Multiplexing Live Video Streams & Voice with Data over a High Capacity Packet Switched Wireless Network

Spyros Psychis, Polychronis Koutsakis and Michael Paterakis

{psycho, polk, pateraki}@ tsinet.gr

Electronic & Computer Engineering Department
&
Telecommunication Systems Institute
Technical University of Crete
73100, Chania, Greece

Abstract **: In this work we propose and evaluate a multiple access control mechanism for the multiplexing and the integrated delivery of Voice, Video and Data Traffic over a Wireless Cellular High Speed Packet Switched Network. We focus on the uplink wireless channel and we investigate the system's performance under a variety of possible loads, which consist of Voice, actual MPEG-4 videoconference streams, and Data Traffic. Our scheme achieves high aggregate channel throughput while preserving the Quality of Service (QoS) requirements of each traffic type.**

1. Introduction.

A Medium Access Control (MAC) Protocol that efficiently integrates voice (Constant Bit Rate, CBR, On/Off Traffic), video (Variable Bit Rate, VBR) and bursty data traffic in high capacity micro cellular environments is presented in this paper. The delivery of high quality live video material is a service with very strict QoS requirements. Consequently the design of MAC strategies that integrate this class of service with others, which often require contradictory QoS guaranties, is needed. Within the microcell, spatially dispersed mobile terminals (MTs) share a radio channel that connects them to a fixed base station or a wireless hotspot. The base station allocates the channel resources, delivers scheduling and feedback information and serves as an interface to the mobile switching center (MSC), which provides access to the fixed network infrastructure and the Internet.

Since the base station is the sole transmitter on the downlink channel, it is in complete control of the downstream traffic, using a MAC protocol to relay information to the users. Thus, we focus on the uplink (mobiles to base station) channel, where a MAC scheme is required in order to resolve the source terminals contention for channel access. The contribution of this work, beyond the design of an efficient MAC scheme, extends to the base station scheduling scheme. Both of these schemes contribute to the achievement of the QoS required by the various traffic streams.

2. Channel Structure, Actions of Terminals and Base Station Scheduling.

The uplink channel time is divided into time frames of equal length. The frame duration is selected such that an active voice terminal (i.e. a terminal in talkspurt) generates exactly one packet per frame. Each frame consists of two types of intervals. These are the request intervals and the information intervals.

Within an information interval, each slot accommodates exactly one, fixed length, packet that contains voice, video or data information and a header.

The voice and data terminals use the request intervals in order to transmit their requests to the BS. Voice terminals are given priority to transmit their requests to the BS. When all the Voice Requests have been transmitted, the Data Request transmission follows. The concept of reserving a minimum bandwidth for voice terminals to make channel reservations helps to keep the voice access delay within relatively low limits and gives clearly better performance than the PRMA [3] and quite a few PRMA-like algorithms, such as DPRMA [4], where the absence of request slots leads to a continuously decreasing probability of finding available information slots as traffic increases, and hence to greater access delays.

The request intervals consist of slots, which are subdivided into mini-slots, and each mini-slot accommodates exactly one, fixed length, request packet. By using more than one minislot per request slot, a more efficient usage of the available request bandwidth is possible. Since we assume that all of the voice source state transitions occur at the frame boundaries, we place all request intervals at the beginning of the frame, in order to minimize the voice packet access delay.

We adopt the idea that the request slots can be shared by voice and data terminals, (first by voice terminals and, after the end of voice contention, by data terminals), in order to optimize the use of the request bandwidth. Furthermore, we assume that the number of request slots per channel frame is variable, depending on the number of the video terminals. [2]

Figure 1 gives an example of the channel frame structure showing the request and information slots within a frame. Notice that in this example each request slot consists of two minislots.

