Call Admission Control for Prioritized Adaptive Multimedia Services in Wireless/Mobile Networks

Sooyeon Kim, Taekyoung Kwon, Yanghee Choi
Dept. of Computer Engineering, Seoul National University, Seoul, Korea
Phone: +82-2-880-1832, Fax: +82-2-872-2045
Email: {sykim,tkkwon,yhchoi}@mmlab.snu.ac.kr

Abstract - Recently, the scarce wireless bandwidth motivated adaptive multimedia services where the bandwidth of an ongoing call can be dynamically adjusted. We consider multiple classes of adaptive multimedia services and assume the prioritization among classes in the bandwidth adaptation. A threshold-type call admission control algorithm is proposed here and product-form solution is derived. Also, we formulate a nonlinear programming model to find out threshold values for optimal solution. Numerical results show that the proposed method guarantees maximum revenue while satisfying the quality of service requirements.

I. Introduction

With the advances in wireless communications technology, there has been a growing interest in deploying multimedia services in wireless/mobile networks. Also, multimedia technology (e.g., compression) has matured and gave birth to diverse multimedia services. In spite of these developments, the link bandwidth of wireless/mobile networks is still a bottleneck. This scarcity in wireless resources motivates the adaptive multimedia services which can operate over a wide range of available bandwidth [1], [2], [3], [4].

Originally, the concept of adaptive multimedia service was introduced in wired networks. In wired broadband networks like ATM, once a call is admitted to the network, a contract between the network and application is established. Then, they both try to maintain the contract throughout the call's lifetime. In such a paradigm, network congestion can cause fluctuations in the availability of network resources and thereby can result in severe degradation of multimedia services. To overcome this problem, many adaptive multimedia schemes are proposed such as hierarchical encoding [6] and network filters [7] to mitigate the effect of fluctuation in the network resources.

An adaptive multimedia paradigm can play an important role to mitigate the highly-varying resource availability in wireless/mobile networks. That is, it is possible to overcome the link overload condition by reducing the bandwidth of individual calls, which we call "bandwidth adaptation." Compared to wired networks, the fluctuation in resource availability in wireless/mobile networks is much more severe and results from two inherent features of such networks: fading and mobility. Here we take into consideration only mobility (equivalently, handoff). Therefore, bandwidth adaptation happens when there is a new call arrival (if accepted), a call completion, or an incoming/outgoing handoff call.

More recently, some call admission control (CAC) schemes have been proposed for adaptive multimedia services in wireless/mobile networks [2], [3], [5]. Bharghavan et al. [2] take into consideration service classes in the Internet and seek to exploit the CAC algorithm and the resource reservation scheme with the assumption of the continuous value of bandwidth; however, the real bandwidth values of adaptive multimedia is more likely to be discrete [1], [10]. Das and Sen [3] assume the discrete values of adaptive multimedia and propose an optimal (from a cell's perspective) CAC and bandwidth adaptation algorithm by considering tradeoff between carried traffic and degradation. Stamatelos and V. Koukoulidis [5] propose a CAC in a wireless ATM environment. However, a call of each class can take only two values in this model.

Over the past few years, the prioritization (or "differentiation") among multiple service classes of service in the Internet is massively investigated (e.g., premium service [9]). We believe that this concept will be reflected in wireless/mobile networks in the near future. Thus, we take into consideration the prioritization in our adaptive multimedia framework.

The rest of this paper is organized as follows. Section II describes our model of adaptive multimedia services. A simple bandwidth adaptation algorithm is described in Section III. Section IV introduces our CAC algorithm. Numerical results are presented in Section V. Finally, Section VI concludes our paper.

II. Model Description

A general model of multi-class adaptive multimedia services in this paper is described. Suppose there are K classes of adaptive multimedia services in wireless/mobile networks which adopt a cellular architecture. New call arrivals of class-i ($i = 1, 2, ..., K$) into a cell are assumed to form a Poisson process with mean arrival rate λ_i. The call holding time (CHT) of a class-i call is assumed to follow an exponential distribution with mean $1/\mu_i$. As a simplifying assumption, handoff call arrivals of class-i also form a Poisson process with mean rate $1/h_i$. Here, the priority of a class-i call is proportional to i, with 1 indicating the lowest priority and with K indicating the highest priority. In our adaptive multimedia framework, the lower the priority of a class is, the more preferably the bandwidth of calls of that class are adapted.

