Computer Communications 25 (2002) 1153–1161

computer communications

www.elsevier.com/locate/comcom

Bandwidth degradation QoS provisioning for adaptive multimedia in wireless/mobile networks

Y. Xiao[a,*], C.L.P. Chen[b], B. Wang[b]

[a]Micro Linear-Utah Development Center, 11734 South Election Road, Suite 200, Draper, UT 84020, USA
[b]Department of Computer Science and Engineering, Wright State University, Dayton, OH 45435, USA

Received 6 June 2001; revised 27 November 2001; accepted 5 December 2001

Abstract

Adaptive multimedia is promising in wireless/mobile networks since it mitigates the fluctuation of resources caused by mobility in these networks. However, bandwidth adaptation entails the bandwidth degradation for some applications. In order to effectively characterize the bandwidth degradation and to provide better quality of service (QoS) to users, we propose two new QoS parameters for adaptive multimedia: the degradation ratio (DR) and the degradation degree (DD), which characterize the frequency of bandwidth degradation and the degree of bandwidth degradation, respectively. Based on DD and DR, a measurement-based call admission control framework and a K-level bandwidth adaptation algorithm are proposed to satisfy the application QoS requirements in terms of the proposed parameters, and to utilize the network resources efficiently. Simulations reveal that DD and DR outperform other QoS parameters in terms of effectively characterizing bandwidth adaptation. Simulations also indicate that the adaptive multimedia framework outperforms the non-adaptive multimedia services. © 2002 Elsevier Science B.V. All rights reserved.

Keywords: Quality of service; Adaptive multimedia; Call admission control; Bandwidth adaptation

1. Introduction

Recently, there have been great demands for multimedia applications with quality of service (QoS) in wireless/mobile networks. QoS provisioning in wireless/mobile networks is more challenging than in wired networks due to channel fading, inherent mobility, and so forth [14]. Even though channel fading can be improved with better transmission and reception systems, mobility may cause severe fluctuation of network resources [10]. For adaptive multimedia networking, the bandwidth of an ongoing call is variable during its lifetime. Adaptive multimedia services are very attractive due to the scarcity in wireless/mobile resources, the availability of a wide range of bandwidth, and their ability to mitigate the resource fluctuation in wireless/mobile networks.

For adaptive multimedia services, existing important QoS parameters for non-multimedia or non-adaptive multimedia, such as the forced termination probability (FTP) and the handoff dropping probability (HDP), become trivial to guarantee. The reason is that the adaptive framework moves the problems of handoff dropping and forced termination into the problem of bandwidth degradation caused by adaptation. We refer to a call as degraded if its assigned bandwidth is below its requested bandwidth. A new QoS parameter, the degradation period ratio (DPR), is proposed by Kwon et al. [5,10]. DPR represents the portion of a call's lifetime during which the call is degraded [5,10]. However, DPR does not characterize the degree of degradation. In this paper, in order to effectively characterize the bandwidth degradation and to provide better QoS to service users, we propose two novel QoS parameters for adaptive multimedia: the degradation ratio (DR) and the degradation degree (DD). The two new QoS parameters characterize both the frequency of bandwidth degradation and the degree of bandwidth degradation. We derive the measurement-based formulas for DR and DD using the time averaging method, which takes into account the observed history of the system resource usage.

One of the key mechanisms for providing QoS guarantees is call admission control (CAC) that enables efficient system resource utilization while application QoS requirements are satisfied. Many CAC schemes in wireless/mobile networks have been proposed in Refs. [1–5,7–10,13,17–19]. Some studies focus on CAC schemes on new calls and handoff calls in a non-multimedia situation [2,3]. Quite a few studies seek CAC schemes for multimedia services in a non-adaptive

* Corresponding author. Tel.: +1-801-617-2518; fax: +1-801-617-2501.
E-mail address: yangxiao@acm.org (Y. Xiao).

0140-3664/02/$ - see front matter
PII: S0140-3664(02)00006-3

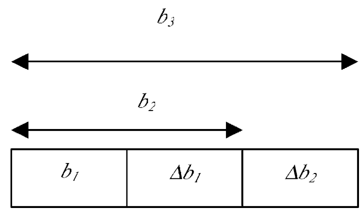

Fig. 1. Example of an adaptive multimedia stream.

multimedia situation [1,8,9]. Most recently, many researches have focused on adaptive multimedia services [4,5,10,17–19]. Kwon et al. [5,10] propose a CAC scheme based on DPR discussed earlier. However, their scheme does not characterize the degree of degradation. Our previous work [17–19] proposes an optimal CAC approach that optimizes revenue for service providers and satisfies QoS requirements in terms of upper bounds of HDPs. The proposed approach is based on the semi-Markov decision process [6,11] approach without considering the bandwidth degradation aspect. In this paper, we propose a distributed and measurement-based CAC scheme based on DR and DD for adaptive multimedia services to satisfy QoS requirements, and to utilize network resources efficiently. The proposed scheme takes into account the observed history of system resource usage, and puts more weight on the current status of the system.

For adaptive multimedia services, a bandwidth adaptation algorithm is needed to adjust the calls' bandwidth in a cell. Therefore, we propose two bandwidth adaptation algorithms based on DR and DD: the two-level bandwidth adaptation algorithm (TL-BAA) and K-level bandwidth adaptation algorithm (KL-BAA). In TL-BAA, there are two bandwidth increment levels and two bandwidth decrement levels, and the two levels represent the level of the requested bandwidth and the level of the minimum bandwidth. In KL-BAA, there are K bandwidth increment levels and K bandwidth decrement levels, and the K levels represent the K layers of multimedia bandwidth. The design philosophy is to minimize DR and DD at any time, and to efficiently utilize the system resources.

The rest of this paper is organized as follows. The traffic model for adaptive multimedia is presented in Section 2. Section 3 proposes two novel QoS parameters. Section 4 describes the proposed measurement-based CAC. Section 5 presents the bandwidth adaptation algorithms. Simulation results and performance comparisons are reported in Section 6. Finally, we conclude this paper in Section 7.

