OPTIMAL CROSS-LAYER SCHEDULING FOR VIDEO STREAMING OVER 1xEV-DO

Tanır Özçelebi[1], Fabio De Vito[1,3], Oğuz Sunay[1], Murat Tekalp[1,2], Reha Civanlar[1], Juan Carlos De Martin[4]
[1]*Koç University, College of Engineering, 34450, Istanbul, Turkey*
[2]*Department of Electrical and Computer Engineering, University of Rochester, Rochester, NY 14627*
[3]*Dipartimento di Automatica e Informatica /* [4]*IEIIT-CNR, Politecnico di Torino, 10129, Torino, Italy*
{tozcelebi,fdevito,osunay,mtekalp,rcivanlar}@ku.edu.tr, juancarlos.demartin@polito.it

ABSTRACT

In wireless video transmission, service fairness is a key point together with video quality and average channel throughput. The resource sharing method used has a major effect on the communication system performance and must utilize information from multiple layers of the OSI protocol stack for better user experience. The semantic and decodability (concealment related) importance of video packets, which is helpful in assigning priorities to these packets, can be considered at the application layer. In this paper, a multiple objective optimized (MOO) opportunistic multiple access design for time slot scheduling in a 1xEV-DO (IS-856) system, where a rate adaptive H.264 encoder is employed, is presented. In this framework, the user that experiences the best compromise between the least buffer fullness, the best channel throughput and the highest video packet importance is served. Hence, losses are forced to occur at the low importance packets. Experimental results show that this system outperforms the state-of-the-art frameworks, guaranteeing better PSNR for high importance regions in the order of 1 or 2 dB's with respect to the CBR case.

I. INTRODUCTION

The increasing bandwidth availability in wireless networks made it possible to distribute also multimedia content to mobile users along with classical applications. In this sense, CDMA networks are particularly useful in the case of video transmission, which is quite demanding in terms of bandwidth. This kind of service in mobile communications requires both speed and buffer capacity in handset devices, and the network resource sharing algorithm has to take into account the wide spectrum of receivers logged into the network, while providing fast access to information content. Quality-of-Service (QoS) is not guaranteed for such applications in most existing systems. Therefore, highly efficient systems that enable high-speed data delivery along with voice support over wireless packet networks are required. Also, there is need for adaptive and efficient system resource allocation methods specific to transmission of such information. Among these methods, opportunistic multiple access schemes [1] in which all system resources are allocated (scheduled) to only one user at a time are known to be optimal due to channel utilization (overall capacity).

In the 1xEV-DO (IS-856) standard [2], opportunistic multiple access is used and all transmission power is assigned to only one user at a time within time slots of length T_s (1.667 ms). The main target is to transmit high speed packetized data to multiple users on CDMA/HDR [3] systems. Adaptive coding and modulation are employed to support various service types (data rates) that can be properly received by a user at all times along the duration of a communication session. It is crucial to choose an appropriate resource (time) scheduling algorithm to achieve the best system performance. Application layer requirements and physical layer limitations need to be well determined, and the scheduler has to be designed accordingly. For example, e-mail and SMS services are tolerant to delay, and intolerant to data loss, while real time streaming applications can tolerate few losses. Hence, cross-layer design is mandatory for video transmission, in order for a scheduling algorithm to be optimal in both physical and application layer aspects.

The state of the art scheduling algorithms for the IS-856 (1xEV-DO) system are maximum C/I (carrier-to-interference ratio) [1], first in first out (FIFO), proportionally fair (PF) [4] and exponential schedulers [5]. All of these algorithms suffer from either service fairness or overall channel throughput.

Since we are interested in video streaming rather than a download-and-play solution, video packets that are delivered later than their playout times are discarded at receiver side and are considered lost; therefore, if the sender detects that the packet will arrive late at the receiver, it can discard (not transmit) the late information at the source side so lowering network congestions.