For the mobility characterization, we assume the following simple model. The cell residence time (CRT), i.e., the amount of time during which a mobile terminal stays in a cell during a single visit, is assumed to follow an exponential distribution with mean $1/h$. We assume that the CRT is independent of class, hence, calls in any class follow the same CRT distribution. Note that the parameter h represents the mean handoff rate.

According to the adaptive multimedia paradigm, a multimedia call can dynamically change its bandwidth throughout its lifetime depending on the situation. We assume that the bandwidth of a class-i call takes its discrete value from the set $B_i = \{b_{i,1}, b_{i,2}, ..., b_{i,s_i}\}$ where $b_{i,j} < b_{i,j+1}$ for $j = 1, ..., s_i - 1$. Here, the number of possible values of bandwidth that a class-i call can be allocated is s_i. Obviously, $b_{i,1}$ and b_{i,s_i} are the minimum and maximum bandwidth for a class-i call.

The total bandwidth (in number of channels) in each cell is the same and denoted by C, assuming a fixed channel allocation (FCA) scheme. Furthermore, it is assumed that the bandwidth of a multimedia call can be scattered across the bandwidth (channel) pool.

The current state of a cell is represented by the vector $\mathbf{x} = (x_1, x_2, ..., x_K)$ where x_i denotes the number of ongoing class-i calls in the system (cell) for $i = 1, 2, ..., K$. Therefore, we should deal with two issues. One is to determine the set of states of the system, which is the role of a CAC. The other is, for each state, to adapt the bandwidth of each call, which is the role of a bandwidth adaptation algorithm.

III. Bandwidth Adaptation Algorithm

In this section, a simple bandwidth adaptation algorithm (BAA) is described where the bandwidth of calls with lower priority are preferably adapted. Figure 1 describes the proposed algorithm that finds the number of degraded class-i calls for given $\mathbf{x}$. Here $\mathbf{b_{max}} = (b_{1,s_1}, b_{2,s_2} ..., b_{K,s_K})$ is a vector of maximum bandwidth values of each class. Also, d_i is the number of class-i calls whose bandwidth is to be adapted, which is the result of this algorithm. As mentioned earlier, class 1 has the lowest priority, while class K has the highest priority.

```
C_req ← x · b_max, degraded ← FALSE, i ← 0
if (C_req > C) degraded ← TRUE
while (C_req > C) {
    i ← i + 1
    if (i > K) d_i's cannot be found; return
    d_i ← x_i
    C_req ← C_req − d_i * (b_{i,s_i} − b_{i,1})
}
if (degraded) {
    while (C_req + b_{i,s_i} − b_{i,1} ≤ C) {
        d_i ← d_i − 1
        C_req ← C_req + b_{i,s_i} − b_{i,1}
    }
}
```

Fig. 1. Bandwidth Adaptation Algorithm

IV. Call Admission Control

In this section, we introduce a threshold-type CAC algorithm where a newly arriving call of class-i is blocked if the current number of ongoing class-i calls is equal to or greater than t_i. On the other hand, an incoming handoff call of class-i is accepted regardless of the number of ongoing class-i calls if the call can be accommodated; if necessary, bandwidth adaptation is performed. The threshold CAC algorithm is chosen because of simplicity; also, the steady state probability is easily tractable by means of the product form [11]. The steady state probability of each state $\mathbf{x}$ is determined by choosing threshold value (t_i) for each class. Finally, we should find out t_i for all classes to maximize the revenue. Here, for each on-going class-i call, revenue is assumed to be accrued at rate r_i regardless of the currently allocated bandwidth. Also, we take into consideration QoS requirements such as the call blocking probability and the call degradation probability. The call degradation probability is the probability that a call is allocated less than its maximum bandwidth.

Using the product form solution, we can calculate the steady state probability of state $\mathbf{x}$ in the threshold CAC algorithm. With the assumptions in Section II, the steady state probability for state $\mathbf{x}$ is given by

$$\pi(\mathbf{x}) = \frac{1}{G}\prod_{i=1}^{K} p_i(x_i) \tag{1}$$

$$G = \sum_{\forall \mathbf{x}} \prod_{i=1}^{K} p_i(x_i) \tag{2}$$

$$p_i(x_i) = I(x_i \le t_i)\left(\frac{\lambda_i + h_i}{\mu_i + h}\right)^{x_i} / x_i! + I(x_i > t_i)\left(\frac{\lambda_i + h_i}{\mu_i + h}\right)^{t_i}\left(\frac{h_i}{\mu_i + h}\right)^{x_i - t_i} / x_i! \tag{3}$$

Here, G is a normalization constant and $I(\cdot)$ is an indicator function which is 1 if its argument is true and 0 otherwise.