2. Adaptive multimedia traffic model

In adaptive multimedia networking, a multimedia call can dynamically adjust its bandwidth depending on the network load situation during its lifetime. We consider only one class of users in this work. However, our proposed algorithms can be easily extended to multiple user classes. We assume that the bandwidth of a call takes its value from $\{b_1, b_2, \ldots, b_i, \ldots, b_K\}$, where $b_j < b_{j+1}$ for all $j = 1, 2, \ldots, K-1$, and K is the number of multimedia layers [5,10,17]. Note that the above 'layer' concept is not the 'layer' concept used in the layered coding approaches [24–27]. Therefore, it does not necessarily stand for a particular video encoding method or any particular multimedia encoding methods, but is a general abstraction of adaptive multimedia that can be expressed in terms of layers. For example, if the bandwidth of a low quality voice call is b_1 and the bandwidth of a high quality voice call is b_2, then the bandwidth of a voice call may take its value from $\{b_1, b_2\}$. If the cell is not overloaded, a call can be assigned b_2, that is, the bandwidth of a high quality voice call. Otherwise, the call may be assigned b_1. Such a scheme can be implemented by the Sub-Rating method [28].

Another example is the layered multiresolution coding approach, where multimedia receivers can selectively choose to receive a subset of the layer-encoded information depending on receivers' capacity [4,5,10,17–19]. Under such a coding approach, $b_{i+1} - b_i$ is the bandwidth of the ith enhanced layer of the multimedia stream. Such layered multiresolution coding techniques could be subband [24–26] or pyramid coding [27]. The substream filtering function used to filter out the higher enhanced layers can be implemented in the base station [14,29]. How the filtering function is implemented is beyond the scope of the paper. Readers are referred to Refs. [14,29] for details.

Fig. 1 shows another example [4] of an adaptive multimedia stream where $K = 3$. If a cell is lightly loaded and sufficient bandwidth is available, the call will be allocated its maximum bandwidth b_3. Otherwise, depending on the cell's load condition, the call may be allocated b_2 or b_1. Δb_1 and Δb_2 are, respectively, the first and the second enhanced segments of the multimedia stream in addition to the minimum bandwidth b_1. Examples could be {*low quality video*, *medium quality video*, *high quality video*}, or {*low quality audio*, *medium quality audio*, *high quality audio*}, etc.

System events include arrival events and service departure events. Arrival events include new call arrival events and handoff call arrival events. Service departure events include call completion events and events of call handoffs to other cells. We consider fixed capacity in each cell as in previous related work. The fixed total number of channels is C. With this background, we propose two novel degradation QoS parameters in Section 3.

3. Novel degradation QoS parameters

The most significant QoS parameter in non-multimedia or non-adaptive multimedia services is the FTP, the probability of terminating an ongoing call before the completion of the service. However, in adaptive multimedia services, FTP can be near zero [14,15,20,21] as long as we design the CAC scheme and the bandwidth adaptation algorithm properly. There are two other common QoS parameters. The first one is the call blocking probability (CBP), the probability of a new arriving call being blocked. The second is the HDP, the probability of a handoff arriving call being dropped. HDP is more important than CBP since it is directly related to FTP, and service users do not like their calls terminated suddenly just because of changing cells. On the other hand, the fact that a new call is rejected is easier to be accepted by service users compared to handoff dropping. We can see from our simulation results that HDP can be near zero too due to the adaptive nature coupled with our proposed CAC scheme and the proposed bandwidth adaptation algorithms. For adaptive multimedia services, the problems of handoff dropping and forced termination in traditional wireless networks are converted to the problem of bandwidth degradation caused by the adaptation.

Kwon et al. [5,10] proposed the DPR, a new QoS parameter that represents the portion of a call's lifetime during which a call is degraded [5,10]. However, DPR does not characterize the degree of bandwidth degradation. In this paper, in order to provide better QoS to service users, new QoS parameters need to be designed to effectively characterize the bandwidth degradation. Therefore, we propose two new QoS parameters: the DR and the DD for adaptive multimedia services. The new QoS parameters characterize both the frequency of bandwidth degradation and the degree of bandwidth degradation.

The requested bandwidth of a user is denoted as b_{req}, where $b_{\mathrm{req}} \in \{b_1, b_2, \ldots, b_i, \ldots, b_K\}$. We assume that all the users in the same class use the same requested bandwidth, and the requested bandwidth is pre-defined. We also assume $b_{\mathrm{req}} > b_1$. Otherwise, the adaptive multimedia service degenerates to a non-adaptive one, and the new QoS parameters will be so trivial that they are always satisfied.

Let $x(t)$ denote the number of calls in a cell at time t. Let $b_{\mathrm{assi}}(i,t)$ denote the assigned bandwidth for call i at time t, where $b_{\mathrm{assi}}(i,t) \in \{b_1, b_2, \ldots, b_i, \ldots, b_K\}$ and $1 \le i \le x(t)$. If $b_{\mathrm{assi}}(i,t) < b_{\mathrm{req}}$, we refer to the call as degraded.

