Multiple-Objective Cross-Layer Optimized Content-Adaptive Scheduling for Video Streaming over 1xEV-DO

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

Abstract

In video transmission over wireless cellular packet networks, service fairness, video quality and channel throughput should all be simultaneously guaranteed. In this regard, a key role is played by the scheduling algorithm, that - to achieve maximum performance- should consider both physical and application layer information. At the application layer, we propose to consider the video perceptual importance, which depends on both the time-varying semantic relevance and the individual importance of each packet. In this paper, a novel multiple objective optimized (MOO) opportunistic multiple access scheme for slot assignment in a 1xEV-DO system is presented; modifications to H.264 codec are also described. Here, the user that experiences the best compromise between the least buffer occupancy level, the best channel condition and the highest packet importance indicator is served at each time slot. Results show that this system outperforms the state-of-the-art techniques.

1. Introduction

In the last decade, users' interest in wireless networks has experienced a fast growth and research in this area has made significant progresses. The increasing bandwidth availability made it possible to distribute also multimedia content to mobile users along with classical applications like e-mail. 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 computational power and buffer capacity in handset devices and the network resource sharing has to take into account the wide spectrum of receivers logged into the network, while providing fast access to information content. Most practical systems do not guarantee Quality-of-Service (QoS) for such applications. Therefore, highly efficient systems that enable high-speed data delivery along with voice support over wireless packet networks are required and 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 in terms of 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 msec). 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 losses, while real time streaming applications can tolerate a few losses. Hence, cross-layer design is mandatory for video transmission, in order for a scheduling algorithm to be optimal in both physical layer 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]. The maximum C/I scheduler is also known as the maximum rate scheduler. Since none of these schemes are cross-layer designed, they are all suboptimal. The maximum rate scheduler gives access to the user with the best channel conditions, in order to maximize only the overall channel throughput. The main drawback of this approach is that, it does not provide fairness among users, since the bandwidth will always be assigned to the closest users to the base station (BS). On the other hand, the FIFO scheduler algorithm is designed to select the active user who experienced the longest delay in access to the network. Consequently, this approach does not optimize the channel throughput. The proportionally fair (PF) scheduler has been proposed in order to find a compromise among the two previous scenarios and selects the user with the best channel improvement. Each user's average available channel

0-7803-9305-8/05/$20.00 © 2005 IEEE.

bandwidth is tracked for a given time window and its present available channel throughput's ratio to its average over that time window is calculated at every time slot. The user with the highest such ratio is scheduled for the present time slot. An evolution of this approach is the *exponential scheduler* of [5] which adds some fairness in terms of service latency.

In packet wireless systems, packet losses are likely to occur and this results in serious reduction in received video quality unless appropriate error protection and/or error concealment techniques are employed. These packet losses are due to bit errors, congestion at intermediate routers or late delivery. 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, hence avoiding network congestion. Automatic Repeat reQuest (ARQ) techniques are not useful for our purposes since they might cause undesired pauses while streaming video information. It is more efficient to rely on some kind of a-posteriori error recovering (FEC), trying to better protect the portions of information we consider more important.

Multimedia streams, and in particular video data, have not uniform importance. In video coding, as a result of inter and intra frame prediction, the importance levels of video packets differ from each other. Also, if we consider the semantic importance levels of different temporal segments, packets that originate from more important segments can be protected better for user convenience. It is necessary to build a network architecture which is able to recognize the different importance of packets signaled by the sources, and behaves accordingly. All of the so far presented slot assignment schemes operate at the physical layer and they do not consider either semantic meaning of the content or decodability importance of network packets (i.e., how well they are concealed). However, overall user utility can be significantly increased using cross layer design, 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 occupancy levels and contribution of the received network packets in terms of visual quality are simultaneously maximized.

This paper is organized as follows: The packet importance concept both under the aspect of decodability and semantic meaning is introduced in Section 2, while the scheduling multi-objective optimization (MOO) formulation is outlined in Section 3. The method used for MOO solution is explained in Section 4. Experimental results with different settings are given in Section 5, and finally, conclusions are drawn in Section 6.

