QoS Constrained Resource Allocation for Multimedia Wireless Networks

Harold Zheng, Sherry Wang, and John A. Copeland
Communications Systems Center, Georgia Institute of Technology, USA
250 14th Street NW, Room 586, Atlanta, Georgia 30332-0490, USA
{zheng, wang, copeland}@csc.gatech.edu

Abstract

To deliver high bandwidth demanding video traffic over the resource limited wireless systems is becoming a challenge issue. To control Quality of Service (QoS) of the delivery is even more difficult. In this paper, a QoS constrained multilevel resource allocation scheme (QML-RA) is proposed to tackle the problem of how to flexibly allocate wireless network resources for carrying video traffic with certain QoS guarantees. This scheme utilizes the layer property of MPEG frames and considers the QoS requirements of different data sections within a video stream. It also takes the advantage of the channel allocation flexibility provided by CDMA/TDD systems and statistically assigns bandwidth to video traffic according to their QoS requirements. Moreover, an optimal smoothing technique is used to preprocess the video traffic so that bit-rate variability could be reduced, which enhances performances of QML-RA. The scheme significantly improves the channel utilization and increases the network throughput.

1. Introduction

Networked real-time multimedia applications in a wireless environment, as suggested by IMT2000, bring many new challenges to network service providers [1-3]. The goal of this research is to tackle a major problem in providing video services over a wireless network. The problem is how to effectively manage limited bandwidth resources for providing real-time video service with certain deterministic guarantees.

It is known that one of the difficulties for video transmission is that an MPEG compressed video stream has multiple Quality of Service (QoS) requirements, such as variable data rate, stringent delay bound, delay variation, and tolerable error rate. Usually, an MPEG video stream conveys three types of pictures [4]: an intra coded frame (I-frame), a predictive coded frame (P-frame) and a Bi-directionally predictive coded frame (B-frame). The combination of the frame types and the corresponding compression algorithms results in a substantial variability in bit rates. In general, I-frames have the highest bit rate and B-frames have the lowest. The peak data rate can be many folds higher than the average rate. In addition, another type of data, the header, plays a vital role in decoding a video stream. A header holds the control information, such as the picture size, the image data start position, and the Group-of-Picture (GOP) structure, etc. Without the control information, it is impossible to decode a video stream properly.

On the other hand, the existing wireless network needs to be enhanced since it is inflexible to deal with various requirements among data streams. For instance, the conventional resource allocation schemes were designed for voice or low bit-rate data traffic. This type of service usually transmits data using fixed bandwidth channels. This allocation scheme is not adequate for video delivery over a wireless environment because of the nature of variable video data rates. The overall bandwidth in a wireless environment is also limited comparing with the wired network. So, an efficient packet-oriented data delivery structure with QoS guarantee should be considered. Some attentions have been focused on channel borrowing to accommodate different traffic load [5-7]. These methods used channel borrowing among up/down links, among channels, and neighboring cells, which can handle the hot point situation where the demand for connection is high. In [5], the resource bunch concept was proposed, which was used in conjunction with the centralized radio resource management for sharing resources more efficiently between cells. Four kinds of channel selection strategies were evaluated as compared with the Fixed Channel Resource Allocation (FC-RA) to demonstrate the capacity improvement. In [6], the necessity of a new system design for multimedia applications was pointed out. It proposed to use different numbers of time slot in a frame to accommodate asymmetric traffic by the up/down link slot borrowing, which maximized the frequency utilization. We call this scheme
as Uplink/Downlink Resource Allocation scheme (UD-RA). In [7], a burst-mode packet access scheme was proposed. In this scheme, high data transmission rates are assigned to mobiles on either the forward or the reverse channel independently for short burst durations. The assignment was based on load and interference measurements.

In this paper, we adopt the “differentiated service” concept to treat different classes of data differently. The proposed scheme is called QoS constrained multilevel resource allocation scheme (QML-RA). This scheme tries to tackle the problem of how to flexibly allocate limited wireless network resource to high bandwidth demanded video traffic with certain QoS guarantees. This scheme takes advantage of the channel allocation flexibility provided by the CDMA/TDD system architecture and statistically assigns available bandwidth to senders according to their traffic loads and QoS priorities. It fully exploits the mobile’s primary channel capacity as well as additional accessible bandwidth, such as the uplink, the downlink, and other users’ under-utilized channels. This scheme also utilizes the layer property of a MPEG video stream to spend more network resources on important frames, such as headers and I frames, instead of transmitting a video stream as a single type of data. In this way, an overall perceptually good video quality is obtained as well as bandwidth efficiency is achieved.