Let $I(\cdot)$ denote the indicator function defined as follows:

$$I(\text{expression}) = \begin{cases} 1, & \text{if expression} = \text{true} \\ 0, & \text{if expression} = \text{false} \end{cases} \tag{1}$$

Let $x_{\mathrm{d}}(t)$ denote the number of degraded calls out of $x(t)$ calls in the cell at time t. Therefore, we have

$$x_{\mathrm{d}}(t) = \sum_{i=1}^{x(t)} I(b_{\mathrm{assi}}(i,t) < b_{\mathrm{req}}) \tag{2}$$

At time t, the instant degraded ratio (IDR) is defined as:

$$\mathrm{IDR}(t) = \frac{x_{\mathrm{d}}(t)}{x(t)} = \frac{\sum_{i=1}^{x(t)} I(b_{\mathrm{assi}}(i,t) < b_{\mathrm{req}})}{x(t)} \tag{3}$$

The amount of bandwidth reduced is $b_{\mathrm{req}} - b_{\mathrm{assi}}(i,t)$, if $b_{\mathrm{assi}}(i,t) < b_{\mathrm{req}}$. At time t, the instant degradation degree (IDD) is defined as:

$$\mathrm{IDD}(t) = \frac{\frac{1}{x_{\mathrm{d}}(t)} \sum_{i=1}^{x(t)} [b_{\mathrm{req}} - b_{\mathrm{assi}}(i,t)] I(b_{\mathrm{assi}}(i,t) < b_{\mathrm{req}})}{\frac{1}{x_{\mathrm{d}}(t)} \sum_{i=1}^{x(t)} (b_{\mathrm{req}} - b_1) I(b_{\mathrm{assi}}(i,t) < b_{\mathrm{req}})} = \frac{\sum_{i=1}^{x(t)} [b_{\mathrm{req}} - b_{\mathrm{assi}}(i,t)] I(b_{\mathrm{assi}}(i,t) < b_{\mathrm{req}})}{(b_{\mathrm{req}} - b_1) \sum_{i=1}^{x(t)} I(b_{\mathrm{assi}}(i,t) < b_{\mathrm{req}})} \tag{4}$$

The denominator of IDD(t) is used for normalization purpose. Both IDD(t) and IDR(t) are random processes. The DR(t) and the DD(t) are defined as the time averages of IDR(t) and IDD(t), respectively, and they reflect the observed history of the system resource usage.

$$\mathrm{DR}(\tau) = \frac{1}{\Delta T} \int_{\tau - \Delta T}^{\tau} \frac{\sum_{i=1}^{x(t)} I(b_{\mathrm{assi}}(i,t) < b_{\mathrm{req}})}{x(t)} \mathrm{d}t \tag{5}$$

$$\mathrm{DD}(\tau) = \frac{1}{\Delta T} \int_{\tau - \Delta T}^{\tau} \frac{\sum_{i=1}^{x(t)} [b_{\mathrm{req}} - b_{\mathrm{assi}}(i,t)] I(b_{\mathrm{assi}}(i,t) < b_{\mathrm{req}})}{(b_{\mathrm{req}} - b_1) \sum_{i=1}^{x(t)} I(b_{\mathrm{assi}}(i,t) < b_{\mathrm{req}})} \mathrm{d}t \tag{6}$$

Here, ΔT is a time interval for measurement, and τ is a time variable. Both DR(t) and DD(t) are also random processes. Since the events for changing $x(t)$ happen discretely in time, the above integrations become discrete summations when implemented. Both DR and DD take values ranging from 0.0 to 1.0. The smaller the values of DR and DD are, the better QoS is.

In order to better understand Eq. (6), if, as a special case, we assume that at all times $b_{\mathrm{assi}}(i,t)$ equals a constant, i.e. $b_{\mathrm{assi}}(i,t) = b_{\mathrm{assi}}$, we have

$$\begin{aligned} \mathrm{DD}(\tau) &= \frac{1}{\Delta T} \int_{\tau - \Delta T}^{\tau} \frac{\sum_{i=1}^{x(t)} [b_{\mathrm{req}} - b_{\mathrm{assi}}] I(b_{\mathrm{assi}} < b_{\mathrm{req}})}{(b_{\mathrm{req}} - b_1) \sum_{i=1}^{x(t)} I(b_{\mathrm{assi}} < b_{\mathrm{req}})} \mathrm{d}t \\ &= \frac{1}{\Delta T} \int_{\tau - \Delta T}^{\tau} \frac{(b_{\mathrm{req}} - b_{\mathrm{assi}}) I(b_{\mathrm{assi}} < b_{\mathrm{req}}) x(t)}{(b_{\mathrm{req}} - b_1) I(b_{\mathrm{assi}} < b_{\mathrm{req}}) x(t)} \mathrm{d}t \\ &= \frac{(b_{\mathrm{req}} - b_{\mathrm{assi}})}{(b_{\mathrm{req}} - b_1)} I(b_{\mathrm{assi}} < b_{\mathrm{req}}) \end{aligned} \tag{7}$$

Under such a condition, DD equals 0.0 if $b_{\mathrm{assi}} \ge b_{\mathrm{req}}$, DD equals 1.0 if $b_{\mathrm{assi}} = b_1$, and $x_{\mathrm{d}}(t)$ equals either 0 or $x(t)$.

We summarize all the QoS parameters for adaptive

Table 1
QoS parameters summary

QoS parameters	Measurement-based CAC, and bandwidth adaptation algorithm	Simulation results	Comments
Forced termination probability			Near zero
Call blocking probability		Used	Not very important
Handoff dropping probability		Used	Near zero
Degradation ratio	Used	Used	Important
Degradation degree	Used	Used	Important

multimedia services to be used in the proposed measurement-based CAC, bandwidth adaptation algorithms, and simulations in Table 1. The measurement-based CAC introduced in Section 4 will be based on QoS parameters: DR and DD. In our simulation results, we will show the results of all QoS parameters except FTP, which is near zero as we discussed earlier in this section.