2. Packet Importance

2.1. Semantic Importance and Codec Modifications

Video contents have not a uniform semantic importance. There are several segments (shots) which can be of different interest for different users within a sequence. In a sport event transmission, play actions and replays are the most important parts. Videos can be segmented into parts with different semantic meanings automatically using existing algorithms (e.g. [6] for soccer games) in the literature. Naturally, the semantic importance depends strongly on the user preferences. Provided that we have a system that can collect each user's preferences for a set of scene types (e.g. play actions, replays and waiting times), 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, preserving the same overall average bitrate for the sequence.

In order to generate this effect, two major modifications have to be implemented within the encoder; first, each semantic region has to contain an integer number of GOPs, so variable GOP size has to be allowed, and moreover the bitrate control should be able to change its target value for each GOP while encoding the sequence.

The state-of-the art encoder works with fixed GOP structure, indicating the number of I/P-frames within a GOP and the number of B-frames between two consecutive P-ones. A semantic region can be formed up by one or more GOPs but, in order to properly change bitrate between regions, each one has to begin with an I-frame; given the variable number of frames that can be contained in each region, it is not possible to specify a fixed structure and the coding of I-frames has to be decided dynamically. Our solution allows to change the total number of frames within the GOP; if a region contains too many frames, we break it into smaller parts. For example, a region lasting 123 frames with a maximum GOP size of 30 frames will be separated into five GOPs, the first four containing 30 frames each, and the last one containing the remaining three frames; next region will then begin with an I-frame as required by the setting. It is also possible to have only one GOP for each region, but in this case, if a transmission error occurs, the effect in the decoded video can last until the end of the region; breaking down regions in GOPs lasting about one second could be a good compromise between coding efficiency and error recovery.

The reference H.264 codec includes bitrate control capabilities; it is built to allow the selection of an overall bitrate for the sequence, and it starts encoding with a user-defined quantization parameter. This implementation can require the time of one or two initial GOPs to converge properly if the initial quantization parameter is not chosen wisely. In this work, we need fast changes in the target bitrate;

in theory the desired per-GOP bitrate can switch between several values in few frames, and convergence time becomes a key point. Given a target bitrate for each GOP as defined by the level of semantic importance of the region it belongs to, we enforce the codec to automatically select the best possible starting quantization parameter from a pre-built table.

Whenever a bitrate switching is required, the codec will start encoding an I-frame choosing the closer quantization parameter; the standard bitrate control routine will then start and encode the GOP at the desired bitrate. The codec is able to work at whichever desired bitrate spanning from 3 Mbps to 15 kbps. Tests show that, whenever a bitrate switching is required, the target value is obtained with an error of 5% in the majority of the cases, even in the first GOP of the region.

This implementation is particularly useful if a sequence contains several scene changes, given its fast convergence to the target value.

2.2. Per-packet Decodability Importance

Defining per-region semantic importance allows selecting the coding bitrate, but 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. This decodability importance depends strongly on the role played in achieving compression by the frame a packet belongs to; packet losses within an I-frame can potentially propagate until the next key-frame is reached. More importance has to be assigned to those packets that would produce higher errors at the decoder, introducing also a relative importance among packets belonging to the same semantic region. The joint usage of both semantic and distortion importance can be mapped in priorities within a scheduling algorithm, to ensure transmission of high-importance packets also in network overload conditions.

Packets belonging to the same GOP have different importance according to the amount of distortion they would introduce in the decoded stream if lost, as defined in [7]. This importance can be computed at low cost if both the position of the frame within the GOP and the concealment technique used are known. The error is injected in the frame the packet comes from, and concealed by the error masking routine; the residual error can then be propagated to following frames my means of motion prediction. Distortion is measured as the MSE introduced by each loss in isolation, computed between the correctly decoded stream and the corrupted version; this number is real and has to be quantized over a given number of levels before being passed to the scheduler. The indicator of the distortion level should be used jointly with the semantic level of importance. The joint importance is computed as the product of the semantic importance level and the quantized decodability importance.

2.3. Definition of Packet Priorities

Traffic sources are scheduled for transmission if they experience good channel and buffer condition and are going to transmit high importance packets; at each time slot, we want to schedule the user that has the highest importance packet ready. Packets are reordered within each GOP according to the product of semantic importance and distortion importance levels; decodability importance is a real number and has to be quantized.

With this setting, we allow a packet coming from an I-frame of a low-importance region to be scheduled instead of a packet coming from a high-importance B-frame.