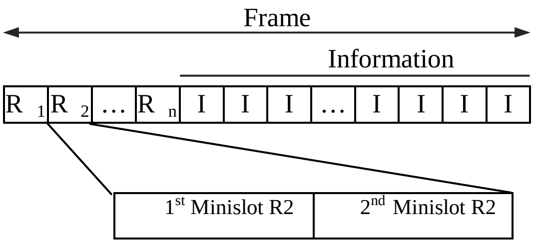

Figure 1: Frame Structure

The video terminals envoy their slot requests to the base station by transmitting them within the header of the first packet of their current video frame. This is possible due to the nature of the video sources, which we assume here that they transmit low activity video content with a few and small changes in bit rate overtime. Thus there is no need to waste channel slots in order to facilitate the transmission of video request packets to the BS.

The BS allocates channel resources at the end of the corresponding request interval, and follows a different allocation policy for video terminals than that for voice terminals.

Video terminals have the highest priority in acquiring the slots they demand. If a full allocation is possible, the BS then proceeds to allocate any still available information slots to the requesting voice terminals. Otherwise, if a full allocation is not possible, the BS grants to the video users as many of the slots they requested as possible (i.e., the BS makes a partial allocation). The BS keeps a record of any partial allocations so that the remaining requests can be accommodated whenever the necessary channel resources become available. In either allocation type case, the BS allocates the earliest available information slots to the video terminals, which, if needed, keep these slots in the following channel frames, until the next video frame (VF) arrives.

Voice terminals, which have successfully transmitted their request packets, do not acquire all the available (after the servicing of video terminals) information slots in the frame. If this happened, voice terminals would keep their dedicated slots for the whole duration of their talkspurt, and thus video terminals, would not find enough slots to transmit in, and this could lead to a violation of the video QoS requirements.

Consequently, in our scheme the BS allocates a slot to each requesting voice terminal with a probability p*. When there are no video terminals in the system, p* is set equal to 1.

Each data user can reserve only one information slot per channel frame and can keep it just as long as it has data packets to transmit.

We do not use preemption of data reservations but still voice users are given priority both in slots and in allocation policy.

3. Voice Source Model.

Our primary voice traffic model assumptions are the following:

a. Voice terminals are equipped with a voice activity detector [3]. Voice sources follow an alternating pattern of talkspurts and silence periods (on and off), and the output of the voice activity detector is modeled by a two-state discrete time Markov chain.

b. All of the voice source transitions (e.g., talk to silence) occur at the frame boundaries. This assumption is reasonably accurate, taking into consideration that the duration of a frame is equal to 12 ms here (as it it will be explained in section 6), while the average duration of the talkspurt and silence periods exceeds 1 sec. [13]

c. The voice delay limit is equal to 40 ms. [4]

4. Data Traffic.

We adopt the data traffic model based on statistics collected on email usage from the Finish University and Research Network (FUNET) [10]. The model simulates email data traffic. The probability distribution function f(x) for the length of the data messages was found to be well approximated by the Cauchy (0.8,1) distribution. The message inter-arrival time distribution for the FUNET model is exponential. In our work, we assume an upper limit on the average data message delay equal to 2 seconds, which is a tolerable delay for email message transmission.

5. MPEG-4 Streams.

MPEG-4 is an ISO standard that provides a standardized and worldwide accepted technological framework that enables the integration of the production, storage distribution and content access paradigms. MPEG-4 leverages existing digital video content by supporting the MPEG-1 and MPEG-2 coding standards. Furthermore, it enables richer development of new digital video applications and services. An MPEG-4 scene is composed by a number of audiovisual objects, which can be either static or time varying in nature. [8]

MPEG-4 Visual provides a natural video coding algorithm that is capable of operating from 5 kbps with a spatial resolution of QCIF (144x176 pixels). It is ITU-T H.263 compatible, in the sense that an MPEG-4 video decoder also correctly decodes an H.263 bit stream. The algorithm scales up to bit rates of some Mbit/s and optimization has been carried for ITU-R 601 resolution pictures (288x720@50Hz and 240x720@59.94 Hz). The recently developed Studio Profile can operate at over 1 Gbit/s. In addition MPEG-4 has a so-called Fine Granularity Scalability mode that allows transmission of the same video content at different bit rates from one coded version of the movie [9].

