An Adaptive Bandwidth Reservation Scheme for High-Speed Multimedia Wireless Networks

Carlos Oliveira†, Jaime Bae Kim‡ and Tatsuya Suda†

† Dept. of Information and Computer Science
University of California, Irvine
Irvine, CA 92697
carlos@ics.uci.edu

‡ Computer Science Department
California State University, Northridge
Northridge, CA 91330

Abstract

In the next generation, high-speed wireless networks, it is important to provide quality-of-service (QOS) guarantees as they are expected to support multimedia applications. This paper proposes an admission control scheme based on adaptive bandwidth reservation to provide QOS guarantees for multimedia traffic carried in high-speed wireless cellular networks. The proposed scheme allocates bandwidth to a connection in the cell where the connection request originates and reserves bandwidth in all neighboring cells. When a user moves to a new cell and a hand-off occurs, bandwidth is allocated in the new cell; bandwidth is reserved in the new cell’s neighboring cells; and reserved bandwidth in more distant cells is released. The amount of bandwidth to reserve is dynamically adjusted, reflecting the current network conditions.

The performance of the proposed scheme is evaluated through simulations of realistic cellular environments. The simulated network consists of a large number of cells; mobile users with various movement patterns are assumed; and a variety of multimedia applications (e.g., audio phone, video conference, video on demand, file transfer, etc.) are considered. It is shown that the proposed scheme provides small hand-off dropping probability (i.e., the probability that hand-off connections are dropped due to lack of bandwidth) and achieves high bandwidth utilization.

1 Introduction

Next generation, high-speed wireless networks are expected to support multimedia applications (video, voice and data). As such, it is important that these networks provide quality-of-service (QOS) guarantees. QOS provisioning for multimedia traffic has been extensively studied for wireline networks, such as B-ISDN networks [1], [2], [3], [4], [5], [6], [7], [8], [9], [10], [11], [12]. However, provision of QOS in...
wireless networks becomes complex due to user mobility. The problem becomes even more challenging as recent wireless networks have been implemented based on small-size cells (i.e., micro-cells or pico-cells [13], [14], [15], [16]) to allow higher transmission capacity and thus to achieve better performance. Small-size cells increase the hand-off rate and result in rapid changes in the network traffic conditions, making QoS guarantees difficult [16], [17], [18].

In this paper, a new admission control scheme is proposed to provide high degrees of QoS guarantees for multimedia traffic carried in micro-cellular, high-speed wireless networks. The proposed scheme combines admission control and bandwidth reservation to guarantee QoS requirements. The proposed scheme considers both local information (e.g., the amount of unused bandwidth in the cell where the user currently resides) and remote information (e.g., the amount of unused bandwidth in the neighboring cells) to determine whether to accept or reject a connection. Since a mobile user is free to move anywhere, an admission control scheme that relies solely on local information cannot guarantee QoS requirements of a connection throughout its lifetime. The proposed scheme thus uses both local and remote information and allocates bandwidth in the cell where a connection request originates and reserves bandwidth in all neighboring cells. When a user moves to a new cell necessitating a call hand-off, the reserved bandwidth in the cell that the user is moving into is used to support the hand-off connection. In addition, every time a user moves to a new cell, bandwidth is reserved in the new neighboring cells, and the reserved bandwidth in the cells which are no longer neighboring to the new cell is released. Further, the proposed scheme distinguishes real-time traffic and non-real-time traffic and reduces the bandwidth assigned to non-real-time connections to provide higher quality of service to real-time connections if necessary. The proposed scheme can also adjust the amount of reserved bandwidth based on the current network conditions. This is done by measuring the average connection dropping probability (i.e., the probability that hand-off connections are dropped due to lack of bandwidth) and the reserved bandwidth usage (i.e., how much of the reserved bandwidth is actually being used) and adjusting the amount of reserved bandwidth accordingly. Therefore, the proposed scheme adapts to various network load conditions. It is also noteworthy that the proposed scheme is performed in each base station in a distributed manner and that no central coordination is necessary.

This paper is organized as follows. Section 2 surveys related work and shows that the proposed scheme is new and original. Section 3 describes the proposed scheme in detail. In Section 4, simulation models are presented. In Section 5, simulation results are presented to investigate the performance of the proposed scheme. Finally, concluding remarks are given in Section 6.

2 Related Work

As described in Section 1, the proposed scheme has the following features:
1. It combines admission control and bandwidth reservation to provide QoS.
2. It uses both local information and remote information to determine whether to accept or reject a new or hand-off connection.
3. It is adaptive. The proposed scheme adjusts the amount of reserved bandwidth based on the current network conditions.
4. It distinguishes new calls and hand-off calls. It gives higher priority to hand-off calls to provide higher connectivity to users who are already accepted in the network.
5. It reduces the bandwidth assigned to non-real time calls (i.e., low priority calls) to provide higher quality of service to real time calls (i.e., high priority calls), if necessary.
6. It is distributed. The proposed scheme is performed at each base station in a distributed manner.

As shown below, the proposed scheme is new and original. Existing schemes possess none of or only a small subset of key features of the proposed scheme.

Channel assignment and bandwidth reservation to support hand-off have been previously studied in [19], [20], [21], [22], [23], [24], [25], [26], [27].

In [19], when an hand-off occurs to a cell where there is not enough bandwidth, a channel currently being used is divided into two (equal size) sub-channels; one to serve the existing calls that are currently using the channel, and the other to serve the hand-off call. Unlike the proposed scheme where a variable amount of bandwidth is allocated or reserved, this scheme uses fixed size subchannels, resulting in potential fragmentation and waste of bandwidth when a call rate does not match the bandwidth of the fixed size channels. Further, this scheme does not apply admission control to guarantee QoS, nor uses information regarding neighboring cells. This scheme only assumes a single type of traffic (i.e., voice calls).