4. Measurement-based call admission control

In this section, we provide a distributed and measurement-based CAC for wireless adaptive multimedia services. The objectives for CAC schemes are to satisfy QoS requirements and to utilize the system resources efficiently [17–19]. We allow a handoff call to be always accepted. For a new call request, the QoS requirements are upper bounds in terms of the DR and the DD introduced in Section 3. The proposed CAC algorithm is distributed since each cell runs an instance of the same program and utilizes information collected locally. The proposed CAC is measurement-based since we calculate QoS parameters, DR and DD, using the time averages given in Eqs. (5) and (6) based on observed system resource usage. Let $\mathrm{DR}_{\mathrm{qos}}$ and $\mathrm{DD}_{\mathrm{qos}}$ denote the pre-specified upper bounds of DR and DD, respectively. The proposed CAC algorithm seeks to maintain the DR value and the DD value in wireless/mobile networks to be statistically less than the pre-defined values of $\mathrm{DR}_{\mathrm{qos}}$and $\mathrm{DD}_{\mathrm{qos}}$, respectively. Momentarily, the system can have DR larger than $\mathrm{DR}_{\mathrm{qos}}$, and/or DD larger than $\mathrm{DD}_{\mathrm{qos}}$, but in the long run, the system will be such that

$$\overline{\mathrm{DR}(\tau)} \leq \mathrm{DR}_{\mathrm{qos}} \tag{8}$$

$$\overline{\mathrm{DD}(\tau)} \leq \mathrm{DD}_{\mathrm{qos}} \tag{9}$$

We measure the system resource usage at regular intervals (every ΔT time units). During each measurement window, we calculate DD and DR using Eqs. (5) and (6). As stated in Section 3, the integrations become discrete summations when implemented. Let us denote DD_k and DR_k the kth measurement of DD and DR, respectively, where $k \geq 1$. The proposed scheme takes into account the history of previous measurements with different weights. A tunable factor α, where $0 < \alpha < 1$, is introduced to reduce the impact of the historical measurements. Initially, let $P_{\mathrm{DD}}(1) = \mathrm{DD}_1$ and $P_{\mathrm{DR}}(1) = \mathrm{DR}_1$. For the jth measurement, where $j > 1$,

$$P_{\mathrm{DD}}(j) = \alpha P_{\mathrm{DD}}(j-1) + (1-\alpha)\mathrm{DD}_j \tag{10}$$

$$P_{\mathrm{DR}}(j) = \alpha P_{\mathrm{DR}}(j-1) + (1-\alpha)\mathrm{DR}_j \tag{11}$$

The weight α determines how fast the estimated average adapts to the new measurement. A larger α results in a faster reaction to the network changes. As a special case when α goes to zero, we have $P_{\mathrm{DD}}(j) = \mathrm{DD}_j$ and $P_{\mathrm{DR}}(j) = \mathrm{DR}_j$, which are the results of current measurement window. In the proposed scheme, the effects of old measurements disappear eventually. With a larger α, such a scheme can reflect quickly the current status of the system.

We need to further consider that ongoing calls may be non-uniformly distributed among the cells so that averaging $P_{\mathrm{DD}}(j)$ ($P_{\mathrm{DR}}(j)$) among neighboring cells is necessary. The averaging is conducted by putting a larger weight on its own cell and smaller weights on the neighboring cells. The measurement-based CAC algorithm is shown in Table 2, where 'average' means averaging over neighboring cells.

We allow the CAC to accept a call momentarily without any regard of its bandwidth limitation. The acceptance by the CAC does not guarantee the call to be finally accepted. The acceptance of a call also depends on whether or not the bandwidth adaptation algorithm, proposed in Section 5, can allocate enough bandwidth for the 'accepted' request.

Table 2
Measurement-based CAC

1. **If** (Handoff arrival) Accept;
2. **else if** (average($P_{\mathrm{DR}}(j)$) $\leq \mathrm{DR}_{\mathrm{qos}}$ and average($P_{\mathrm{DD}}(j)$) $\leq \mathrm{DD}_{\mathrm{qos}}$) Accept;
3. **else** Reject;

5. Bandwidth adaptation algorithms

A bandwidth adaptation algorithm decides how to adjust the calls' bandwidth in a cell adaptively. The algorithm is activated whenever there is a call arrival acceptance event or a service departure event. With respect to different QoS objectives, several bandwidth adaptation algorithms [4,17–19,25,27] have been proposed and studied. In this work, our

Table 3
Two-level bandwidth adaptation algorithm (TL-BAA) (A: available bandwidth; &&: "and" operator)

Case 1: Call Departure
While ($A > 0$ && degraded calls exist) {
Find the most degraded call;
Assign_at_most (b_{req});
if (its bandwidth does not increase) **break**;}
While ($A > 0$) {
Find the call with the smallest bandwidth;
Assign_at_most (b_K);
if (its bandwidth does not increase) **break**;}

Case 2: Call Arrival
if ($A \geq b_{req}$) Assign_at_most (b_K);
else {
Reduce (b_{req});
if ($A \geq b_{req}$) Assign_at_most (b_K);
else {
Reduce (b_1);
if ($A < b_1$) Reject the call;
else Assign_at_most (b_{req});}}

Reduce(int level)
While ($A < b_{req}$ && Exist a call's bandwidth >*level*) {
Find call with the largest bandwidth;
Reduce its bandwidth to *level*;}

Assign_at_most (int level)
Assign a bandwidth as much as possible but at most *level*;

Table 4
K-level bandwidth adaptation algorithm (KL-BAA) (A: available bandwidth; &&: "and" operator)

Case 1: Call Departure
While ($A > 0$ && degraded calls exist) {
Find the most degraded call;
Increase_one_level ();
if (fail to increase one level) **break**;}
While ($A > 0$) {
Find the call with the smallest bandwidth;
Increase_one_level ();
if (fail to increase one level) **break**;}

Case 2: Call Arrival
if ($A \geq b_{req}$) Assign_at_most (b_K);
else {
Reduce_level_by_level ();
if ($A < b_1$) Reject the call;
else Assign_at_most (b_{req});}}

Increase_one_level ()
/*Assume that the call's bandwidth is b_i*/
if ($b_i < b_K$ && $A \geq b_{i+1} - b_i$) change its bandwidth to b_{i+1};

Reduce_level_by_level ()
While ($A < b_{req}$ && Exist call's bandwidth $>b_1$) {
Find the call with the largest bandwidth;
Reduce_one_level ();}

Reduce_one_level ()
/*Assume that the call's bandwidth is b_i*/
if ($b_i > b_1$) change its bandwidth to b_{i-1};

Assign_at_most (int level)
Assign a bandwidth as much as possible but at most *level*;

objectives are to minimize DR and DD at any time instant, and to efficiently utilize the system resources.