In our study, we use the trace statistics of actual MPEG-4 streams from [1]. The video streams have been extracted and analyzed from a camera showing the events happening within an office. We have used the high quality version of the movie, which has a mean bit rate of 400 Kbps, a peak rate of 2 Mbps, and a standard deviation of the bit rate equal to 434 Kbps.

All the streams are encoded at 25 video frames per second, which means that a new video frame is generated every 40 msecs. In our study we have made the assumption that the maximum transmission delay for video packets is 40 msecs, with packets being dropped when this deadline is reached. That is, all video packets of a VF must be delivered before the next VF arrives. The allowed video packet dropping probability is set to 0.0001 [4].

6. System Parameters.

The channel rate is 20Mbps (from [5]). The frame duration is chosen to be equal to the time a voice terminal needs to generate a new voice packet [6]. Assuming that the speech codec rate is 32 Kbps and that the packet length is equal to the size of an ATM cell, yields the frame duration of 12 ms. The 12 ms of frame duration accommodate 566 slots. Consequently our channel's information payload rate is slightly higher than 18Mbps since we encapsulate the ATM headers and the reservation slots. Due to the high channel rate of 20 Mbps which leads to a slot duration of only 0.021 msecs, it would be impossible to accommodate more than 2 minislots in each request slot because we need to allow for guard time and synchronization overheads, for the transmission of a generic request packet, and for the propagation delay within the picocell.

The number of the request slots in our system varies from 5 to 30 (i.e. from 10 to 60 request minislots). As we have already mentioned the actual number of request slots depends on the number of video sources admitted into the system (see Table 1). Consequently the bandwidth reserved for the request intervals varies from 0.88% to 5.3 % of the total available bandwidth. The number of the R-slots in Table 1, was specified through simulation.

We adopt the two-cell stack reservation random access algorithm for use by the voice terminals when transmitting their request packets, due to its operational simplicity, stability and relatively high throughput [7].

The two-cell stack blocked access collision resolution algorithm is adopted for transmitting the request packets of the data users [11].

7. Simulation Results and Discussion.

An extensive simulation study was carried out. Each run simulated one hour of actual network activity (300005 channel frames). All the results shown correspond to averages of 10 independent runs (Monte Carlo method). The lack of similar work in the literature (i.e., proposed mechanisms for handling actual Video Streams of this quality, together with performance evaluation of such mechanisms) prevents us from making result comparisons (the only similar work, to the best of our knowledge, is presented in [12], however it considered a 9 Mbps Wireless Channel).

Initially we simulated the system under all possible movie loads from 0 to 22 movies with no active data terminals. These runs helped us to specify the boundaries where the number of the request slots must change in order to accommodate the generated voice and data request packets. Additionally the value of the probability p* was chosen equal to 0.1 (10%), since other values of p* have also been tried out through simulation, and it has been found that the chosen value gives very satisfactory results for all the examined cases of video load. Table 1 shows the results of these runs.

VIDEO	VOICE	R-SLOTS	AVERAGE THROUGHPUT (%)
0	1258	30	94.57
1	1092	25	84.30
2	1062	25	84.25
3	1003	25	82.03
4	896	25	76.19
5	870	25	76.45
6	847	25	76.93
7	763	20	72.82
8	734	20	72.85
9	712	20	73.40
10	591	20	66.51
11	560	20	66.39
12	514	20	65.14
13	441	12	61.86
14	412	12	61.89
15	363	12	60.42
16	296	5	57.59
17	265	5	57.47
18	212	5	55.69
19	130	5	51.73
20	98	5	51.54
21	70	5	51.64
22	0	5	48.59

Table 1

A careful reviewer of the results shown above will notice that the reduction of the Voice Terminals capacity as the number of Video Terminals increases does not appear in a completely linear manner. This is also shown in Figure 2 where the results are graphically presented. Each different shadow on the bars corresponds to a different value of R-slots