Papers [20], [21], [22], [23], [24] discuss schemes based on dynamic channel (re)assignment. In these dynamic channel (re)assignment schemes, channels are (re)assigned to different neighboring cells to reduce interference and to increase overall system capacity. For example, channels assigned to calls in progress may be reallocated to avoid neighboring cells using the same channels simultaneously. Some of the bandwidth may be removed from a call in progress and reassigned to a new call or a hand-off call. Existing dynamic channel (re)assignment schemes, however, require a centralized control for optimal channel assignment and thus result in significant control traffic overhead. In addition, unlike the scheme proposed in this paper, they consider little or no information regarding the neighboring cells. Further, existing dynamic channel (re)assignment schemes only address channel reassignment; they do not address admission control nor reserve bandwidth for hand-off calls. Therefore, desired call dropping probabilities for hand-off calls may not be guaranteed. They also assume a single traffic type (voice traffic).
Papers [25], [26], [27] consider bandwidth reservation for hand-off calls to guarantee high connectivity of admitted calls. In two similar bandwidth reservation schemes proposed in [25] and [26], a fixed number of channels in each cell is reserved exclusively for hand-offs. This scheme also allows queueing of hand-off requests when none of the reserved channels are available. Allowing queueing of hand-off requests is a worthwhile idea and complements the scheme proposed in this paper; however a trade-off between an increased delay (by queueing hand-off requests) and smaller dropping probabilities (due to a higher likelihood that reserved channels may become available for hand-off calls) needs to be considered. Unlike the proposed scheme, schemes described in [25], [26] do not apply admission control, nor consider information regarding neighboring cells. In addition, these schemes reserve a fixed number of channels and do not adapt to changes in the network conditions. Only voice traffic is assumed in these schemes.

Note that none of the above papers applies admission control to guarantee QoS, unlike the scheme proposed in this paper. Papers [28], [29], [30], [31] address the application of admission control to QoS guarantees; however, they do not apply bandwidth reservation. In the following, detailed comparisons of the proposed scheme with existing schemes are provided.

In [28] admission control for integrated voice and data traffic in packet radio networks is investigated. The admission control policy is based on a pre-determined threshold value of either the mean delay or the packet loss probability for data traffic and on the long-term blocking probability for voice traffic. This scheme does not reserve bandwidth in neighboring cells and does not adapt to changes in network conditions.

In the admission control scheme proposed in [29], different resource sharing schemes are employed to allocate resources to different classes of traffic. Although this scheme is interesting and useful, it does not consider reserving bandwidth in advance. It only assumes a simple, Poisson traffic model.

A distributed admission control scheme is proposed in [30]. In this scheme, the admission control is based on both the number of existing connections in the cell where a connection request is generated and the number of connections in the adjacent cells. Admission control is performed at each base station in a distributed manner. Only a single traffic type is considered. Unlike the proposed scheme, this scheme does not reserve bandwidth and is not adaptive to changes in network conditions. Therefore, the scheme proposed in [30] may not work well if the network carries diverse types of traffic with varying bandwidth requirements since the number of connections alone may not provide accurate information regarding the network traffic load.

In [31], an admission control based on dynamic channel assignment is proposed. In this scheme, network traffic conditions are first evaluated in the admission control phase, and then channels are assigned accordingly to new calls. Although this scheme is adaptive to changes in network conditions, it does not reserve bandwidth in neighboring cells. Therefore, this scheme risks an inability of meeting QoS requirements when hand-off occurs.
Other notable schemes exist to address efficient support of hand-off calls are described below.

In [32], [33], [34], a virtual tree architecture is proposed to support high-rate hand-offs between cells. A virtual tree represents connections of mobile users who remain within a greater geographical area covering some number of cells. In this architecture, once admitted, mobile users can freely hand-off to any cell within the virtual tree without being subject to a new admission control. It has been shown that this virtual tree architecture achieves small cell overload probability. This scheme assumes single traffic type and considers only the total number of connections to decide whether to accept or reject a connection.

In [35], the shadow cluster concept is used in admission control to allow predictive resource allocation. In this scheme, mobile terminals inform the base stations in neighboring cells of their bandwidth requirements, position and movement parameters at call set-up time. Based on this set of information, base stations predict future demands, reserve resources accordingly, and admit only those mobile terminals which can be supported adequately. The drawback of this scheme is that it requires detailed knowledge regarding the mobile terminal movement patterns such as with what probability a mobile terminal will be active in a given cell at a given time. The accuracy of prediction depends on a number of factors such as behavioral patterns of mobile terminals and characteristics of mobile terminals’ physical surroundings. However, it is difficult to accurately estimate movement patterns. An idea of using user mobility patterns is employed in the proposed scheme, and its effectiveness has been demonstrated in the simulation study. Paper [35] complements our paper in that it presents how movement patterns are estimated.

3 Proposed Admission Control Scheme Based on Adaptive Bandwidth Reservation

In this paper, two types of traffic are assumed in wireless networks: (a) Class I – real-time traffic and (b) Class II – non real-time traffic. Class I traffic includes video and voice traffic from users equipped with an adjustable rate codec [36], [37], [38], [39]. In case of congestion, such users can gracefully adjust the coding rate such that the quality of video/audio received at destination(s) is still acceptable. However, if the coding rate is reduced below some threshold, the quality of received video/audio becomes unacceptable. Class II traffic includes non-real-time data traffic such as email and other TCP/IP traffic. In case of congestion, it is acceptable to buffer non-real-time data at a network node (such as a base station) or at a user station and transmit them at a slower rate. For Class II traffic, it is assumed that there is no minimum required bandwidth since it can tolerate relatively large delays.

In the proposed scheme, it is assumed that when a mobile user requests a new connection or roams into a new cell, it provides the following information: (a) class of traffic (Class I or Class II), (b) desired amount of bandwidth of the connection, (c) minimum required bandwidth of the connection (only for the...
Class I connection), and (d) maximum acceptable connection dropping probability. The minimum required bandwidth for a Class I connection is the minimum amount of bandwidth that a source requires to maintain acceptable quality (e.g., such as the minimum coding rate at the source). The maximum acceptable connection dropping probability is the maximum acceptable probability that a connection is terminated during hand-off due to the lack of bandwidth.

For a new connection, the proposed scheme works as follows. The scheme attempts to allocate the desired amount of bandwidth in the cell where the new connection is generated. If the desired amount of bandwidth is not available, the new connection is rejected. If the desired amount of bandwidth is available, and the new connection is of Class II, then it is accepted and the desired amount of bandwidth is allocated. If the new connection is of Class I, the proposed scheme allocates the desired amount of bandwidth in the cell where the connection originates and reserves bandwidth in all neighboring cells in anticipation of hand-offs. (Algorithms to determine the amount of bandwidth to reserve are described later in this section.)