Ideally, every call in a cell should be allocated the maximum bandwidth (b_K) whenever possible. However, if the cell is over-loaded, some of the calls in the cell might receive a bandwidth lower than b_{req}. When a new call or a handoff call arrives, some of the calls already in the cell might be forced to lower their bandwidth (the minimum bandwidth is b_1) to accommodate the newly arrived call. On the other hand, when a call completes or handoffs to other cells, some of the remaining calls in the cell might increase their bandwidth (the maximum bandwidth is b_K). We can use a long couch with a fixed capacity in a meeting room as an analogy [17–19]. We assume that this couch is the only place where people can sit. If there are more people coming to the meeting room, everyone will sit much closer together so that newcomers can be accommodated. If someone leaves, the people remaining on the couch can get more space. The proposed bandwidth adaptation algorithms make use of above ideas to allocate, increase, and decrease bandwidth for the calls in a cell. For accepted arrivals, if there is enough available bandwidth, allocate the amount of bandwidth requested by the call. Otherwise, some calls' bandwidth will be decreased to accommodate the new arrivals. For call departures, the available bandwidth increases, and the algorithm will then selectively increase the bandwidth of some calls in the cell.

The TL-BAA is shown in Table 3, and the KL-BAA is shown in Table 4.

In TL-BAA, there are two bandwidth increment levels and two bandwidth decrement levels. The first level bandwidth increment is to increase the degraded calls' bandwidth to b_{req}. The second level bandwidth increment is to increase some calls' bandwidth to b_K. On the other hand, the first level bandwidth decrement is to decrease some calls' with sufficient bandwidth to b_{req}. The second level bandwidth decrement is to further decrease some calls' bandwidth to b_1.

In KL-BAA, if we assume a call's bandwidth being b_i, where $b_i \in \{b_1, b_2, \ldots, b_j, \ldots, b_K\}$, we define one level decrement as when the call's new bandwidth becomes b_{i-1}, and similarly, we define one level increment as when the call's new bandwidth becomes b_{i+1}. By 'K-level', we mean that a call's bandwidth can have at most K increment or decrement levels.

6. Performance study

In this section, we present simulation results to show how the proposed approaches satisfy the desired QoS for adaptive multimedia services. We conduct simulations to compare our approaches with non-adaptive multimedia services with respect to utilizations, CBPs, and HDPs in Section 6.1. In Section 6.2, we show by simulations that

Table 5
Simulation parameters

Experiment parameters	Values	Experiment parameters	Values
C	100	$1/\mu$	0.0005 s
K	3	$1/h$	0.0002 s
$\{b_1, b_2, b_3\}$	$\{1, 5, 10\}$	T	500 s
b_{req}	5	α	0.7
b_n	9	DR_{qos}	0.01
DD_{qos}	0.01	UP	93

DD and DR outperform other QoS parameters [4,5,10] in terms of effectively characterizing bandwidth adaptation. We also compare the proposed approach with the optimal CAC approach [17–19] in Section 6.2.

The cellular network studied is assumed to consist of 10 cells forming a ring, a common topology used in many previous studies [2,22,23].

In this simulation, we assume that call requests arrive according to a Poisson distribution. Each call has a service time (call holding time) that is exponentially distributed. The arrival distribution and the service distribution are independent of each other. Let λ_n denote the new call arrival rate; λ_h denotes the handoff call arrival rate; μ denotes the service rate; h denotes the rate of call handoff to other cells. Simulations are carried out under traffic load ρ. We assume that the handoff arrival rate is proportional to the new call arrival rate, $\lambda_h = \beta\lambda_n$ [5,10,19].

6.1. Case study one: performance study and comparison with non-adaptive multimedia

In this section, we present simulation results to show how the proposed approaches satisfy the desired QoS for adaptive multimedia services. We conduct simulations to compare our approaches with non-adaptive multimedia services. Simulation parameters are illustrated in Table 4. T is the simulation time, C is the capacity, ΔT is the measurement time period, b_n is the bandwidth for non-adaptive multimedia calls. For non-adaptive multimedia services, CAC and bandwidth allocation algorithm are defined as follows. For a handoff arrival, accept the call if the amount of available bandwidth is enough; for new arrivals, the upper limit (UP) method [13] is used, in which a new call request is blocked if the number of the calls is greater than an upper limit value. The upper limited value, UP, is defined in Table 5.

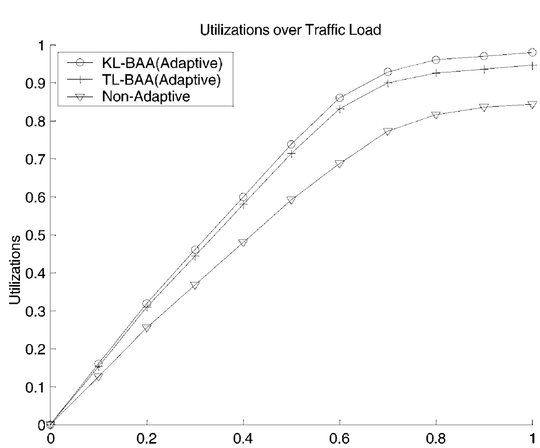

Fig. 2. Utilization versus traffic load for adaptive/non-adaptive multimedia.

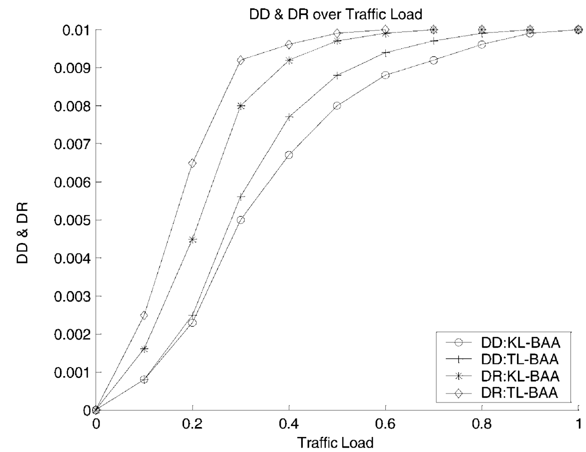

Fig. 3. DD and DR versus traffic load.