In addition, we modified a video-smoothing algorithm [8] for preprocessing the input Variable Bit Rate (VBR) video traffic. This technique dramatically reduces bit-rate variability so as to the bandwidth requirement, which further improves the effectiveness of the QML-RA. So, it becomes feasible to provide video service over a resource-limited network. The system configuration of this proposal is based on Wideband Packet-CDMA (WP-CDMA) [9] proposed by Ericsson.

The rest of the paper is organized as follows. In Section 2, the frame structure and the slot arrangement of a WP-CDMA system are briefly introduced. After that, three data classes are defined. These definitions reflect the differentiation of QoS requirements and handling methods for MPEG video data traffic. Then the QML-RA is proposed and explained in detail. In Section 3, the computer simulation results are presented. Finally, Section 4 concludes our study.

2. QoS Constrained Multilevel Resource Allocation scheme (QML-RA)

2.1 Wideband packet CDMA system (WP-CDMA)

The physical foundation of the QML-RA is based on CDMA technology. CDMA uses the whole spectrum to deliver information, and the mobile unit can detect its own signal as long as it uses the right code words at the right time duration. There are several proposed CDMA system configurations for the third generation wireless networks. One of them is Wideband Packet-CDMA system proposed by Ericsson for IMT-2000 Radio Transmission Technology (RTT) [9]. Figure 1 is the frame structure of WP-CDMA. Data are transmitted via uplinks and downlinks using frames. Each frame has sixteen slots. A slot is the basic time unit for Time Division Duplex (TDD). During a frame time, sixteen slots can be assigned to either an uplink or a downlink depending on needs, but at least one slot is reserved for uplink direction and one slot for downlink direction. The direction of each slot can be randomly decided to go in either uplink or downlink. The Figure 2 shows three patterns of TDD arrangement. Each slot time can have multiple, up to eight, code words. Each code word is called a burst. A burst is the smallest resource element.

![Figure 1: WP-CDMA physical channel frame structure](image)

<table>
<thead>
<tr>
<th>Pilot</th>
<th>TPC</th>
<th>RI</th>
<th>Data</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

0.625 ms

![Figure 2](image)

(a) Symmetric cases; (b) Asymmetric case
The combination of CDMA and TDD provides great flexibility for resource allocation. One of them is that it can deliver unbalanced traffic easily. It also gives the possibility of adapting to VBR video traffic.

2.2 Data classes definition

The traffic used in this research is the MPEG-1 compressed video streams that are of the layered properties. In a layered video application, if lower layers have minor losses, they will not adversely affect the perceived video quality. Instead, timing is a crucial requirement for comfortable viewing. Therefore, the QML-RA adopts the concept of "differentiated services" so that ongoing video traffic can be categorized into different data classes with correspondent QoS levels. These data classes are

- Class1 data - "No loss data": It demands no loss in transmission. The important data, such as headers and frame control data belong to this class.
- Class2 data - "Assured data": It requires a timing guarantee with tolerance of minor data loss, which is appropriate for both I- and P- types of frames.
- Class3 data - "best effort data": It demands a timing guarantee transmission with high tolerance of data loss, which is suitable to B frames.

The loss tolerance is relaxing in the descending order while the delay bound is still restrict. Therefore, instead of treating an overall video stream as a single type of data, resource allocation schemes spend more network resources on the more important components within a MPEG stream. In this way, an overall perceptually good video quality is obtained as well as bandwidth efficiency is achieved.

2.3 Proposed QML-RA

Although the advantages of CDMA/TDD have been considered in many resource allocation designs, currently they are only limited to low bit-rate data transmission or voice data service. In a conventional system, the voice channel bandwidth usually is designed to be small. This design can maximize the number of concurrent users. However, it is hard to accommodate variable bit-rate data service. On the other hand, the new IMT-2000 system tries to achieve 2Mbps per channel for high bandwidth traffics, but it may be overkill for low bit rate traffic such as low bit-rate videos. Therefore, a dynamic resource allocation scheme is necessary to achieve high utilization of the available bandwidth for both high and low bit-rate applications.

The proposed QML-RA scheme defines three levels of resource that can be shared among traffic loads:

1. **Primary-channel level.** Each mobile owns a primary uplink channel and a downlink channel. The primary channel duration is one slot time. The minimum channel capacity is one burst during a slot time. If the mobile or the base station has traffic load more than one burst, the sender has the highest priority to allocate other bursts during its primary channel slot time. If the sender cannot consume all the bursts in its primary channel duration, the unused bursts can be shared by other traffic that needs extra bandwidth during the following other two levels of resource allocations. The purpose of the primary channel level allocation is to provide a guaranteed amount of bandwidth for the video traffic. This level of resource is reservation based. It is necessary to arrange it carefully to avoid over subscription, otherwise, bandwidth waste may occur. Since it is a guaranteed resource, this level of resource can be suitable to Class1 data.