In video coding, the importance levels of video packets differ from each other due to variance of temporal semantic importance, and also inter and intra frame prediction. The overall user utility can be significantly increased using cross layer design and appropriate packet priority assignment and content (semantic relevance) analysis. In this paper, a novel cross-layer multi-objective optimized (MOO) scheduler for video streaming over 1xEV-DO system is presented. The overall channel throughput, individual buffer

This work is supported by 3DTV-NoE funded by the European Commission 6th Framework Information Society Technologies Programme, and by Centro Supercalcolo Piemonte (CSP) - Torino, Italy.

occupancy levels and contribution of the received network packets in terms of visual quality are simultaneously maximized.

This paper is organized as follows: The scheduling multi-objective optimization (MOO) formulation is outlined in Section 2. The method used for MOO solution is explained in Section 3. The experimental results with different settings are given in Section 4, and finally, the conclusions are drawn in Section 5.

II. PROBLEM FORMULATION

Video contents have not a uniform semantic importance; within a sequence, there are several runs of shots which can be of different interest for different users. It is possible to encode each region at a different bitrate, allowing low-importance frames to be coded at much less bitrate than important ones, thus saving bits for high-importance scenes. In this way, we can obtain better PSNR for semantically important regions. In order to generate this effect, each semantic region has to contain an integer number of GOPs. Therefore, GOP size needs to be flexible, and moreover the bitrate control should be able to change its target value for each GOP while encoding the sequence. Packets belonging to the same region have not the same decodability importance; usually, packets coming from I- and P-frames have higher impact on the decoded video if lost, given the possibility of error propagation by means of motion prediction, as described in [6]. The joint importance is computed as the product of the semantic importance level and the quantized decodability importance. Decodability importance is a real number and has to be quantized.

Transmission of video content over low bandwidth channels requires pre-fetching of data stream at the receiver side, so that distortion and pauses caused by buffer underflows or overflows in the duration of video playout can be avoided. This pre-roll (initial buffer) delay can not be excessive for any particular user due to buffer limitations and customer convenience. High visual quality, low pre-roll delay and continuous playout of the content are the most important requirements from a video streaming system, and appropriate scheduling algorithms are desirable.

Both physical layer feedback (C/I ratios) and application layer feedback (decoder buffer level) are used in the proposed framework. In the 1xEV-DO scheme, the back-channel is used to report the current C/I ratio experienced by mobile users, so that the transmitter is aware of the maximum rate that can be achieved for each user within a probability of error range. Channel statistics history is stored and used at the transmitting site for better performance. In our framework, the client buffer occupancy levels are also reported back to the base station.

Assume that there exist K users within the wireless network, demanding videos from the base station with a certain bitrate distribution, $R_V(t)$. Here t ($0 \le t \le \infty$) denotes the discrete time slot index. Our aim is to maximize the overall average channel throughput at each time slot, R(t), while guaranteeing fair and satisfying quality of service for each of these K users. Fairness can be provided by maximizing the buffer levels of individual candidates for scheduling at each time slot. If buffer underflows are inevitable, the video quality can still be protected by careful priority assignment to video packets according to per-packet decodability and semantic importance. In this way, since the video packets with high decodability and semantic importance are transmitted with priority, *packet losses are forced to occur at the less important parts.* Therefore, the group of objective functions to be optimized among users at time t is $\{B_i(t), R_i(t), imp_i(t)\}$, where $B_i(t)$ denotes the buffer fullness level, $R_i(t)$ represents the effective channel throughput, and $imp_i(t)$ is the per-packet importance for user i at time t. The average channel throughput up to time slot t can be calculated as below:

$$\overline{R}(t) = \frac{1}{t} \times \sum_{1 \le i \le k} \sum_{1 \le t' \le t} s_i(t') \cdot R_i(t') \qquad (1)$$

where $s_i(t)$ is a binary variable taking the value 1 if user i is scheduled at time slot number t, 0 otherwise. The buffer occupancy level of user i at time t is given by

$$B_i(t) = \max\{B_i(t-1) + T_s \times (s_i(t) \cdot R_i(t) - R_v(t)), 0\} \quad (2)$$