For a Class I hand-off connection, the proposed scheme works as follows. If the amount of available bandwidth in the cell that the hand-off connection is moving into is less than the minimum required bandwidth of the connection, the hand-off connection is dropped. Otherwise (i.e., if the amount of available bandwidth is greater than or equal to the minimum required bandwidth of the connection), the lesser of the available bandwidth in the cell and the connection’s desired amount of bandwidth is allocated to the hand-off connection. At the same time, bandwidth is reserved in the new neighboring cells, and the reserved bandwidth in old neighbor cells is released. (Algorithms to determine the amount of bandwidth to reserve are described later in this section.) If bandwidth reservation succeeds in all new neighboring cells, the hand-off connection is accepted. If reservation fails in any of the new neighboring cells, the hand-off connection is dropped. As mentioned earlier, Class I connections such as those from users with an adjustable rate video/voice codec can adjust the coding rate based on the bandwidth availability and, if necessary, reduce the coding rate down to its minimum level, still maintaining an acceptable quality at the destination.

For a Class II hand-off connection, it is accepted as long as there is some bandwidth available in the cell that the connection is moving into. A Class II hand-off connection is dropped only when there is no bandwidth available in the cell that the connection is moving into. Bandwidth reduction in a Class II hand-off connection results in a slower transmission rate and, thus, a longer transmission delay. However, Class II is for non-real-time traffic, and the impact of increased delay should not be significant.

Note that in the proposed scheme, when the available bandwidth in the cell is less than the desired amount of bandwidth, a new connection is always rejected, whereas a hand-off connection may be accepted if the minimum required bandwidth can be provided. In other words, the proposed scheme provides priority to a user that is already in the network (i.e., a hand-off connection) over a new user (i.e., a new connection).

As described above, bandwidth is reserved at connection setup and during hand-off for Class I connec-
Figure 1: Example of Bandwidth Reservation Procedure

tions. Each time a mobile user moves to a new cell, bandwidth is reserved in the new neighboring cells, and the reserved bandwidth in the cells which no longer neighbor the new cell is released. Refer to Figure 1 for an example. Assume that a mobile user initiates a new Class I connection in cell A. In the proposed scheme, bandwidth is allocated to the connection in the current cell A, and bandwidth is reserved in neighboring cells B, 1, 2, 6, 7 and 8. When the user moves from cell A to cell B, the reserved bandwidth (and additional unused bandwidth, if there is any) in cell B is used to accommodate the hand-off connection, and the reserved bandwidth in cells 1, 7 and 8 (i.e., cells which are not adjacent to cell B) is released. At the same time, bandwidth in cells A, 3, 4 and 5 (i.e., new neighboring cells) is reserved.

Various algorithms may be used to determine the amount of bandwidth to reserve for Class I connections. A simple approach is to reserve the desired amount of bandwidth of a Class I connection. However, this approach may result in reserving a huge amount of bandwidth, since the sum of the desired amount of bandwidths of a large number of Class I connections may need to be reserved. As an example, refer to Figure 1 again. Following this approach, it is necessary to reserve in cell B the sum of the desired amount of bandwidth of Class I connections in cells A, 2, 3, 4, 5 and 6. Since it is very unlikely that all Class I connections in cells A, 2, 3, 4, 5 and 6 will move to cell B at the same time, this approach results in inefficient bandwidth usage. A more efficient approach is to reserve only a fraction of the sum, allowing Class I connections to share a common pool of reserved bandwidth. The amount of bandwidth to reserve may be calculated as a function of the requested bandwidth by Class I connections (e.g., the average or the largest of all the requested bandwidths from the neighboring cells) or as a function of the number of Class I connections (e.g., the number of Class I connections in the neighboring cells). Both methods are
considered in this paper, and their performance is compared through simulation. The actual algorithms used in simulation to determine the amount of bandwidth to reserve are described in Section 4.1.

The proposed scheme is summarized in the following pseudo code:

```plaintext
IF new connection THEN
  IF desired amount of bandwidth ≤ unused bandwidth THEN
    IF Class I THEN
      accept connection
      allocate desired amount of bandwidth
      reserve bandwidth in neighboring cells
      /* if reservation is not successful in all */
      /* neighboring cells, the new connection is rejected */
    ELSE /* Class II */
      allocate desired amount of bandwidth
    END IF /* not enough bandwidth */
  ELSE /* hand-off */
    IF Class I THEN
      IF minimum required bandwidth ≤ Class I reserved bandwidth + unused bandwidth THEN
        accept connection
        allocate min{ desired amount of bandwidth, Class I reserved bandwidth + unused bandwidth }
        reserve bandwidth in new neighboring cells
        /* if reservation is not successful, drop the connection */
        release no longer needed bandwidth
      ELSE /* not possible to allocate minimum */
        drop connection
      END IF /* hand-off */
    ELSE /* Class II */
      IF unused bandwidth > 0 THEN /* as long as there is some available bandwidth */
        accept connection
        allocate min{ desired amount of bandwidth, unused bandwidth }
      ELSE /* no bandwidth available */
        drop connection
      END IF /* Class II */
  END IF /* hand-off */
END IF /* new connection */
```

After bandwidth is reserved (for Class I connections), network conditions may change. Therefore, the proposed scheme adjusts the size of the reserved bandwidth pool based on the current network conditions. It is assumed that each base station constantly monitors the hand-off dropping probability and the percentage of reserved bandwidth which is actually being used (reserved bandwidth pool utilization). In the proposed scheme, the hand-off dropping probability is the primary measure used to adjust the size of the reserved bandwidth pool. The reserved bandwidth pool utilization is used only when the hand-off dropping probability is not available (i.e., when there is no dropping of hand-off connections). The algorithm to control the size of the reserved bandwidth pool is summarized in the following pseudo code:

```plaintext
IF monitored dropping probability ≠ 0 THEN
  IF monitored dropping probability ≥ \( \text{thres}_{up1} \times \text{requested dropping probability} \) THEN
    size of reserved bandwidth pool = \( \text{min}\{ up_1 \times \text{size of reserved bandwidth pool, reserved bandwidth + unused bandwidth} \} \)
    /* increase the size of reserved bandwidth pool by \( up_1(>1) \) */
  ELSE
    IF monitored dropping probability ≤ \( \text{thres}_{down1} \times \text{requested dropping probability} \) THEN
      size of reserved bandwidth pool = \( \text{down}_1 \times \text{size of reserved bandwidth pool} \)
      /* decrease the size of reserved bandwidth pool by \( down_1 (0 < down_1 < 1) \) */
  END IF
END IF
```
As the pseudo code illustrates, the proposed scheme increases (decreases) the reserved bandwidth pool size by a factor $up_1$ ($down_1$) when the monitored hand-off dropping probability is larger (smaller) than a fraction $thres_{up1}$ ($thres_{down1}$) of the requested dropping probability, or by a factor $up_2$ ($down_2$) when the utilization of the reserved bandwidth pool is above (below) a threshold value $thres_{up2}$ ($thres_{down2}$).