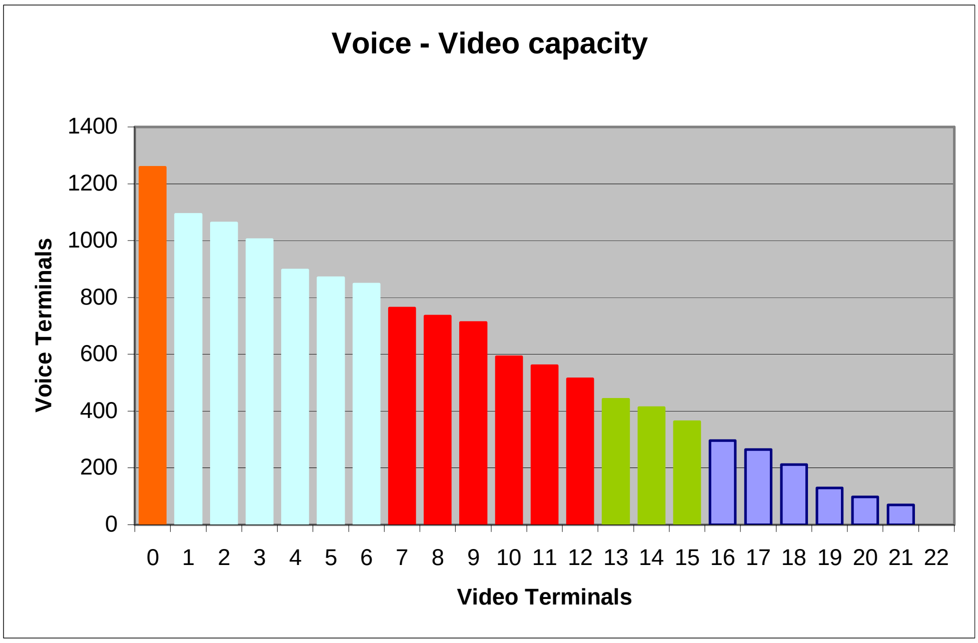

Figure 2: Voice – Video Capacity

This non-linear reduction appears due to the relatively few changes in the values of R. The reason we do not change the number or R slots each time a new Video Terminal enters the system, is that such an approach would increase the system complexity since the Base Station would have to change the channel frame structure quite often. Therefore, we assume that the Base Station has a predefined small set of operating modes, which derive from the number of R slots reserved. As shown in Table 1 and as expected from the bursty nature of the video traffic and the very strict packet video dropping probability, the average channel throughput decreases dramatically as the number of video sources increases. Still, our scheme is shown to achieve high throughput (over 60%) for most cases of video traffic load.

During the second set of experiments we simulated the system under Voice and Data traffic only. Since there is no Video in the system the value of R was set to 30 slots. We tried to accommodate into the system as many voice terminals as possible given that the data throughput should match the data load value. Table 2 shows the difference between the theoretical calculation of the Channel Voice capacity and the actual one. The theoretical calculation of the voice capacity was made with the equation:

[(total number of information slots-estimated data throughput) / probability that a voice terminal is in talkspurt].

We observe that the actual Voice Capacity is smaller than the one predicted theoretically. Our simulations also show that the data mean message delay is significantly lower than the assumed upper bound of 2 secs. The average data message delay does not exceed 1 sec, except in the case ë=0.25 where the number of voice terminals is quite high.

λ	Voice Terms Capacity (Theoretical)	Voice Terminals Capacity	Difference
0,25	1212	1212	0
0,5	1165	1135	-30
0,75	1118	1078	-40
1	1071	1030	-41
1,25	1024	985	-39
1,5	978	938	-40
1,75	930	890	-40
2	883	845	-38

Table 2

During the third set of experiments we simulated the system under certain mixtures of Voice, Video and Data traffic. Table 3 shows some representative results.