We can also calculate the channel throughput in a recursive manner in terms of previous value as given below:

$$\begin{aligned} \overline{R}(t) &= \frac{1}{t} \times \left((t-1) \times \overline{R}(t-1) + \sum_{1 \le i \le k} s_i(t) \cdot R_i(t) \right) \\ &= \frac{(t-1) \times \overline{R}(t-1)}{t} + \frac{1}{t} \cdot \sum_{1 \le i \le k} s_i(t) \cdot R_i(t) \end{aligned} \qquad (3)$$

For large values of t, the first term on the right hand side of the above equation becomes approximately equal to $\overline{R}$(t-1). Then, the throughput enhancement due to scheduling the i^{th} user at time slot t, $\Delta\overline{R}_i$ (t), is calculated as follows:

$$\overline{\Delta R_i}(t) = \overline{R}(t) - \overline{R}(t-1) \cong \frac{1}{t} \cdot R_i(t) \qquad (4)$$

Ideally, the server side must schedule the user that experiences the best compromise between the least buffer occupancy level, the best available throughput enhancement and the most important network packet to be delivered. Hence, our optimization formulation for choosing the user to schedule at time slot t is given by

$$\arg\max_i (\overline{\Delta R_i}(t)) = \arg\max_i \left\{ \frac{1}{t} \cdot R_i(t) \right\} \qquad (5)$$

$$\arg\min_i (B_i(t)) \qquad (6)$$

$$\arg\max_i (imp_i(t)) \qquad (7)$$

jointly subject to

$$B_i(t+1) \le BufferSize(i)$$

where *BufferSize(i)* denotes the available decoder buffer size at the i^{th} client. The last constraint is necessary to guarantee that a user whose buffer will overflow after a possible slot assignment is never scheduled. This constraint can indeed cause performance drops in terms of channel capacity especially in the case of maximum rate scheduler, since the user with the highest available rate can not be scheduled all the time.

It is not possible to suggest a direct relationship between the values of instantaneous buffer level and available channel rate for a specific user. In fact, a user's buffer level gives no obvious hint about the current channel condition and visa versa. Therefore, the exhaustive multi-objective optimization method given in Section 3 needs to be applied.

III. MULTI-OBJECTIVE OPTIMIZATION (MOO)

For the objective function set $F=\{f_1,f_2,...,f_N\}$, a solution s^* is called globally Pareto-optimal if any one of the objective function values cannot be improved without degrading other objective values. For single objective optimization problems, one can come up with one or more optimal solutions resulting in a *unique* optimal function value, and a Pareto-optimal solution is also a globally optimal solution. In contrast, this uniqueness of the optimal function value is not valid for multi-objective optimization (MOO) problems since two or more of the objective functions may be either conflicting or uncorrelated. Hence, there may exist many Pareto-optimal solutions and one has to discriminate between these solutions to determine which one is better and to come up with a best compromise solution. For this, one needs to determine the relative importance of objective functions. In case of equally important objectives, the individual objective values need to be rescaled to an appropriate range in order to compensate for range differences as follows:

$$f_{i,scaled} = \frac{f_i - f_{\min}}{f_{\max} - f_{\min}} \quad (8)$$

In our method, the throughput enhancement, the decoder buffer occupancy level and the per-packet importance are normalized to take real values between 0 and 1 as shown in Figure 1 for the two dimensional case, which is also described in [7].

In solution of such problems, an *infeasible* point that optimizes all of the objective functions *individually* is called the utopia point. The utopia point, *U(t)*, on the throughput-buffer-importance space is set as follows:

$$U(t) = \left(\overline{\Delta R}(t)_{\max}, B(t)_{\min}, imp_i(t)\right) \quad (9)$$

The best compromise optimal solution is found as the *feasible* point that is closest to the utopia point in the Euclidian-distance sense. A more detailed explanation of the multiple-objective optimization (MOO) techniques used in the literature can be found in [8]-[9].