The scheme proposed in this paper provides high degrees of QOS (i.e., connection dropping probability) guarantees for multimedia traffic. The proposed scheme allows trading-off the connection blocking probability of new connections and the connection dropping probability of hand-off connections. The connection blocking probability (CBP) is the probability that a new connection is rejected, whereas the connection dropping probability (CDP) is the probability that a hand-off connection is dropped due to insufficient bandwidth. Reserving bandwidth for hand-off connections results in less available bandwidth for newly arriving connections. This, in turn, results in the higher blocking probability for new connections. This trade-off will be examined in Section 5 (Numerical Result section).

The proposed scheme dynamically adapts the amount of the reserved bandwidth pool based on the current network conditions (i.e., based on the average connection dropping probability and the reserved bandwidth usage) to improve the bandwidth utilization and the connection blocking probability. None of the previously proposed schemes has this adaptive capability. While there is some processing overhead associated with measuring the average connection dropping probabilities and the reserved bandwidth usage, the amount should be minimal.

4 Simulation Model

4.1 Three Models of the Proposed Scheme

There are some key design options for the proposed scheme. First, in reserving bandwidth for Class I connections, if user movement patterns are known, more bandwidth may be reserved in the cells that users are likely to move to. If such patterns are not known, the same amount of bandwidth may be reserved in each of the neighboring cells. Second, in determining the amount of bandwidth to reserve, various functions can be used. For instance, the amount of bandwidth to reserve may be calculated based either on
the requested bandwidth of existing Class I connections or on the number of existing Class I connections. Through simulations, the trade-offs of the following three models of the proposed scheme are investigated.

**Uniform and Bandwidth-Based Reservation Model**

In this model, it is assumed that user movement patterns are unknown, and the same amount of bandwidth is reserved in all neighboring cells. In addition, the amount of bandwidth to reserve is calculated based on the requested bandwidth of existing Class I connections. Specifically, the largest of all the requested bandwidths, \( \text{the largest} \), is used as the amount of bandwidth to reserve.

**Movement-Based and Bandwidth-Based Reservation Model**

In this model, it is assumed that user movement patterns are known, and different amounts of bandwidth are reserved in different neighboring cells based on user movement patterns. (A particular user’s movement patterns can be provided either by the user or by the network through techniques that predict a user’s movement [40], [41], [42].) In the simulations, it is assumed that the user movements are highly directional. That is, a user is likely to move to one of the neighboring cells, say cell \( D \), with a large probability \( p_D \). The largest of all the requested bandwidths, \( \text{the largest} \), is reserved in cell \( D \) as is done in the first model, and a smaller amount of bandwidth \( \text{the largest} \times f \), where \( f \) is a reduction factor, is reserved in each the remaining neighboring cells.

**Movement-Based and Number-of-Connections-Based Reservation Model**

This model is similar to the second model. The only difference is that in this model, the amount of bandwidth to reserve is calculated as a function of the number of existing Class I connections instead of the requested bandwidths. In the simulations, the amount of bandwidth to reserve is determined using Table 1. The values in Table 1 were calculated by multiplying the average bandwidth of Class I connections assumed in the simulation (shown in Table 2) and the average number of Class I connections for each range (i.e., 0 to 5, 6 to 10, 11 to 20, 21 or more). Table 2 shows traffic characteristics of Class I and Class II connections assumed in the simulation. (This table will be explained later in section 4.3)

**4.2 Other Schemes Simulated for Comparisons**

In order to evaluate the performance of the proposed scheme, two generic bandwidth allocation schemes (Scheme A and Scheme B) that use only local information (i.e., information regarding the cell where the user currently resides) are used as a basis for comparison. These two schemes are described below.
Table 1: Reserved Bandwidth Based on The Number of Class I Connections

<table>
<thead>
<tr>
<th>Number of Class I Connections</th>
<th>Reserved Bandwidth</th>
</tr>
</thead>
<tbody>
<tr>
<td>0 - 5</td>
<td>512 kbps</td>
</tr>
<tr>
<td>6 - 10</td>
<td>1024 kbps</td>
</tr>
<tr>
<td>11 - 20</td>
<td>2048 kbps</td>
</tr>
<tr>
<td>21 or more</td>
<td>3072 kbps</td>
</tr>
</tbody>
</table>

Scheme A represents an extension of the current cellular system [13], [15] with support for multimedia traffic. Scheme A is a generic scheme where no control (e.g., no bandwidth reservation) is employed. This scheme thus provides high bandwidth utilization and small blocking probabilities for new connections. However, this is at the expense of high dropping probabilities for hand-off connections. Scheme B represents a class of schemes that use dynamic resource (re)assignment, an alternative approach to the bandwidth reservation used in the proposed scheme. Similar concept has been used in [20], [21], [22], [24], [23]. Both Scheme A and Scheme B are used to investigate the trade-off between the dropping probability of hand-off connections, the blocking probability of new connections, and bandwidth utilization in the proposed scheme and to provide a baseline for comparisons.

Scheme A - No Reservation Scheme

As with the proposed scheme, Scheme A distinguishes between new and hand-off connections. However, unlike the proposed scheme, Scheme A does not reserve bandwidth in neighboring cells. In Scheme A, a new connection (regardless of whether it is a Class I connection or a Class II connection) is accepted when its desired amount of bandwidth is available in the cell where the new connection is generated. Otherwise, it is rejected. A Class I hand-off connection is accepted when at least its minimum required bandwidth is available in the cell that the hand-off connection is moving into. Unlike the proposed scheme, new and hand-off Class I connections are accepted without reserving bandwidth in neighboring cells. A Class II hand-off connection is accepted as long as there is some unused bandwidth in the cell that the hand-off connection is moving into as in the scheme proposed in this paper.