Fig. 2 shows the utilization versus the traffic load ρ for adaptive multimedia and non-adaptive multimedia services. Clearly, the utilization for adaptive multimedia is better than that for non-adaptive multimedia. When the traffic load becomes higher, the advantage is more evident. The reason for multimedia services' better resource utilization is due to the very adaptive nature of these services that allows the network to offer services whenever there is sufficient amount of resources by intelligently adjusting resource allocation. Moreover, the utilization is higher in KL-BAA scheme than in TL-BAA scheme, which indicates that KL-BAA offers more flexibility of bandwidth adjustment, thus resulting in better resource utilization.

Fig. 3 shows the QoS parameters, DD and DR, versus traffic load ρ for adaptive multimedia services. The requested QoS bounds for DD and DR are DD_{qos} (0.01) and DR_{qos} (0.01), respectively. From the figure, we observe that the desired QoS requirements for adaptive multimedia services are satisfied. Moreover, the KL-BAA scheme offers better assurance than the TL-BAA scheme does.

Fig. 4 depicts the CBP versus the traffic load for adaptive multimedia (using KL-BAA) and non-adaptive multimedia services, respectively. Clearly, CBP performance for adaptive multimedia is better than that for non-adaptive multimedia due to the CAC scheme and the bandwidth adaptation algorithm employed.

Fig. 5 shows the HDP for adaptive multimedia (using KL-BAA) and non-adaptive multimedia services. The figure indicates that HDP is near zero under low traffic loads. This is trivially true by the nature of the algorithm as per our discussion in earlier sections. However, if the

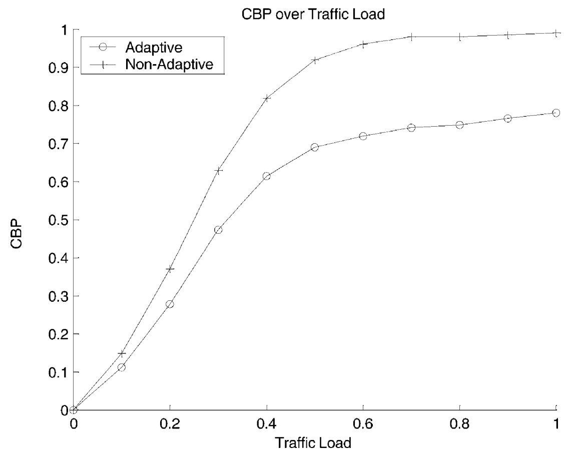

Fig. 4. CBP versus traffic load for adaptive/non-adaptive multimedia.

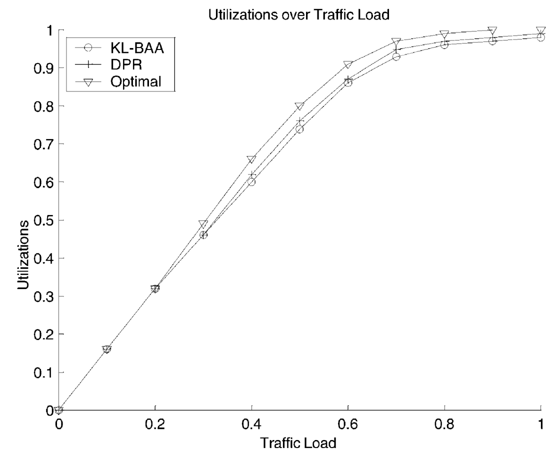

Fig. 6. Utilizations for three adaptive approaches.

traffic is very heavy, HDP is non-zero. For non-adaptive multimedia, HDP is significant even under low traffic loads.

In summary, from a service provider's point of view, the proposed methods outperform the non-adaptive multimedia services in terms of utilization and HDP while QoS requirements are satisfied. One reason for this is that we allow the multimedia service to be adaptive in order to mitigate the fluctuation of the system resources. Other reasons, such as the CAC scheme chosen and the bandwidth adaptation algorithm used also play very important roles. On the other hand, from a service user's point of view, QoS is improved by adopting new QoS parameters to effectively characterize the bandwidth degradation caused by adaptation.

6.2. Case study two: comparison with other adaptive multimedia schemes

In this section, we compare the performance of DD and DR with that of the QoS parameter DPR [4,10], which represents the portion of a call's lifetime during which the call is degraded. A performance metric is introduced to compare the effectiveness of characterizing bandwidth adaptation. We adopt the degradation metric, degraded area size (DAS) [30–31], which stands for the normalized mean of the product of the proportion of a call's degraded time and its degree of degradation. We also compare the proposed approach with the optimal CAC approach [17–19]. In Refs. [17–19], the CAC adopts the semi-Markov decision process approach to formulize the utilization and QoS requirements with linear programming (LP). Simulation parameters are the same as listed in Table 5.

Fig. 6 shows the utilization versus the traffic load ρ for the proposed approach (when KL-BAA is used), the DPR based approach, and the optimal approach. When the traffic load is low, the utilization for three approaches are the same. When the traffic load gets higher, the utilization of the optimal approach is the best. However, we observe that the proposed approach approximates the optimal approach quite well, and the three approaches are almost the same.

Fig. 7 shows that the DAS for the three approaches versus the traffic load. We observe that the proposed approach outperforms all other approaches in terms of effectively characterizing bandwidth adaptation.