For each state $\mathbf{x}$, the number of degraded class-i calls is denoted by $d_i(\mathbf{x})$ which is determined by our BAA. As a result, the call degradation probability of class-i is expressed by

$$\sum_{\forall \mathbf{x}} \pi(\mathbf{x}) \frac{d_i(\mathbf{x})}{x_i} \ . \tag{4}$$

Also, in our CAC, the call blocking probability of a new class-i call is given by

$$\sum_{\forall x_i \ge t_i} \pi(\mathbf{x}) \ . \tag{5}$$

Note that the call blocking probability applies to only new call. An incoming handoff call is always accepted; if the available bandwidth is insufficient, the bandwidth adaptation is performed to make room for the handoff call. (We implicitly assume that the case in which a handoff call should be forced to be terminated will not happen.)

Considering the above properties, our CAC can be formulated as a nonlinear programming (NLP) problem as shown below. Here, the values of decision variables $t_1, t_2, ..., t_K$ are determined by solving the NLP:

Maximize

$$\sum_{\forall \mathbf{x}} (\mathbf{r} \cdot \mathbf{x}) \pi(\mathbf{x})$$

subject to

$$\sum_{\forall \mathbf{x}, x_i > 0} \pi(\mathbf{x}) \frac{d_i(\mathbf{x})}{x_i} \le P_{D_i} \ ,$$

and

$$\sum_{\forall x_i \geq t_i} \pi(\mathbf{x}) \leq P_{B_i} \ .$$

Here, P_{D_i} and P_{B_i} denote the upper bounds of the call degradation probability and the call blocking probability of a class-i call, respectively. Also, $\mathbf{r} = (r_1, r_2, ..., r_K)$ is a revenue vector.

To solve the above NLP problem, we adopt the "separable programming" technique, which investigates every possible case of threshold values of all classes. Then, all the nonlinear terms in the above NLP can be regarded as constants and thereby the NLP can be changed into LP model.

Let $\mathbf{t} = (t_1, t_2, ..., t_K)$ be a vector of threshold values of each class. Finally, introducing boolean variables $A_{(0,0,...,0)}$, $A_{(1,0,...,0)}$, ..., $A_{(max(t_1),max(t_2),...,max(t_K))}$ instead of $\mathbf{t}$ changes the nonlinear expressions (1)-(3) into the following linear expressions. Here $max(t_i)$ means the maximum possible integer value of threshold of class-i. Note that only one of the above boolean variables is 1 while the others are 0; for example, in three classes, $A_{2,3,5} = 1$ means that $\mathbf{t} = (2, 3, 5)$ is the threshold vector for optimal solution.

$$\sum_{\forall \mathbf{t}} A_{\mathbf{t}} = 1 \tag{6}$$

$$\pi(\mathbf{x}) = \sum_{\forall \mathbf{t}} A_{\mathbf{t}} \left[\frac{1}{G(\mathbf{t})} \prod_{i=1}^{K} p_i(x_i, \mathbf{t}) \right] \tag{7}$$

$$G(\mathbf{t}) = \sum_{\forall \mathbf{x}} \prod_{i=1}^{K} p_i(x_i, \mathbf{t}) \tag{8}$$

$$\begin{aligned} p_i(x_i, \mathbf{t}) = & I(x_i \leq t_i) \left(\frac{\lambda_i + h_i}{\mu_i + h} \right)^{x_i} / x_i! + \\ & I(x_i > t_i) \left(\frac{\lambda_i + h_i}{\mu_i + h} \right)^{t_i} \left(\frac{h_i}{\mu_i + h} \right)^{x_i - t_i} / x_i! \end{aligned} \tag{9}$$

Consequently, the threshold values for the maximum revenue can be found by solving the above linear programming (LP) problem.

V. Numerical results

We performed comprehensive simulation experiments to verify the correctness of our transformation from NLP to LP. In the following results, we simulated the environment in which there are three classes of adaptive multimedia services: voice, low-quality video, and high-quality video. Table I shows the bandwidth values (in kbps) and the revenue rate for each class. Note that the voice class has the highest priority, while the high-quality video class has the lowest quality. The upper bounds of call blocking probabilities are equally set to 0.5 for three classes. Moreover, the upper bound of call degradation probability is set to 0.1 for voice, 0.2 for low-quality video, and 0.3 for high-quality video. The overall bandwidth capacity of a cell is assumed to be 3 Mbps. The new call arrival rate varies between 0.01 and 0.05. Also, the handoff call arrival rate is assumed to be proportional to the new call arrival rate by $h_i = \alpha\lambda_i$, where α is set to 0.5 in the experiments. Throughout the experiments, the mean values of the CHT and the CRT are assumed to be 500 seconds and 100 seconds, respectively.