2. **Intra-channel level.** This is the most frequently used level of resource. At this level, high bandwidth demanding traffic can use the unoccupied slots after the primary channel level assignment. The traffic borrows a number of unused slots and bursts in these slots. All or partial of bursts in these slots can be borrowed. The condition of using this level of resource is that the borrowed slots must be in the same frame that has the borrower’s primary channel slot. The borrowed slots could be set in the same direction as the primary slot as long as they satisfy the condition that at least one slot is reserved for the uplink and one slot is reserved for the downlink in a frame. QML-RA will use this level of resource to accommodate the asymmetric traffic load between the uplink and the downlink since each side can allocate slots and bursts proportionally to its traffic loads. The purpose of this level of resource allocation is to take advantage of TDD by borrowing unused slots to support high bandwidth traffic. The purpose of allocating slots within a frame is for reducing the communication overheads. Traffic resource will be allocated more aggregated in limited number of frames. This level of resource is suitable to Class2 data.

3. **Inter-channel level.** This is the largest scale of borrowing bandwidth for high bandwidth demanded traffic. It allocates unused resources from other channels. The channel here means slots and bursts in different physical channels within one cell. The unused resources are the slots and bursts that are not used at the primary-channel level and the intra-channel level. This level of borrowing will fully utilize all available bandwidth within a cell to support traffic demands. Receivers can decode the incoming traffic as long as they use the right codes and time slots.
Three resource levels are integrated into a resource bundle. The resource bundle is then controlled by a dynamic control mechanism resided in a base station that has enough power and computation capacity. This resource bundle has the QoS priority for performing resource distribution. A sender gives a base station bandwidth requirements and the base station sends back the setting parameters such as burst ID and slot ID to mobile via Dedicated Control Channel (DCCH) [9].

In QML-RA, the different levels of resource can be used to handle different classes of data. Since the bandwidth-borrowing scheme is a statistical multiplexing process, the success of getting resource is based on the competition with other traffic. To achieve QoS guarantees for different types of data, we prioritized the data based on their classes and designated bandwidth to them in the sequence from Primary-Channel level to Intra-channel level and eventually to Inter-channel level. According to the data class hierarchy, the Class1 data have the highest priority to occupy channels, then Class2 data, and finally Class3 data. If multiple connections have the same class of data, then they will be given the same priority for resource allocation. QML-RA can guarantee that the most important data would not be lost. The pseudo code of channel assignment scheme of ML-RA is showed in Figure 3.

2.4 Performance analysis

**Theorem:** The data loss of QML-AR is less than that of UD-RA and FC-RA.

**Proof:** Let: $W$ = total bandwidth, $B$ = total bit-rate, and:

$$W = \sum_{i=1}^{N} w_i, \quad B = \sum_{i=1}^{N} b_i$$

Where, $w_i$ is the $i^{th}$ available bandwidth corresponding to the $i^{th}$ frame’s bit-rate $b_i$. Then: Data Loss (DL) for different schemes will be:

$$DL_{FC-RA} = \sum_{i=1}^{N} \max (b_i - w_i, 0)$$

$$DL_{QML-RA} = \max \left[ \left( \sum_{i=1}^{N} b_i - \sum_{i=1}^{N} w_i \right), 0 \right]$$

Assume that within a video trace, there are $M$ instances where $b_i > w_i$. Then,

$$DL_{FC-RA} = \sum_{i=M}^{N} \max (b_i - w_i, 0) + \sum_{i=(N-M)}^{N} \max \left( b_i - w_i, 0 \right)$$

$$= \sum_{i=M}^{N} (b_i - w_i)$$

$$DL_{QML-RA} = \max \left[ \left( \sum_{i=M}^{N} b_i - \sum_{i=(N-M)}^{N} w_i \right), 0 \right]$$

Since $b_i < w_i$ in the $(N-M)$ items, the second addend is less than zero, so $DL_{QML-RA} < DL_{FC-RA}$. Using the same notations and token, it can be proved that $DL_{QML-RA} < DL_{UD-RA}$.