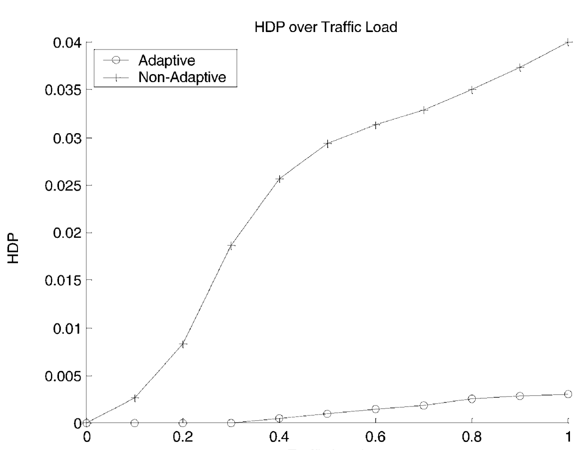

Fig. 5. HDP versus traffic load for adaptive/non-adaptive multimedia.

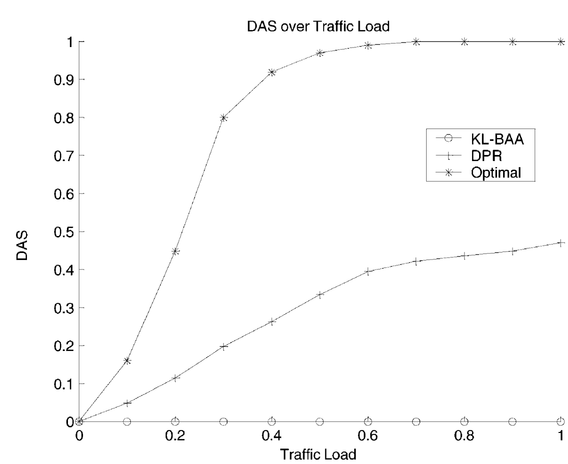

Fig. 7. DAS for three adaptive approaches.

Moreover, the computational complexity of the optimal approach is exponential [19] while the computational complexity of the proposed approach is polynomial.

7. Conclusion and future work

In this paper, we propose two new QoS parameters: the DR and the DD to effectively characterize the bandwidth degradation caused by adaptation in an effort to provide better QoS for adaptive multimedia users. Based on the new QoS parameters, we propose a measurement-based CAC scheme and two bandwidth adaptation algorithms to provide QoS assurance for adaptive multimedia services in wireless/mobile networks. The proposed CAC scheme takes into account observed history of system resource usage, and puts more weights on the current status of the system. Simulations reveal that DD and DR outperform other QoS parameters in terms of effectively characterizing bandwidth adaptation. Simulation results show that, under different traffic loads, QoS requirements are satisfied while high network utilization is achieved K-level bandwidth allocation algorithm outperforms the TL-BAA both in terms of network utilization and QoS assurance. We also compare results with those from the non-adaptive multimedia services. Our findings show that the adaptive multimedia framework significantly outperforms the non-adaptive multimedia services. The proposed approach is compared with the optimal CAC scheme, and the simulation results show that the utilization difference between the optimal approach and the proposed approach is small, and the proposed scheme can keep the bandwidth degradation small while the optimal CAC approach cannot.

We believe that monitoring DD and DR is an innovative policy that can greatly aid cellular network providers in providing better QoS assurance while achieving higher resource utilization. Our work can be extended in several avenues. One straightforward future work is to extend our schemes to networks that support multiple classes of adaptive multimedia services.

References

[1] D. Ayyagari, A. Ephremides, Admission control with priorities: approaches for multi-rate wireless systems, Mobile Networks and Applications 4 (1999) 209–218.

[2] C. Ho, C. Lea, Improving call admission policies in wireless networks, Wireless Networks 5 (1999) 257–265.

[3] R. Ramjee, D. Towsley, R. Nagarajan, On optimal call admission control in cellular networks, Wireless Networks 3 (1997) 29–41.

[4] T. Kwon, I. Park, Y. Choi, S. Das, Bandwidth adaptation algorithms with multi-objectives for adaptive multimedia services in wireless/mobile networks, Proceedings of ACM Workshop on Wireless Mobile Multimedia, WOWMOM'99, pp. 51–58.

[5] T. Kwon, Y. Choi, C. Bisdikian, M. Naghshineh, Call admission control for adaptive multimedia in wireless/mobile network, Proceedings of ACM Workshop on Wireless Mobile Multimedia, WOWMOM'98, Dallas, October 1998, pp. 111–116.

[6] H.C. Tijms, Stochastic Modeling and Analysis: A Computational Approach, Wiley, New York, 1986.

[7] J. Choi, T. Kwon, Y. Choi, M. Naghshineh, Call admission control for multimedia service in mobile cellular networks: a Markov decision approach, IEEE ISCC'00, Antibes, July 2000.

[8] T. Kwon, Y. Choi, M. Naghshineh, Optimal distributed call admission control for multimedia service in mobile cellular network, Proceedings of Mobile Multimedia Conference, MoMuc'98, Berlin, October 1998.

[9] D.I. Yoon, B. Lee, A distributed dynamic call admission control that supports mobility of wireless multimedia users, Proceedings of ICC'99, pp. 1442–1446.

[10] T. Kwon, Y. Choi, C. Bisdikian, M. Naghshineh, QoS provisioning for adaptive multimedia in wireless/mobile networks, Wireless Network, in press.

[11] S.M. Ross, Introduction to Probability Methods, Academic Press, New York, 1997.

[13] S. Biswas, B. Sengupta, Call admissibility for multirate traffic in wireless ATM networks, IEEE INFOCOM'97 (1997) 650–659.

[14] M. Naghshineh, W. Willebeek-LeMair, End-to-end QoS provisioning in multimedia wireless/mobile networks using a adaptive framework, IEEE Communications Magazine 35 (11) (1997) 72–81.

[15] A. Alwan, R. Bagrodia, N. Bambos, M. Gerla, L. Kleinrock, J. Short, J. Villasenor, Adaptive mobile multimedia networks, IEEE Communications Magazine 34 (4) (1996) 34–51.