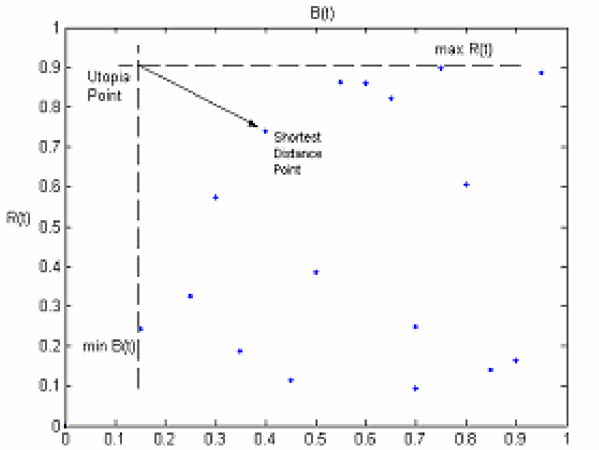

Figure 1. The proposed algorithm schedules the user whose corresponding point is closest to the utopia point.

IV. EXPERIMENTAL RESULTS

We encoded a 2250-frame test sequence (part of a soccer match), whose time duration is 90 seconds. Semantically more important regions are coded at a higher rate than low-importance GOPs, at bitrate ratios 1-to-2 and 1-to-3 as shown in Table 1 and Table 2, resulting in an average bitrate of 100 kbps in each case. The decodability importance has been quantized using two levels.

The resulting bitstreams are fed into the scheduler using the product of semantic and decodability importance indicators. Losses are introduced only by late packet delivery. Twelve users require the same video at random times within a time period of 5 seconds. Packets are ordered on a GOP-basis at the source side, according to their importance. In this way, we transmit important packets of each GOP first, and packets discarded due to late delivery will be concentrated in easily-concealed and low-importance regions. The maximum allowed initial buffering time is set to 5 seconds. Alternatively, a user stops pre-fetching after half of the decoder buffer is already full, where the decoder buffer size is set to 1 Mbits.

Results for overall PSNR obtained are shown in Table 1. In the 1-to-1 rate allocation case, the whole video is encoded and played at constant bitrate avoiding peaks in the rate distribution. Here, the network is highly loaded with 1200 kbps (12x100 kbps) peak rate almost all the time during transmission, causing excessive packet loss rate. On the other hand, in the 1-to-2 ratio case, the rate distribution is not uniform. Considering that the users are accessing the network at random times, some of the users will be draining data at 122 kbps, while others are streaming at 61 kbps, smoothing out the peak required transmission rate values. Hence, this unequal rate distribution is actually useful for reducing packet losses. In the 1-to-3 rate ratio case, the bitrates of the semantically important segments are themselves too high, resulting in packet losses higher than in the case of 1-to-2 ratio case.

The overall PSNR of the sequences decreases as the gap of semantic importance increases along segments. This is natural due to the effect of non-linear mapping between bitrate and PSNR. Here, the low importance parts have much lower PSNR than the 1-to-1 case, while high importance levels gain 1 or 2 dB's with respect to the same level. If we focus our attention only on the high importance parts, the PSNR increases with wider gap in the rate allocation of different regions.

V. DISCUSSION

In this paper, we proposed a novel cross-layer optimization technique for determining the best allocation of channel resources (time slots) across users over 1xEV-DO wireless channels. The novelty of this framework comes from the usage of decodability and semantic importance feedback from the application layer to the scheduler. The modifications to the H.264 codec have been described as well as the optimized scheduling algorithm. Network simulations show that noticeable improvements can be obtained with respect to the scheduler which does not consider packet importance, especially under strict requirements such as very short pre-roll delays. Experimental results show that, this approach ensures higher video PSNR with respect to constant bitrate coding. Furthermore, to better simulate the actual user behavior, we introduced random initial access times for users. As a result, received video PSNR was further improved.