Scheme A is summarized by the following pseudo code:

```plaintext
IF new connection THEN
    IF desired amount of bandwidth \leq unused bandwidth THEN
        accept connection
        allocate desired amount of bandwidth
    ELSE /* not enough bandwidth */
        reject connection
ELSE /* hand-off */
    IF Class I THEN
        IF minimum required bandwidth \leq unused bandwidth THEN
            accept connection
```

11
allocate min{ desired amount of bandwidth, unused bandwidth }
ELSE /* not possible to allocate minimum */
drop connection
ELSE /* Class II */
    IF unused bandwidth > 0 THEN /* as long as there is some available bandwidth */
        accept connection
        allocate min{ desired amount of bandwidth, unused bandwidth }
    ELSE /* no bandwidth available */
        drop connection

Note that the proposed scheme becomes equivalent to Scheme A when its bandwidth reservation process is removed.

**Scheme B - Bandwidth Reassignment Scheme**

Scheme B differs from Scheme A only in the way that a Class I hand-off is handled. In Scheme A, a Class I hand-off connection is dropped when its minimum required bandwidth is not available. In Scheme B, if the minimum required bandwidth is not available, bandwidth from already existing Class II connections is “borrowed” (reassigned) so that the minimum required bandwidth of the Class I hand-off connection can be allocated. It is assumed in the simulation that, in “borrowing” bandwidth from existing Class II connections, Scheme B proportionally borrows bandwidth from all existing Class II connections in the destination cell of a Class I hand-off.

**4.3 Simulation Model Description and Parameter Values**

The simulation model is composed of $N$ cells, each cell keeping contact with its six neighboring cells. Refer to Figure 2. Each cell contains a base station, which is responsible for the connection setup and tear-down of new connections and hand-off connections, as well as the reservation of bandwidth in neighboring cells. Each cell has a maximum bandwidth capacity of $B$ bps.

Two types of new connections are assumed in the simulation: a user connection, which is initiated by a mobile user, and a network connection, which is initiated by a user in the wireline network and destined to a mobile user. The interarrival times of new connection requests are assumed to follow a geometric distribution with mean $1/\lambda_N$ for network connections and mean $1/\lambda_U$ for user connections. It is also assumed that each connection may experience multiple hand-offs in its life time. The probability that a connection experiences its first hand-off is assumed to be $p_h$, and this probability is assumed to decrease exponentially for successive hand-off’s of the same connection. Namely, the hand-off probability of a connection is equal to $p_h/2^n$, where $n$ is the number of hand-offs already experienced by the connection.

In order to represent various multimedia applications, six different application groups are assumed based on the connection duration, bandwidth requirement and class of service (Class I or Class II). The different
application groups include constant bit rate (CBR), variable bit rate (VBR), and data traffic sources (unspecified bit rate - UBR). Table 2 shows the six application groups used in the simulation. These are typical applications seen on existing networks, and their parameter values are chosen from [2], [4], [43], [44], [45], [46], [47], [48]. The bandwidth required for VBR and UBR connections is assumed to follow a geometric distribution between the minimum and the maximum values shown in Table 2. The average bit rates of application groups 3, 4, 5 and 6 are assumed to be \( \frac{3}{98} \), \( \frac{4}{98} \), \( \frac{5}{98} \) and \( \frac{6}{98} \) bps, respectively (see Table 2). The connection duration is also assumed to follow a geometric distribution between the minimum and maximum values shown in Table 2, with average durations of \( \frac{1}{116} \), \( \frac{2}{116} \), \( \frac{3}{116} \), \( \frac{4}{116} \), \( \frac{5}{116} \) and \( \frac{6}{116} \) for application groups 1 through 6, respectively. In the simulation, it is assumed that new connections from all six application groups are generated with equal probability.

As described in Section 3, the proposed scheme uses several parameters to dynamically adjust the size of the reserved bandwidth pool based on current network conditions. First, two sets of threshold values, \((\text{thres}_{\text{up}1}, \text{thres}_{\text{down}1})\) and \((\text{thres}_{\text{up}2}, \text{thres}_{\text{down}2})\), should be chosen such that \(\text{thres}_{\text{up}1} > \text{thres}_{\text{down}1}\) and \(\text{thres}_{\text{up}2} > \text{thres}_{\text{down}2}\), creating a hysteresis effect. This effect generates a stable operation region for the scheme (i.e., between \(\text{thres}_{\text{up}}\) and \(\text{thres}_{\text{down}}\)). Second, larger values of \(\text{up}_1\) and \(\text{up}_2\) (increment factors), and smaller values of \(\text{down}_1\) and \(\text{down}_2\) (decrement factors) make the algorithm more aggressive, i.e., more responsive to changes in the network conditions. However, larger values for increment factors may result in a waste of bandwidth, whereas smaller values for decrement factors may result in a lack of bandwidth.
Two different user movement patterns are simulated: random movement pattern (i.e., the user moves to all possible directions with equal probability), and highly directional movement pattern (i.e., the user moves to a specific neighboring cell $D$ with very high probability $p_D$). Furthermore, to investigate how the proposed scheme adapts to changing network conditions, a special configuration is also simulated. This configuration includes a cell (called a source cell) where the arrival rate of hand-off connections is much smaller than the network average and another cell (called a sink cell) where the arrival rate of hand-off connections is much larger than the network average. In other words, it is very unlikely that mobile users in the surrounding cells move into the source cell. Therefore, the probability that a mobile user requests a hand-off connection in the source cell, $p_{source}$, is very small. On the other hand, it is very likely that mobile users in surrounding cells move into the sink cell. Therefore, the probability that a mobile user requests a hand-off connection in the sink cell, $p_{sink}$, is very large.

Table 3 summarizes the simulation parameters. The values for the simulation parameters are chosen carefully in order to closely represent realistic scenarios [2], [4], [43], [44], [45], [46], [47], and yet make the simulation feasible. The hand-off probability $p_{hand-off}$ represents the system mobility (i.e., hand-off rate), and thus, three values, 0.25, 0.5 and 0.75, are used to represent low mobility, medium mobility and high mobility in the system.

