

Adaptive Buffer Sensitive Scheduling for Transmitting Video Streams in a Mobile Multimedia System¹

Joe Yuen, Kam-Yiu Lam, and Edward Chan
Department of Computer Science
City University of Hong Kong
Tat Chee Avenue, Kowloon, Hong Kong
{csjyuen|cskylam|csedchan}@cityu.edu.hk

Abstract

In this paper, we propose the Adaptive Buffer Sensitive (ABS) algorithm for scheduling the transmission of video packets from multiple video streams in a mobile multimedia system. Two important considerations in the design of the ABS are: (1) to maximize the mobile network bandwidth utilization and at the same time to minimize the impact of transient overloading on video playback; and (2) to minimize the impact of a video playback by other video requests. The second objective is very important to mobile multimedia systems since the clients may request videos with great differences in workload characteristics. In the ABS, the allocation of network bandwidth to serve client requests is divided into two phases. Firstly, each client receives equal share of network bandwidth or at least enough bandwidth to support the average playback of its requested video. Secondly, the extra bandwidth after the first allocation will be distributed to clients based on its playback status, i.e., buffer playback duration, to deal with the problem of transient overloading. A selective packet transmission scheme is adopted in ABS to deal with heavy overloading situation.

1. Introduction

Recent advances in mobile communication technologies and mobile devices have greatly increased the functionality of mobile information services and have made many mobile computing applications a reality. An important application is mobile video system where real-time video streams will be transmitted through a mobile network, and being playback at mobile client machines. However, owing to the unique characteristics of the systems, the design of an efficient mobile multimedia system require the resolution of additional constraints due to the mobile network and mobile machines. For example, the asymmetric bandwidth of a mobile network poses challenges on the flow-control mechanisms for video transmission. Such limited bandwidth together with the high error rate of communication makes different feedback techniques inefficient. It is difficult to guarantee the playback quality in such a poor network environment, and the mobile network is likely to be the bottleneck resource of the system. The pre-buffering techniques for conventional distributed multimedia systems need further modifications before they can

be effectively applied in a mobile multimedia system since different mobile clients, i.e., PC notes and PDAs, may have different video buffer sizes. Furthermore, different types of mobile clients may request videos with great differences in sizes, resolution and workload characteristics. In such as mixed and dynamic environment, how to provide a fair service in serving the video requests from different clients is therefore an important consideration. In this paper, we design an efficient scheduling algorithm, called *Adaptive Buffer Sensitive* (ABS), for scheduling of the transmission of real-time video packets over a mobile network when the media server has to serve multiple video requests at the same time.

2 The Adaptive Buffer Sensitive (ABS)

The objectives of the ABS are to minimize the impact of transient workload on a video playback, and at the same time to minimize the impacts of a video request on other video requests. To achieve the second objective, we need to allocate a minimum bandwidth to each request. After providing minimum bandwidths to all the requests, the remaining bandwidth will be allocated in such a way that the request which playback situation is poorer will receive more bandwidth. The situation of a video playback can be defined based on its buffer playback duration of the client buffer. If the situation is so poor that even if allocating additional bandwidth to serve a request still cannot solve the problem, the server may start to drop frames at the server side by using a selective packet transmission scheme with attempt to increase the playback duration at the client buffer. Although this will affect the smoothness of the playback temporarily, it is hope that after the selective packet transmission, the playback status and buffer level can be restored to normal. The ABS consists of three steps: (1) admission control; (2) bandwidth allocation; and (3) selective packet transmission.

2.1 Admission Control

An important objective in admission control is to maximize the bandwidth utilization. Although various lossless admission control mechanisms have been proposed, they are not suitable to mobile multimedia systems since they usually define a tight control for admission such that the result bandwidth utilization will be low. Such a tight condition is highly undesirable in a

¹ The work described in this paper was partially supported by a grant from the Research Grants Council of Hong Kong SAR, China [Project No. CityU 1097/99E].

mobile environment since the network bandwidth is already limited. One simple way to defined the admission condition is:

$$\frac{\sum_{i=1}^n BW_Consume_i}{BW_Total} \leq 1 \quad (1),$$

such that each client should have sufficient bandwidth in average during a sufficient long period of time for playing its videos except when its buffer is emptied. However, equation (1) does not consider the problem of high error rate of a mobile network. Due to the errors in packet transmission, the actual sizes of packet transmitted can be much less than the bandwidth allocated for transmitting the packets. In order to include the problem of re-transmission, an error factor α is introduced to define the average percentage of bandwidth allocated for retransmission of lost and corrupted packets. Therefore, the available bandwidth for packet transmission excluding those error transmissions, $BW_Available$, may be defined as:

$$BW_Available = BW_Total \times (1 - \alpha) \quad (2)$$

Therefore, Equation (1) can be re-defined as:

$$\frac{\sum_{i=1}^n BW_Consume_i}{BW_Total \times (1 - \alpha)} \leq 1 \quad (3)$$

A new video request j will be accepted if adding its average workload, $BW_Consume_j$ to the summation workloads of all the existing streams in the system can still satisfy the above inequality. Otherwise, it will be rejected.

