) THEN
       * L_H = 0; L_{H D} = 0;
   * QoS_state=IN;

Fig. 2. Enhanced adaptive control algorithm (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 \(^2\) 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 the threshold normalized by the cell capacity, i.e., \(T(j) / C(j)\).

\(^2\)The threshold normalized by the cell capacity, i.e., \(T(j) / C(j)\).
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 increase and decrease will result in the soft balancing of thresholds among neighboring cells, alleviating the inter-cell unfairness.

However, in some cases such as non-uniform 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 $A_j$ are not. Then, cell $i$ is more likely to be overloaded and have more handoff drops than $A_j$, causing the decrease of $T(i)$. In fact, this decreased $T(i)$ is close to $T_{opt}(i)$. The averaging effect, however, will not increase $T(i)$ above $T_{opt}(i)$, since $T(i)$ is increased only if the long-term QoS is satisfied. In addition, if it is above $T_{opt}(i)$, $P_d(i)$ will be higher than $P_{QoS}$, making the adaptive algorithm decrease $T(i)$ properly. We will call the admission control scheme that uses $T_1$ and $A2$ as $AC3$.

IV. COMPARATIVE PERFORMANCE EVALUATION

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

A. Simulation Environment and Parameters

We consider a two-dimensional cellular system, which is the cluster of 19 cells. 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 $\lambda$ (calls/s/cell).
- 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 $\mu^{-1}$ (= 120 s).
- The capacity of each cell is $C$ (=100 BUs), and the cell’s diameter is 1 km.

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

$$L = (1 - F_1 + 4(1 - F_1)) \lambda \mu^{-1} / C.$$  

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. 3 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 $P_{QoS}$=0.01, it was observed to suffer from inter-cell unfairness problem in $P_{QoS}$=0.001. It is due to the fact that the threshold is much hard to restore its optimal value once it is decreased by bursty handoffs in $P_{QoS}$=0.001. Fig. 4 shows the status of each cell at the end of simulations with the offered load 3.0 and $P_{QoS}$=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 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, as explained in Section III-A. In AC2 and AC3, this unfairness problem is effectively resolved.

Now let us go back to Fig. 3 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. Thus we choose the 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 [4] and [6], 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 inter-cell fairness. Fig. 5 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 inter-cell 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.

Fig. 6 shows $P_b$, $P_d$ and utilization versus offered load with high user mobility for $F_1=1.0$ and 0.5. Both schemes satisfy the QoS constraint; $P_d$’s are kept below 0.01. In terms of $P_b$ and the utilization, both schemes show higher $P_b$ and lower utilization for $F_1=0.5$ than for $F_1=1.0$, since the more video calls exist, the more bandwidth is needed. We can also see that the utilization in AC3 is higher than that in CS98 for each value of $F_1$.

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 $N_{\text{guard}}$, which is the size of cached history used for mobility estimation. Fig. 7(a) shows the average numbers of numerical operations (i.e., summations and multiplications) and comparisons used by an admission decision. For CS98, $N_{\text{guard}}=1$ is used, the simplest case. 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 are required for an admission decision in a cell. On the other hand, in AC3, signaling messages are comprised of (i) increase $T$ messages, (ii) decrease $T$ messages and (iii) $T$ information messages. (iii) is needed only when a BS increases or decreases $T$ by receiving (i) or (ii). Fig. 7(b) shows the average number of messages sent at each cell per minute. 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.

V. CONCLUSION

In this paper, we proposed and evaluated practical adaptive admission control algorithms to keep the handoff dropping probability below a pre-defined level while maximizing utilization. We investigated the inter-cell unfairness problem as a new performance evaluation criterion. We compared our best scheme AC3 with other existing competitive bandwidth reservation methods, in particular, CS98 and OKS98. It solved the inter-cell 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.

REFERENCES