Using the simulation model described above, the proposed scheme is evaluated. The performance measures obtained through the simulation are the blocking probability of new connections, the dropping probability of hand-off connections, and the bandwidth utilization. In most of the simulation results that follow,
<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>$N$</td>
<td>100 cells</td>
<td>Number of cells in the system</td>
</tr>
<tr>
<td>$B$</td>
<td>30 Mbps</td>
<td>Maximum Bandwidth Capacity of a Cell</td>
</tr>
<tr>
<td>$1/\lambda_N$</td>
<td>variable</td>
<td>Mean interarrival time of network connection</td>
</tr>
<tr>
<td>$1/\lambda_U$</td>
<td>variable</td>
<td>Mean interarrival time of user connections</td>
</tr>
<tr>
<td>$p_{hi}$</td>
<td>0.25/0.5/0.75</td>
<td>Hand-off Probability</td>
</tr>
<tr>
<td>$p_{D}$</td>
<td>random ($p_i = p_j$ for all $i,j$) $0.8(\sum p_i = 0.2 \ (i \neq D)$)</td>
<td>Probability to move to a specific neighboring cell $D$</td>
</tr>
<tr>
<td>$p_{source}$</td>
<td>0.25</td>
<td>Probability to move to a source cell</td>
</tr>
<tr>
<td>$p_{sink}$</td>
<td>0.95</td>
<td>Probability to move to a sink cell</td>
</tr>
<tr>
<td>$thres_{up_1}$</td>
<td>1.0</td>
<td>upper threshold for dropping probability</td>
</tr>
<tr>
<td>$thres_{down_1}$</td>
<td>0.75</td>
<td>lower threshold for dropping probability</td>
</tr>
<tr>
<td>$thres_{up_2}$</td>
<td>0.75</td>
<td>upper threshold for reserved pool utilization</td>
</tr>
<tr>
<td>$thres_{down_2}$</td>
<td>0.25</td>
<td>lower threshold for reserved pool utilization</td>
</tr>
<tr>
<td>$up_1, up_2$</td>
<td>1.1</td>
<td>increment factor for reserved bandwidth pool size</td>
</tr>
<tr>
<td>$down_1, down_2$</td>
<td>0.9</td>
<td>decrement factor for reserved bandwidth pool size</td>
</tr>
</tbody>
</table>

Table 3: Simulation Parameters

these performance measures are plotted as a function of the connection arrival rate. The connection arrival rate is an arrival rate of new connections measured as the average number of new connection requests per second per cell. New connections include both network connections and user connections, and the arrival rate of new connections is $\lambda_N + \lambda_U$.

5 Numerical Results

In this section, simulation results are presented to evaluate the proposed scheme. In Section 5.1, a simple version of the proposed scheme (i.e., the proposed scheme with its feature of adapting to the network conditions disabled) is compared with Scheme A and Scheme B described in Section 4.2. In Section 5.2, three models of the proposed scheme discussed in Section 4.1 are simulated, and their trade-offs discussed in Section 4.1 are investigated. The adaptive feature of the proposed scheme is disabled in Section 5.2. The adaptive feature of the proposed scheme is investigated in Section 5.3 by comparing adaptive and non-adaptive versions of the proposed scheme. In Section 5.4, the performance of the adaptive version of the proposed scheme for Class II is investigated.

5.1 Comparison between the Proposed Scheme and Schemes A and B

In order to objectively and fairly compare the proposed scheme and Schemes A and B, the adaptive feature of the proposed scheme is disabled in this subsection. This is because Schemes A and B do not have this
capability. Since the proposed scheme simulated in this subsection does not adjust the amount of reserved bandwidth based on the network conditions, it provides the lower bound of the performance of the proposed scheme with adaptive feature. Further, it is assumed that information regarding the users’ movement patterns are unknown. Thus, only the Uniform and Bandwidth-Based Reservation model of the proposed scheme is used (see Section 4.1). The hand-off probability $p_{h}$ is assumed to be 0.75, representing a highly mobile environment.

Figure 3 shows the connection blocking probability (CBP) of new Class I connections and the connection dropping probability (CDP) of Class I hand-off connections. In this figure, we assume that the maximum amount of bandwidth that can be reserved per cell is a percentage $r$ of the total bandwidth of the cell, with $r = 20\%$. In other words, the maximum size of the reserved bandwidth pool is limited to 20\% of the total bandwidth of the cell. This figure shows the trade-off between the CBP and CDP in the proposed scheme. Since the proposed scheme reserves bandwidth in neighboring cells for hand-offs, it achieves smaller CDP than Schemes A and B. However, it comes at the expense of a larger CBP than Schemes A and B since less bandwidth is available for newly arriving connections. Although the proposed scheme produces a larger CBP than Schemes A and B, the percentage of decrease in the CDP is greater than the percentage of increase in the CBP. For example, at 0.1 connection arrival rate, the proposed scheme shows the CBP that is 45.5\% higher than Scheme A and 22.5\% higher than Scheme B. On the other hand, the proposed scheme achieves the CDP that is 89\% lower than Scheme A and 66\% lower than Scheme B.

It is also shown in this figure that for the proposed scheme and Scheme B, the CDP decreases as the connection arrival rate increases beyond a certain value of the arrival rate. In these schemes, as the connection arrival rate increases, the average amount of unused bandwidth in a cell decreases, and thus, new connections which require a large amount of bandwidth are likely to be rejected. Therefore, the average bandwidth allocated to new connections decreases, and when these connections are handed-off, they are likely to be accepted since their minimum required bandwidth is likely to be smaller than the desired bandwidth of new calls. This results in a smaller CDP. On the other hand, as the connection arrival rate increases, the number of hand-offs also increases increasing the CDP. Overall, this trade-off causes a decrease in the CDP beyond a certain value of the connection arrival rate.

To further investigate the trade-off between the CBP, CDP and bandwidth utilization and to make a fair comparison between different schemes, the CBP value is fixed for all three schemes in the following two figures (Figure 4 and Figure 5).

Figure 4 shows the CDP of Class I hand-off connections. The CBP values are fixed for all three schemes. This figure shows that the proposed scheme achieves the smallest CDP for the same CBP value.