2.2 Bandwidth Allocation

Bandwidth allocation in the ABS is divided into two rounds. In the first round, the bandwidth allocated to serve a client is the minimum of the bandwidth that is equally divided from the available bandwidth, and the mean bandwidth consumption of its requested video such as:

$$BW_Allocated_i = \min\left(\frac{BW_Available}{n}, BW_Consume_i\right)$$

where $BW_Allocated_i$ is the bandwidth allocated to serve request i and n is the number concurrent requests. The rationale of such allocation scheme is that we want to provide a fair treatment to all the clients independent of the workload characteristics of their requested videos such that each client may receive up to $(BW_Available/n)$. If the average required bandwidth of a client is smaller than the mean value, $(BW_Available/n)$, we only provide its average bandwidth with the attempt to maintain the playback duration at the client buffer to the preferred level. Although the size of the next packet for transmission may be larger than the average workload of the video, allocating the average bandwidth to serve a client should not significantly affect the playback quality if the client buffer level is high enough. The buffer level should be able to be maintained at a preferred level in a sufficient long period of time when the client receives the average workload of its requested video. The problem of transient overloading on its performance will be solved by the bandwidth allocated in the second round.

Under such allocation scheme, the total bandwidth allocated at the first round, $BW_Allocated_{First_Round}$, will be:

$$BW_Allocated_{First_Round} = \sum_{i=1}^n \min\left(\frac{BW_Available}{n}, BW_Consume_i\right)$$

After first round allocation, the extra bandwidth, if any, will be allocated to the clients based on their buffer playback

duration, which is calculated as the playback time of the latest packet at the client buffer minus the current time. Obviously, the playback status is less affected by transient workload if its playback duration is longer. Therefore, in the allocation of the extra bandwidth, more bandwidth will be allocated to the client which playback duration is shorter such as:

$$BW_Allocated_Second_i = \frac{Buffer_Playback_Duration_i}{\sum_{i=1}^n Buffer_Playback_Duration_i} \times (1 - BW_Allocated_{First_Round})$$

It is hoped that by allocating more bandwidth to serve a client whose video buffer is low can restore its buffer level to the preferred buffer level.

2.3 Selective Packet Transmission

If the playback duration approaches to zero even though extra bandwidth has been allocated to it in the second round, the server may need to selectively send packets to the client for playback, i.e., some packets will be skipped. The detection of buffer empty at a client can be done simply by comparing the playback time of the last transmitted packet with the current time. If the playback time of the packet is smaller than the current time, the buffer is empty. When this situation occurs, allocating more bandwidth to serve the client may not be able to restore its playback status to normal in a short period of time. Therefore, the server has to invoke the selective packet transmission scheme. Under the scheme, each packet (or GOP) will be defined with a transmission deadline. For packet i of a stream, the transmission deadline, D_i is defined as:

$$D_i = \text{Playback time of packet } i - (S_i, BW_Available)$$

where S_i is the size of the packet

In selecting the next packet for transmission, all the packets, which have missed their transmission deadlines, will be skipped. The server looks at the packets queued in server buffer for the video. Starting from the first packet in the queue, the set of packets within a selection window will be examined and the packet with the lowest *resource consumption rate* will be selected. The size of the selection window is a pre-defined tuning parameter. Resource consumption rate of a packet is defined as: Resource consumption rate of packet i = size of packet $i \times$ (play rate of the video / no. of frames in the GOP)

Once a packet has been selected, the packets in front of the selected packet in the queue will be skipped. The selection for the next packet will start from the next packet after the selected packet up to the size of the selection window. By skipping packets, it is hoped that the buffer playback duration at the client buffer can be extended. The selective packet transmission scheme will stop when the buffer level is restored to the preferred buffer level. As astute reader may notice that if the selection window is larger, the number of packets skipped will also be more. Therefore, initially, it should use a small value and be increased when the playback duration cannot be increased after skipping the packets in each round.

3 Conclusion

In this paper, we design a scheduling algorithm called *Adaptive Buffer Sensitive* (ABS) for the scheduling of packets from multiple video streams to use the limited mobile bandwidth. Our objective is to minimize the impact of transient overloading on the playback quality of videos and at the same time to minimize the interference between the playback of different videos.