Because of the paper length limitation, the details of this part are omitted here. Since the QML-RA assigned the resource level based on the traffic component priorities, the loss rates for three data classes will be different. The Class1 data loss will be minimum. The Class3 data loss is the biggest. By this way, the overall video quality will be improved.

3. Computer simulation results
3.1 Preprocess data

One of the challenges for video delivery over wireless channel is how to reduce bandwidth requirement of video traffic. Smoothing techniques can greatly reduce the burstiness of a VBR video stream so as to the bandwidth requirement. In [10], the pros and cons of several smoothing techniques have been evaluated. The general method of smoothing stored video streams is that by initiating a transmission earlier, the sender can send high bit-rate frames at a lower rate. The earlier arrived frames are stored in a fixed-size receiver buffer. The receiver can then retrieve, decode, and display frames at the stream frame rate. In this research, we adopt an *optimal smoothing* technique [8] by adding data class identifiers. This method is directly addresses the problem of reducing bit-rate variability for reducing bandwidth requirement. This optimal smoothing algorithm uses *work-ahead* principle so that video data can be sent ahead of schedule with respect to its playback time. The algorithm takes a list of the video frame size, the client buffer size and the tolerable initial delay as inputs. It off-line computes an optimal server transmission schedule that minimizes both the variance and peak rate at which data is send to the client. It does not introduce any delay. The result of smoothing is a list of piecewise CBR frame size with data-class identifier.

Considering the limited memory size possessed by a mobile receiving device, throughout this research, we only used 64k bytes buffer and 480ms initial delay for smoothing. Even with this small size of buffer, the effect of bandwidth reduction is impressive. The bandwidth allocation was conduct at the each bandwidth change point of the bandwidth profile.

3.2 Computer simulation configuration

The experiment system configuration is based on Wideband Packet-CDMA (WP-CDMA) proposed by Ericsson. In this research, we focus on the communications within a single cell. The physical settings are as the follows: slot duration is 10 ms, and the payload of a burst is 160 bits. Each channel has a maximum of 2 Mbps capacity. The number of channel is 15. The video traces are MPEG compressed data with average bit rate of around 300 kbps.

3.3 Simulation results

The simulations compared the performances of the proposed QML-RA scheme with an up/down link borrowing resource allocation (UD-RA) scheme as well as a Fixed Channel scheme (FC-RA). The FC-RA only shares resource in a symmetric pattern that at most half slots in a frame can be allocated for the downlink. Multiple MPEG video traces were used to explore the system capacity. The simulation results have shown the performances improvement. In Figure 4, we explored the system utilization. When more video traces feed into the system, the system bandwidth resource utilization increases. The utilization grows linearly at beginning, and then it starts to bend. This means that not all data traffic have been carried through the system. Some of the data are lost. We then test the block rate that is defined as the numbers of severely lost videos vs. the numbers of input videos. We defined the connection is blocked if 2% of data loss occurred. The Figure 5 shows that QML-RA and QML-RA with smoothing are significantly outperformed the less shared FC-RA and UD-RA schemes. QML-RA supports more video connections with the same block rate. Figure 6, 7 and 8 are the data-loss rate for Class2, Class3 data and the aggregated traffic. Once again, QML-RA shows better performance.

4. Conclusions

In this paper, we respond to the challenges of video delivery over a wireless environment and discuss the problems of the conventional channel allocation schemes for high bandwidth demanded multimedia applications. To overcome the drawbacks of the existing schemes and to meet new service requirements, a QoS constrained Multi-level Resource Allocation scheme was proposed. This scheme reserves certain amount of bandwidth for supporting critical data (such as Class1 data) and shares resource with other connections. Therefore, it keeps not only the QoS guarantees of circuit-oriented network but also the high efficiency of the packet-oriented resource allocation style. To satisfy the QoS requirements, three data classes for categorizing MPEG traffic and three resource levels for bandwidth sharing were defined. Though a primary channel for each mobile connection is kept, the overall resource allocation is performed over different resource levels. The preprocessed traffic also improved the system performance. The experiment results have shown that smoothed video traffic can increase channel utilization as well as reduce loss rates. It is worthwhile to point out that using QML-RA in conjunction with CAC can provide QoS guarantees to deliver video traffic in the wireless environment.

5. References

vice provider’s Perspective”, IEEE Personal Communications, August 1997, pp. 8-13


Figure 4. The system bandwidth utilization vs. the video traces

Figure 5. The video connection block rate vs. the number of video traces

Figure 6. The Class 2 data loss rate vs. the number of video traces

Figure 7. The Class 3 data loss rate vs. the number of video traces
Figure 8. The aggregate data loss rate vs. the number of video traces