Figure 5 shows bandwidth utilization as a function of the connection arrival rate. Bandwidth utilization is the percentage of the total bandwidth actually being used by connections in a cell. Again, the CBP values
Figure 3: Connection Blocking Probability and Connection Dropping Probability (Class I)

Figure 4: Connection Dropping Probability (Class I)
are fixed for all three schemes. It is shown that for low connection arrival rates, the bandwidth utilization is virtually the same for all schemes. However, as the connection arrival rate increases, the proposed scheme results in lower bandwidth utilization than Schemes A and B as expected. This is because the proposed scheme reserves bandwidth for hand-off connections, whereas Schemes A and B do not. Figure 5 also shows that the bandwidth utilizations for Schemes A and B are virtually identical. This is because both schemes do not reserve any bandwidth, and thus no bandwidth is wasted.

Simulation results for a low mobility environment ($p_H = 0.25$) and a medium mobility environment ($p_H = 0.5$) showed similar behaviors as seen in a high mobile environment. Due to the space limitation those results are not included in the paper.

In this subsection, the proposed scheme without the adaptive feature is compared with Scheme A and Scheme B. As mentioned in Section 4.2, Scheme A represents a class of schemes which do not reserve any bandwidth to avoid the dropping of hand-offs, whereas Scheme B represents a class of schemes which dynamically (re)assigns bandwidth, an alternate approach to the proposed scheme. The proposed scheme suffers from overhead for making reservations. The overhead includes, for instance, overhead due to exchange of signaling packets between a base station of the cell where a user currently resides and base stations of the neighboring cells in addition to processing overhead at base stations to process reservation requests. Scheme B suffers from overhead of reassigning bandwidth from existing calls to new or hand-off
calls. This overhead includes, for instance, removing bandwidth from a call in progress (e.g., signaling packet exchange between the base station and the user, buffering overhead for storing data while the bandwidth assignment is being adjusted) and processing overhead at the base station. In the simulation, it was assumed that the overhead of the proposed scheme is comparable to that of the alternative scheme (Scheme B). Since Scheme A does not employ any control, there is no control overhead associated with Scheme A. Simulation results show that even though the proposed scheme results in lower bandwidth utilization than other schemes due to the bandwidth reservation, it achieves the lower CDP than the other schemes for the same CBP value. In estimating the performance of the proposed scheme in a real network environment, the impact of the bandwidth reservation overheads needs to be carefully evaluated.

5.2 Trade-offs between Three Models of the Proposed Scheme Discussed in Section 4.1

In this subsection, trade-offs between three models of the proposed scheme discussed in Section 4.1 are investigated. Three models investigated are: (1) Uniform and Bandwidth-Based Reservation Model, (2) Movement-Based and Bandwidth-Based Reservation Model and (3) Movement-Based and Number-of-Connections-Based Reservation Model. Refer to Section 4.1 for details of each model.

Figure 6 shows the CDP of Class I hand-off connections for each of the three models of the proposed
scheme. The maximum size of the reserved bandwidth pool is assumed to be 10% of the cell’s total bandwidth (i.e., $r = 10\%$). In this figure, the arrival rate of new Class II connection requests is kept constant at 0.05 requests per second per cell, and the CDP is obtained by varying only the arrival rate of new Class I connection requests. This figure shows that as the connection arrival rate increases, the schemes which use the user’s movement patterns (Movement-Based Reservation Models) generally produce smaller CDPs than the scheme which does not use such information (Uniform Reservation Model). When the connection arrival rate is very small, Movement-Based and Number-of-Connections-Based Reservation Model results in the largest CDPs.

Figures 7 and 8 show the bandwidth utilization and CDP of Class I hand-off connections, respectively, as a function of $r$ (the maximum percentage of cell bandwidth to be reserved). In these figures, the connection arrival rate is kept constant at 0.1 connection requests per second per cell composed of 0.05 Class I and 0.05 Class II connection requests per second. Note that as more bandwidth is reserved (i.e., as $r$ increases), CDP is reduced although the bandwidth utilization decreases.

Figures 7 and 8 also show that the Movement-Based and Bandwidth-Based Reservation Model exhibits better overall performance than the other two models; it produces very small CDPs with reasonable bandwidth utilization. The Uniform and Bandwidth-Based Reservation Model produces poor bandwidth utilization, and Movement-Based and Number-of-Connections-Based Reservation Model produces poor
In comparing three different models of the proposed scheme in this subsection, simulation results show that the proposed scheme provides significantly smaller CDPs when the amount of bandwidth to reserve is a function of the bandwidth requested by all Class I connections than when it is a function of the number of Class I connections. This is because multimedia traffic is an aggregation of heterogeneous traffic whose bandwidth requirements can differ significantly.

Simulation results also show that use of user movement patterns results in higher bandwidth utilization. However, overhead associated with obtaining accurate user movement patterns need to be carefully considered in implementing the proposed scheme in a real network environment. Applying user movement patterns allow reserving bandwidth in selective cells and thus reduce the number of control packets to be exchanged between base stations. At the same time, it also requires a significant overhead if user movement patterns need to be monitored on a real-time basis. In implementing the proposed scheme, the overheads of various design options need to be carefully evaluated.
5.3 Investigation of Adaptive Feature of The Proposed Scheme

The proposed scheme is capable of adapting the size of the reserved bandwidth pool based on the network conditions. This feature may be used to optimize the bandwidth utilization and to provide the desired CDP under different network conditions. In order to investigate how well the proposed scheme adapts to network conditions, the adaptive and non-adaptive versions of the proposed scheme are compared. Please note that the non-adaptive version of the proposed scheme was simulated in Sections 5.1 and 5.2. Since the Non-Adaptive scheme does not adjust the size of reserved bandwidth pool based on network conditions, the value of the parameter $\tau$ must be adjusted to meet the maximum acceptable dropping probability of Class I connections. Therefore, in the following simulations, the results from Figure 8 are used to determine the value of $\tau$ in order to meet the maximum acceptable dropping probability for Class I connections. In the following simulation results, the Movement-Based and Bandwidth-Based Reservation Model is used, since it presents better overall performance than the other two models (see Section 5.2).