TABLE I

Bandwidth Values and Revenue Rates

service	B_i	r_i
voice	{64, 80, 96, 112, 128}	1
low-quality video	{128, 160, 192, 224, 256}	3
high-quality video	{256, 288, 320, 352, 384}	8

Table II presents all t_i tuples, when the new call arrival rates are 0.3, which satisfy the upper bounds of both the call blocking probability and the call degradation probability. Here, the tuple (8, 6, 14) marked by * is the solution of our LP model, and it shows that we find the tuple that maximizes the revenue and satisfies the QoS requirements at the same time.

VI. Conclusion

It is anticipated that demands for adaptive multimedia services will grow in the future wireless/mobile networks, especially considering highly fluctuating bandwidth availability. In our adaptive multimedia frame-

TABLE II

Revenues of Different Threshold Sets

(t_1, t_2, t_3)	revenue
(7, 5, 13)	42.805441
(7, 5, 14)	43.929670
(7, 6, 13)	43.509179
(7, 6, 14)	44.633431
(7, 7, 13)	43.949723
(8, 5, 13)	43.349376
(8, 5, 14)	44.473667
(8, 6, 13)	44.053130
(8, 6, 14)*	45.177361
(8, 7, 13)	44.493615

work, the bandwidth of a call takes a value from a set of discrete values depending on situations. We also take into account the prioritization among multiple classes of services; that is, bandwidths of lower priority calls are preferably reduced in overload condition. A threshold-type call admission control algorithm is proposed and product-form solution is derived. Also, we formulate a nonlinear programming (NLP) model to find out threshold values for optimal solution and show that the NLP model can be transformed into a linear programming (LP) model using separable programming. Numerical results verify that the result of the proposed method guarantees the maximum revenue while satisfying the quality of service requirements.

References

[1] M. Naghshineh and M. Willebeek-LeMair, "End-to-End QoS Provisioning in Multimedia Wireless/Mobile Networks Using an Adaptive Framework," IEEE Communications Magazine, Vol. 35, No. 11, pp. 72-81, Nov. 1997.

[2] V. Bharghavan, K. Lee, S. Lu, S. Ha, J. Li, and D. Dwyer, "The TIMELY Adaptive Resource Management Architecture," IEEE Personal Communications Magazine, Vol. 5, No. 8, Aug. 1998.

[3] S. K. Das and S. K. Sen, "Quality-of-Service Degradation Strategies in Multimedia Wireless Networks," IEEE VTC '98, Ottawa, Canada, May 1998.

[4] K. Lee, "Adaptive Network Support for Mobile Multimedia," *ACM MobiCom '95*, pp. 62-74, 1995.

[5] G. Stamatelos and V. Koukoulidis, "Reservation-Based Bandwidth Allocation in a Radio ATM Network," *IEEE/ACM Transactions on Networking*, Vol. 5, No. 3, Jun., 1997.

[6] Nachum Shacham, "Multipoint Communication By Hierarchically Encoded Data," IEEE INFOCOM '92, pp. 2107-2114, 1992.

[7] Nicholas Yeadon, "Filters: QoS Support Mechanisms for Multipeer Communications," *IEEE Journal on Selected Areacs in Communications*, Vol. 14, No. 7, pp. 1245-1262, Sept. 1996.

[8] A. K. Talukdar, B. R. Badrinath, and A. Acharya, "Rate Adaptation Schemes in Networks with Mobile Hosts," ACM/IEEE MobiCom'98, Oct. 1998.

[9] K. Nichols, V. Jacobson, and L. Zhang, "A Two-bit Differentiated Services Architecture for the Internet," Internet Draft, draft-nichols-diff-svc-arch-00.txt, November, 1997

[10] T. Kwon, Y. Choi, C. Bisdikian, and M. Naghshineh, "Call Admission Control for Adaptive Multimedia in Wireless/Mobile Networks," ACM WoWMoM '98, pp. 111-116, Dallas, Oct. 1998.

[11] K. W. Ross, Multiservice Loss Models for Broadband Telecommunication Networks, Springer, 1995.

[12] E. Horowitz et al., Computer Algorithms/C++, Computer Science Press, 1996.