3. Problem Formulation

Due to additional bandwidth limitations, wireless communications require more careful managing of system resources compared to its wired counterpart. Visual quality of the received video is crucial as far as the mobile subscribers are concerned; hence over-compression of video information is not feasible for service providing companies. Therefore, 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 needed in order for a scheduling algorithm to work efficiently. 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 as shown in Figure 1.

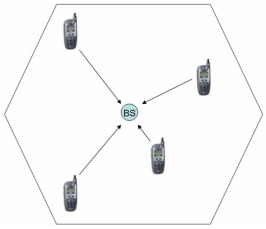

Figure 1. All users in the cell provide channel and buffer status feedback to the base station (BS).

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

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

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 4 needs to be applied.

4. Multi-Objective Optimization (MOO)

For single objective optimization problems, one can come up with one or more optimal solutions resulting in a *unique* optimal function value. 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.

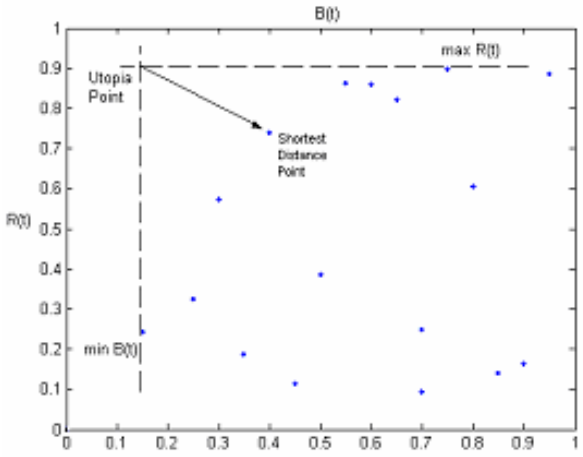

Figure 2. The proposed algorithm schedules the user whose corresponding point is closest to the utopia point.

In the proposed 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 2 for the two dimensional case, which is also described in [8]. The utopia point, *U(t)*, on the throughput-buffer-importance space is set as follows:

$$U(t) = \left(\overline{\Delta R}(t)_{\max}, B(t)_{\min}, imp_i(t)\right) \qquad (8)$$

A more detailed explanation of the multiple-objective optimization (MOO) techniques used in the literature can be found in [9]-[10].

5. Experimental Results

We encoded a 2250-frames test sequence (part of a soccer game) at 100 kbps average bitrate, where the time duration of the sequence is 90 seconds. Semantically important regions are coded at ten times the bitrate used for low-importance GOPs, obtaining 150 kbps versus 15 kbps ratio; this wide difference in bitrate has been chosen to assign as much as possible of the resources to semantically important regions, and corresponds to a scenario in which all of the users declare that they prefer to receive a very good stream in the high-importance part and nearly do not care of the low-importance regions (the ratio of importance is 1 to 10). The decodability importance has been quantized using two levels.

The resulting stream is fed into the scheduler using a 4-level (joint semantic and quantized decodability) importance indicator. One percent random packet losses are added to simulate bit errors, while other losses can be introduced by late packet delivery. To further stress the system, eight users require the same video at the same time, generating a peak joint bitrate of 1.2Mbps when semantically important frames are transmitted; this will result in packets not transmitted since their deadline expired. 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 regions. The maximum allowed initial buffering time is set at half or one second.

The PSNR results obtained for one second and half a second of pre-roll delays are shown in Table 1. PSNR losses are in the order of 1 dB for 1 second initial waiting time, and 2 dB for half a second, since the decoded video PSNR is lower due to higher packet loss rates in the latter.

To better show how packet priority increases system performance, packet loss rates and PSNR values obtained using the same compressed video and a "plain" scheduler (i.e., a scheduler that does not take into account packet importance) are shown in Table 2.

Table 1. Packet loss rates and PSNR for the test sequence, using two levels of semantic importance (150 and 15 kbps) and two levels of decodability importance.

User	1 s pre-roll		½ s pre-roll	
	PLR (%)	PSNR (dB)	PLR (%)	PSNR (dB)
no-loss	0	32.69	0	32.69
1	2.74	30.91	3.65	30.56
2	2.05	31.17	4.13	30.71
3	2.26	31.07	4.31	30.76
4	1.39	31.88	2.86	31.45
5	2.61	31.16	4.55	30.77
6	2.19	31.54	4.90	30.83
7	1.98	31.37	3.41	31.56
8	2.16	31.22	3.52	30.89

Table 2. Packet loss rates and PSNR for the test sequence, using no importance; the video is still encoded switching between two levels at 150 and 15 kbps.