Figure 9 shows the CBP and CDP of Class I connections as a function of the connection arrival rate. The maximum acceptable CDP of Class I connections is assumed to be 0.05. In this figure, the maximum acceptable CDP of Class II connections is kept constant at 0.1. This figure shows that when the connection...
arrival rate is small, the Adaptive and Non-Adaptive schemes result in similar CBP and CDP. However, as the connection arrival rate increases, the performance difference between the two schemes becomes significant. When the connection arrival rate is very large (e.g., arrival rate equal to 1 connection request per second), the Non-Adaptive scheme results in a CDP much smaller than the maximum acceptable CDP of 0.05. This implies that with the Non-Adaptive scheme, more bandwidth is reserved than what is needed to provide the maximum acceptable CDP of Class I hand-off connections, resulting in a larger CBP for the new arriving connections. The Adaptive scheme utilizes the bandwidth more efficiently, resulting in smaller CBP values than the Non-Adaptive scheme.

Since the Non-Adaptive scheme reserves bandwidth excessively for Class I connections, it leaves less bandwidth for Class II hand-off connections and results in a larger CDP of Class II connections. Figure 10 shows the CDP of Class II hand-off connections as a function of the connection arrival rate. As expected, the Non-Adaptive scheme results in a larger CDP of Class II hand-off connections.

In order to show how much bandwidth each of two schemes actually reserves, Figure 11 plots the average amount of bandwidth reserved as a function of the connection arrival rate. Again, two values, 0.05 and 0.01, are used for the maximum acceptable CDP of Class I hand-off connections. The maximum acceptable CDP of Class II connections is kept constant at 0.1. Note that in Figure 11, the average amount of bandwidth reserved by the Non-Adaptive scheme increases as the connection arrival rate increases up to
0.01 and then remains constant. This is because the ceiling on the reserved bandwidth pool, $r$, is reached when the connection arrival rate becomes larger than 0.01. On the other hand, the average amount of bandwidth reserved by the Adaptive scheme varies according to the connection arrival rate. This is because the Adaptive scheme adjusts the size of reserved bandwidth pool based on the network conditions, namely, it reserves bandwidth only when necessary. Note that in Figure 11, the bandwidth reserved by the Non-Adaptive scheme beyond the amount reserved by the Adaptive scheme corresponds to the bandwidth which is more than what is needed to provide the requested CDP.

To further investigate the adaptability of the proposed scheme, the source/sink configuration described in Section 4.3 is also simulated. Figure 12 shows the average amount of bandwidth reserved in three types of cells (a source cell, a sink cell and a regular cell) when the source/sink model is used. The maximum acceptable CDP of Class I hand-off connections is assumed to be 0.05, and the maximum acceptable CDP of Class II connections is kept constant at 0.1. Note that in this figure the Non-Adaptive scheme reserves approximately the same amount of bandwidth in three different types of cells. On the other hand, the Adaptive scheme reserves significantly different amounts of bandwidth. The Adaptive scheme reserves the least amount of bandwidth in a source cell since the probability that a connection is handed-off to the source cell is very small. The Adaptive scheme reserves the largest amount of bandwidth in a sink cell since the probability that a connection is handed off to the sink cell is very large. This figure illustrates that the
proposed Adaptive scheme adapts to the current network conditions and changes the amount of reserved bandwidth accordingly.

In this subsection, the Adaptive and Non-Adaptive versions of the proposed scheme are compared. Both schemes (Adaptive scheme and Non-Adaptive scheme) reserve bandwidth to provide QOS guarantees. Even though bandwidth reservation is an effective method to achieve QOS guarantees, it may result in low bandwidth utilization, especially when the network conditions change rapidly. The Adaptive scheme, therefore, is designed to reserve bandwidth dynamically adjusting the amount of reserved bandwidth based on the measured CDP and reserved bandwidth utilization to achieve high bandwidth efficiency. Furthermore, the Adaptive scheme can quickly respond to changes in network load, by increasing or decreasing the amount of reserved bandwidth based on the current network load conditions as well. Simulation results presented in this subsection show that the Adaptive scheme adapts to the current network conditions and achieves high bandwidth utilization. In the Adaptive scheme, there is extra processing overhead associated with measuring the CDP and the reserved bandwidth usage. This processing is not required in the Non-Adaptive scheme since the Non-Adaptive scheme does not perform any real-time measurements. This overhead, however, should not be significant.
5.4 Quality of Service Guarantee for Class II Connections

It is shown in the previous sections that the proposed scheme achieves the requested CDP of Class I connections. This is done by reserving bandwidth for Class I hand-off connections, although this reduces the available bandwidth for Class II connections. In order to provide a certain level of performance for Class II hand-off connections, it may be necessary to set aside a fixed amount of bandwidth in each cell for Class II hand-off connections. Such scheme is simulated in this subsection. This scheme is exactly same as the proposed (Adaptive) scheme except that a fixed amount of bandwidth is pre-reserved in each cell for Class II hand-off connections.

Figure 13 shows the performance of the scheme described above (denoted as “fixed set-aside bandwidth scheme” in the figure) obtained through simulations. In the simulation, the average of the desired amount of bandwidth for all Class II connections is set aside for Class II hand-off connections. It is also assumed that the bandwidth set aside for Class II hand-off connections can also be used by Class I hand-offs when needed. This figures shows the CDP of Class II connections, for two different requested Class I CDP values, 0.05 and 0.01. The requested CDP of Class II connections is assumed to be 0.1. It can be seen that the fixed set-aside bandwidth scheme provides the requested CDP of Class II connections, while maintaining the requested CDP of Class I connections. This is because that the adaptive feature of the fixed set-aside bandwidth scheme allows efficient bandwidth utilization even when some bandwidth is set aside for Class II hand-off connections.

6 Conclusions

The provision of QOS guarantees in future personal communication systems is a complex problem. In this paper, an admission control scheme based on adaptive bandwidth reservation has been proposed to provide QOS guarantees in multimedia personal communication systems. The proposed scheme provides the QOS guarantee by reserving bandwidth in cells surrounding the cell where a connection is established. When a user moves to a new cell, the reserved bandwidth is utilized and the bandwidth reservation process is repeated. Bandwidth is reserved in the new neighboring cells, and the reserved bandwidth in the cells which are no longer neighboring to the new cell are released.

It is shown through simulations that the proposed scheme provides a substantially lower connection dropping probability under realistic circumstances than the schemes without bandwidth reservation. It is also shown that by adjusting the amount of reserved bandwidth based on the current network conditions, the proposed scheme can achieve higher bandwidth utilization.
References