VI. REFERENCES

[1] R. Knopp and P. A. Humblet, "Multiple accessing over frequency selective fading channels," in Proceedings of IEEE PIMRC 1995, Toronto, Canada, October 1995.

[2] TIA/EIA IS-856-2, Cdma2000 High Rate Packet Data Air Interface Specification, 2002.

[3] P. Bender, P. Black, M. Grob, R. Padovani, N. Sindhushayana, and A. Viterbi, "Cdma/hdr: A bandwidth effcient high-speed wireless data service for nomadic users," IEEE Communications Magazine, vol. 38, no. 7, pp. 70-77, July 2000.

[4] A. Jalali, R. Padovani, and R. Pankaj, "Data throughput of cdma-hdr: A high efficiency high data rate personal communications system," in IEEE 51st Vehicular Technology Conference, Tokyo, Japan, May 2000.

[5] S. Shakkottai and A. Stolyar, "Scheduling algorithms for a mixture of real-time and non-real-time data in HDR," Proc. 17th International Teletraffic Congress (ITC-17), Salvador de Bahia, Brazil, 2001.

[6] F. De Vito, D. Quaglia, J.C. De Martin, "Model-based Distortion Estimation for Perceptual Classification of Video Packets" , Proceedings of IEEE Int. Workshop on Multimedia Signal Processing (MMSP), vol. 1, pp.79-82, Siena, Italy, September 2004.

[7] T. Özçelebi, O. Sunay, A. M. Tekalp, M. R. Civanlar, "A Cross-Layer Optimized Minimum-Delay Scheduling Algorithm for Real Time Video Streaming to Multiple Users over 1xEV-DO," to appear in IEEE GlobeCom 2005.

[8] H. Papadimitriou, M. Yannakakis, "Multiobjective query optimization," Proceedings of Symposium on Principles of Database Systems (PODS), pp. 52-59, California, 2001.

[9] Y.-il Lim, P. Floquet, X. Joulia, "Multiobjective optimization considering economics and environmental impact," ECCE2, Montpellier, 5-7 Oct. 1999.

Table 1. Packet loss rates and overall PSNR for the test sequence, using 2 levels of semantic importance and 2 levels of decodability importance.

User	1 to 1 rate allocation		1 to 2 rate allocation		1 to 3 rate allocation	
	PLR (%)	PSNR (dB)	PLR (%)	PSNR (dB)	PLR (%)	PSNR (dB)
1	0.89	34.24	0	34.16	1.49	33.54
2	0.14	34.45	0	34.16	1.97	33.32
3	0.90	34.24	1.18	33.85	2.01	33.45
4	0	34.504	0.96	33.90	0.95	33.58
5	1.85	34.24	1.17	33.87	3.53	33.24
6	1.70	34.09	0.17	34.13	0.75	33.65
7	0.11	34.47	0.17	34.13	0.81	33.59
8	0.20	34.40	0.61	34.03	0.75	33.71
9	0.35	34.39	0	34.16	0.34	33.77
10	0.87	34.34	0	34.16	1.21	33.55
11	0.86	34.27	1.18	33.86	1.14	33.60
12	0	34.504	0.96	33.91	0.47	33.75

Table 2. Packet loss rates and PSNR for the high importance region of the test sequence, using 2 levels of semantic importance and 2 levels of decodability importance.

User	1 to 2 rate allocation		1 to 3 rate allocation	
	PLR (%)	PSNR	PLR (%)	PSNR
1	0	36.03	1.84	35.99
2	0	36.03	1.74	36.00
3	0.23	35.97	2.21	35.96
4	0	35.97	0.19	36.45
5	0.77	35.77	4.35	35.52
6	0.22	35.92	0.91	36.19
7	0.12	35.97	0.14	36.41
8	0.09	35.96	0.93	36.29
9	0	36.03	0.41	36.35
10	0	36.03	0.50	36.38
11	0.40	35.94	1.41	36.12
12	0	35.97	0.58	36.32