User	1 s pre-roll		½ s pre-roll	
	PLR (%)	PSNR (dB)	PLR (%)	PSNR (dB)
no-loss	0	32.69	0	32.69
1	3.56	30.97	5.17	30.33
2	3.08	31.20	4.87	30.60
3	3.77	31.00	5.31	29.70
4	2.48	31.54	3.50	31.08
5	4.15	30.32	6.31	30.42
6	3.51	31.21	5.95	30.51
7	2.96	31.24	4.9	30.59
8	2.97	31.23	4.96	30.53

The results given in Table 2 show that for the given sequence required simultaneously by eight users, the use of packet priorities within the scheduler can ensure up to nearly 1 dB gain (for user 5) over the plain version. Furthermore, this approach gains even more if the maximum pre-roll delay is fixed at half a second (about 1 dB gain is observed for users 3 and 7).

Since we are more interested in the high-importance regions rather than low importance ones, we also encoded the sequence without differentiating over semantic importance, at the constant bitrate of 100kbps, and transmitted it over the simulated scheduler. Table 3 shows the comparison of the obtained packet loss rates and PSNR values for high-semantic importance frames.

Table 3. Packet loss rates and PSNR for semantically important frames, obtained by taking into account semantic importance during encoding or not.

User	With semantic importance		Without semantic	
	PLR (%)	PSNR (dB)	PLR (%)	PSNR (dB)
no-loss	0.00	37.01	0.00	34.50
1	1.45	35.60	1.16	33.30
2	1.67	35.27	0.71	33.92
3	1.74	35.13	1.33	33.37
4	0.86	36.42	0.86	33.93
5	2.00	35.87	1.19	33.29
6	1.65	35.83	1.00	33.64
7	1.52	35.45	0.94	33.64
8	1.45	35.49	0.73	33.86

Table 4. Packet loss rates and PSNR for the overall sequence, setting the same request time for all users and adding a randomized delay between 0 and 10 seconds.

User	All users at the same time		Randomized access	
	PLR (%)	PSNR (dB)	PLR (%)	PSNR (dB)
no-loss	0	32.69	0.00	32.69
1	2.74	30.91	0.97	32.22
2	2.05	31.17	1.29	31.42
3	2.26	31.07	0.99	31.94
4	1.39	31.88	0.84	32.02
5	2.61	31.16	1.30	31.63
6	2.19	31.54	1.97	31.16
7	1.98	31.37	0.76	32.02
8	2.16	31.22	3.65	31.10

Results show that coding the sequence at constant bitrate gives an average PSNR of 33.5 dB, which is higher than the values of tables 1 and 2. If we look only at the PSNR of the high-semantic importance regions, our performance is in the order of 35.5 dB, so in those shots the proposed coding outperforms the constant bitrate coding even if we experience higher loss rates.

The previously presented results are a lower bound to the performance, since all of the users require the video at the same time. Table 4 compares simultaneous access with a random access delay, uniformly distributed between zero and ten seconds; in both cases the users start playback one second after they require the video. Results show that in this case we get better PSNR mainly due to lower loss rates. With this simple randomization, only the last users accessing the network experience lower performance and higher loss rates.

6. 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.

7. References

[1] R. Knopp and P. A. Humblet, "Multiple accessing over frequency selective fading channels," in Proceedings of IEEE PIMRC 1995, vol. 3, pp. 1326-1330, Canada, September 1995.

[2] IS-856-2 cdma2000 High Rate Packet Data Air Interface Specification, TIA Std., Rev. 2, October 2002.

[3] P. Bender, P. Black, M. Grob, R. Padovani, N. Sindhushayana, and A. Viterbi, "CDMA/HDR: A bandwidth efficient 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), Brazil, 2001.

[6] A. Ekin, A. M. Tekalp and R. Mehrotra, "Automatic soccer video analysis and summarization," IEEE Trans. on Image Processing, vol. 12, no. 7, pp. 796-807, June 2003.

[7] 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, Italy, Sept. 2004.

[8] 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," submitted to IEEE GlobeCom 2005.

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

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