R-Slots	λ	Video Terms	Voice Terms	Average Throughput (%)
5	0,5	20	0	51,24
12	0,5	15	288	61,85
12	1,5	15	94	61,40
20	0,5	10	496	66,44
20	1,5	10	298	65,69
25	0,5	1	971	76,75
25	1,5	1	771	76,30
25	0,5	5	780	82,27
25	1,5	5	586	81,37
30	1,5	0	938	91,72

Table 3

We notice that, for the same number of Video Terminals, when ë increases by 1 (e.g. from 0.5 to 1.5, with 15 active Video Terminals), the system's voice capacity decreases by approximately 194 voice terminals. The reason for this is that a voice terminal is in talkspurt with probability 0.425 and each voice terminal in talkspurt requires one information slot per channel frame. Therefore, the voice capacity decrease of 194 terminals corresponds on average to approximately 83 slots, which is almost equal to the average increase of 80 slots needed by the data terminals because of the increase in the value of ë by one message per frame (i.e. on average 80 packets per frame). This means that our protocol achieves an almost stable channel throughput for a given number of video terminals, as shown in the last column of Table 3. The first and the last rows of the Table present the minimum (no voice, many bursty video sources are present) and the maximum (no video, many CBR voice sources are present) channel throughputs achieved by our system, respectively.

8. Conclusions.

In this work, we have proposed and investigated the performance of a mechanism for transmitting videoconference streams and voice with data over a high capacity wireless channel. We evaluated its performance through an extensive simulation study in which we used actual MPEG-4 stream traces together with data generated by a synthetic model fitted to measurements of email traffic from a National University and Research Network (FUNET). Our scheme, is shown to achieve high aggregate channel throughput, expressed in terms of the maximum number of video and voice terminals with the maximum data load sustained by the system, in many cases of traffic load, while preserving the Quality of Service (QoS) requirements of each traffic type.

REFERENCES

[1] http://peach.eas.asu.edu/index.html

[2] P. Koutsakis and M. Paterakis, *"On Multiple Traffic Type Integration over Wireless TDMA Channels with Adjustable Request Bandwidth"*, International Journal of Wireless Information Networks, Vol. 7, No.2, pp.55-68, 2000.

[3] S. Nanda, D. J. Goodman and U. Timor, *"Performance of PRMA: A packet voice protocol for cellular systems"*, IEEE Transactions on Vehicular Technology, Vol. 40, pp. 584-598, 1991.

[4] D.A. Dyson and Z.J. Haas, *"A Dynamic Packet Reservation Multiple Access Scheme for Wireless ATM"*, ACM / Baltzer MONET Journal, 4(2), pp. 87-89, Jan 1999.

[5] N.Passas D.Skyrianoglou and L.Merakos, *"Traffic Scheduling in Wireless ATM Networks"*, in Proceedings of the IEEE ATM'97 Workshop, Lisbon, Portugal, May 1997.

[6] P. Koutsakis, *"Medium Access Control and Traffic Policing for Multimedia Traffic Integrated Access in High Capacity Wireless Channels"*, PhD Dissertation ECE Dept, Technical University of Crete, Chania 2002.

[7] A. C. Cleary and M. Paterakis, *"On the Voice-Data Integration in Third Generation Wireless Access Communication Networks"*, European Transactions on Telecommunications and Related Technologies, Vol. 5, No. 1, Jan.-Feb. 1994, pp. 11-18.

[8] A. Basso, S. Varakliotis, R. Castagno, *"Transport of MPEG-4 over IP/RTP"*, in Proceedings of the PV Workshop, 2000.

[9] Leonardo Chiariglione, *"MPEG-4, why use it?"* Telecom Italia Lab–Italy, http://leonardo.telecomitalialab.com/paper/mpeg-4/

[10] Q. Pang, A. Bigloo, et al., *"Service Scheduling for General Packet Radio Service Classes"*, in Proc. of the IEEE Wireless Communications and Networking Conference 1999 (WCNC '99), New Orleans, USA.

[11] M. Paterakis and P. Papantoni – Kazakos, *"A Simple Window Random Access Algorithm With Advantageous Properties"*, IEEE Trans. on Inform. Theory, Vol. IT- 35, No. 5, September 1989, pp. 1124-1130.

[12] P. Koutsakis, S. Psychis and M. Paterakis, *"On the Integration of MPEG-4 Video Streams with Voice and E-mail Data Packet Traffic over Wireless Picocellular Networks"*, in Proceedings of the IEEE PIMRC 2001 Conference, San Diego, USA

[13] P.T. Brady, *"A technique for investigating on-off patterns of speech"*, Bell Sys Tech Journal, 1965.