[17] Y. Xiao, C.L.P. Chen, Y. Wang, An optimal distributed call admission control for adaptive multimedia in wireless/mobile networks, Proceedings of the Eighth International Symposium on Modeling, Analysis and Simulation of Computer and Telecommunication Systems, MASCOTS 2000, San Francisco, CA, USA, August 29–September 1, 2000, pp. 477–482.

[18] Y. Xiao, C.L.P. Chen, Y. Wang, Quality of service provisioning framework for multimedia traffic in wireless/mobile networks, Proceedings of The IEEE Ninth International Conference on Computer Communication and Networks, IEEE IC3N'00, Las Vegas, NV, USA, October 16–18, 2000, pp.644–648.

[19] Y. Xiao, C.L.P. Chen, Y. Wang, Optimal admission control for multiclass of wireless adaptive multimedia services, IEICE Transactions on Communications, Special Issue on Mobile Multimedia communications E84-B (4) (2001) 795–804.

[20] V. Bharghavan, K. Lee, S. Lu, S. Ha, J. Li, D. Dwyer, The TIMELY adaptive resource management architecture, IEEE Personal Communication Magazine 5 (8) (1998).

[21] S.K. Das, S.K. Sen, Quality-of-service degradation strategies in multimedia wireless networks, IEEE Vehicular Technology Conference, VTC'98, Ottawa, Canada, May 1998.

[22] M. Sidi, D. Starobinski, New call blocking versus handoff blocking in cellular networks, Proceedings of IEEE INFOCOM 1996, pp. 35–42.

[23] H. Takanashi, S.S. Rappaport, Dynamic base station selection for personal communication systems with distributed control schemes, Proceedings of Vehicular Technology Conference, 1997, pp. 1787–1791.

[24] M. Vitterli, Multi-dimensional subband coding: some theory and algorithms, Signal Processing 6 (1984).

[25] D. Taubman, A. Zakhor, Multi-rate 3-D subband coding of video, IEEE Transactions on Image Processing 3 (5) (1994) 572–588.

[26] J. Woods (Ed.), Subband Image Coding, Kluwer, Dordrecht, 1991.

[27] P.J. Burt, E.H. Adelson, The Laplacian pyramid as a compact image code, IEEE TransCom COM-31 (4) (1983) 532–540.

[28] Y.-B. Lin, A.R. Noerpel, D.J. Harasty, The sub-rating channel assignment strategy for PCS hand-offs, IEEE Transactions on Vehicular Technology 45 (1) (1996) 122–130.

[29] N. Yeadon, Filters: QoS support mechanism for multipeer communications, IEEE Journal on Selected Areas in Communications 14 (7) (1996) 1245–1262.

[30] Y. Xiao, C.L.P. Chen, Y. Wang, Quality of service and call admission control for adaptive multimedia services in wireless/mobile networks,

Proceedings of the IEEE 2000 National Aerospace and Electronics Conference, pp. 214–220.

[31] Y. Xiao, C.L.P. Chen, Y. Wang, A measurement-based call admission control for adaptive multimedia services in mobile cellular networks, Proceedings of The Conference on Network and Application Performance, OPNETWORK2000, August 28–September 1, Washington, DC, USA.

Dr Yang Xiao received his Ph.D. in computer science and engineering from Wright State University, USA in 2001, M.S. degree in computer science and engineering from Wright State University, USA in 2000, M.S degree concentrating on Computer Graphics from JILIN University, China in 1991, and a B.S. in Mathematics from JILIN University, China in 1989. From 1991 to 1996, he had been a senior software engineer working in the computer industry for five years. From 1996 to 2001, he had been awarded the DAGSI Ph.D. Fellowship. He was a recipient of the 1999 Graduate Student Excellence Award in recognition of outstanding academic achievements in the program of Computer Science and Engineering Ph.D. at Wright State University. Dr Xiao currently works in the Micro Linear Corporation as a MAC architect involving IEEE 802.11 standardization enhancement work. He is a member of ACM, IEEE and IEICE. His current research interests include, but are not limited to, Wired/Wireless Networks and Communications, Analytic Modeling, Performance Evaluation, and Distributed Algorithms.

Dr. Chen received his BSEE from NTUT, Taiwan, R.O.C., an M.S. degree from the University of Michigan, Ann Arbor, Michigan, in 1985, and a Ph.D. degree from Purdue University, West Lafayette, Indiana, in Dec. 1988. In 1988–1989, he was a visiting Assistant Professor at the School of Engineering and Technology, Purdue University, Indianapolis, Indiana. Since September 1989, he has been at the Computer Science and Engineering Department, Wright State University, Dayton, Ohio, where he is currently a Professor. Dr. Chen is also a visiting research scientist at the Materials Directorate, Wright Laboratory, Wright-Patterson Air Force Base. He has been a senior research fellow sponsored by the National Research Council, National Academy of Sciences. He is on sabbatical leave to Purdue University and Case Western Reserve University from fall 1996 – fall 1997. His current research interests and projects, supported by the Air Force Office Scientific Research, the U.S. Air Force, State of Ohio, include computer networks, networking, neural networks, fuzzy-neural systems, intelligent systems, robotics, and CAD/CAM. His current research area with NASA, Glenn Research Center includes data mining on aircraft flight and maintenance data, aircraft engine life prediction, and life extending control.

Dr. Bin Wang received his Ph.D. in electrical engineering from the Ohio State University in 2000, M.S. degree in computer science from the University of Louisville in 1994, and bachelor of engineering degree from Zhejiang University, China in 1992. From September 1994 to August 1996, he was a research assistant in the department of electrical and computer engineering, the University of Wisconsin-Madison. Since September 1996, he has been with the department of electrical engineering, the Ohio State University. He spent the summer of 1998 at Panasonic Information and Networking Technology Laboratory, Princeton, NJ. He joined the department of computer science and engineering, the Wright State University, in September 2000 as an assistant professor. His research interests are communication networks with emphasis on providing quality of service assurance in high-speed networks, wireless and mobile networks, DWDM optical networks, stochastic modeling and queueing analysis of systems, simulation optimization, and network protocol development.