A CONTENT-AWARE SCHEDULING SCHEME FOR VIDEO STREAMING TO MULTIPLE USERS OVER WIRELESS NETWORKS

Peshala V. Pahalawatta, Randall Berry, Thrasyvoulos N. Pappas, and Aggelos K. Katsaggelos

EECS Department, Northwestern University
2145 Sheridan Rd, Evanston, IL 60208, USA
phone: + (1) 847-491-3039, fax: + (1) 847-491-4455, email: (pesh, rberry, pappas, aggk)@ece.northwestern.edu

ABSTRACT

There is a rapidly growing interest in high speed data transmission over digital cellular networks. This interest is fueled mainly by the need to provide multimedia content to mobile users. In this paper, we present a packet scheduling scheme that can be used for real-time streaming of pre-encoded video over downlink packet access wireless networks. We consider a gradient-based scheduling scheme in which user data rates are dynamically adjusted based on their channel quality as well as the gradients of a utility function. The utility functions are designed by taking into account the distortion of the received video. They allow for content-aware packet scheduling both within and across multiple users. Simulation results show that the gradient-based scheduling framework, when combined with the distortion-aware utility functions, significantly outperforms conventional content-independent packet scheduling schemes.

1. INTRODUCTION

Since the advent of digital cellular networks, there has been a rapid growth of interest in the area of high speed data transmission to mobile users. The growing demand stems mainly from the need to supply multimedia content, such as on-demand streaming video, to mobile clients. The emerging 3G cellular standards, such as *HSDPA* (High Speed Downlink Packet Access) [1], and possibly 4G standards, such as IEEE 802.16 (*WiMAX*) [2], aim to provide sufficiently high data throughputs that will enable the introduction of a broad range of wireless multimedia applications.

The search for higher data throughputs has greatly benefited from an understanding of the *multi-user diversity gain* [3], achieved when scheduling to multiple users. Systems that exploit multi-user diversity dynamically schedule greater resources to users with better channel quality, while delaying transmissions to users with degraded channels. To avoid consistently favoring users with better average channels (e.g., users closer to the base station), proportionally fair scheduling schemes have been proposed that take into account the average throughput for each user, in addition to the channel conditions. Other scheduling schemes, such as the *Exponential Rule* [4], attempt to stabilize the queues at the transmission buffer. In [5], a more general, gradient-based scheduling scheme is used, such that the user rates weighted by the gradients of a given system utility function will be maximized.

While the above schemes increase the overall data throughput, they do not explicitly consider the transmitted media content when making scheduling decisions. The quality of a decoded video stream, however, is only loosely dependent on the raw data throughput. Other factors, such as the error concealment schemes used at the decoder, can determine the relative importance of individual data packets. Therefore, a scheduling technique that takes into account the relative importance of video packets, ought to outperform the content independent techniques discussed above. In this paper, we build on the general gradient-based scheduling technique in [5], to develop a content-aware scheduling scheme for a wireless multi-user video streaming system.

Related work on video streaming has appeared in [6], [7], [8], [9], and [10]. In [6], [7], and [8], the resource allocation is performed based on the queue lengths at each user's transmission buffer, or based on packet deadlines. The information content of the data packets is not considered at the scheduler. In [9], a heuristic approach is used to determine the importance of frames across users based on the frame type (I, P, or B), or its position in a group of pictures. In [10], a concept of additive distortion among video packets, introduced in [11], is used to determine the importance of video packets for each user. Scheduling across users, however, is performed using conventional, content-independent techniques.

A content-aware resource allocation scheme is proposed in [12], where the priority across users is determined as a combination of an importance measure similar to that in [10], and the delay of the *Head Of Line* (HOL) packet for each user. At each time slot, all the resources are dedicated to the user with the highest priority. In our scheme, we consider 3G wireless architectures, such as HSDPA, that allow for a combination of TDM, and CDMA, to allocate resources simultaneously to multiple users at a given transmission opportunity. The scheme also departs from [12] in that it is performed at each transmission time slot based on the instantaneous channel fading states of each user, thus enabling a tighter coupling between the channel resource allocation and content-dependent priority metric. The key idea in this scheme is to order individual video data packets by their information content, and to specify a distortion-based utility function, which will enable the prioritization of the data packets across multiple users as well as within a single user. We consider error robust data packetization schemes, at the encoder, and realistic error concealment schemes, at the decoder.

In Sec. 2, we provide a brief overview of the system and discuss the resources and constraints inherent to the problem. The problem is formulated in Sec. 3, and the solution summarized in Sec. 4. In Sec. 5, we discuss our proposed distortion based utility function, which integrates the gradient-based scheduling framework into a content-aware scheduling scheme. The effects of error concealment schemes used at the decoder on the distortion estimation are considered in Sec. 6. We provide some simulation results in Sec. 7.

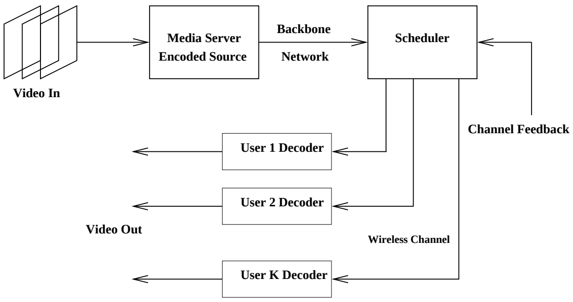


Figure 1: Wireless multiuser video streaming system

2. MODEL

2.1 System Overview

Figure 1 illustrates the type of system considered in this paper. The system begins at a media server, which contains multiple pre-encoded video sequences. We assume that the video is packetized into independently decodable packets where each packet represents a *slice* (group of macroblocks) of the video frame. Once a particular video stream is requested by a client, the video packets are transmitted over a backbone network, to the scheduler servicing multiple clients. We assume that the backbone network is lossless and of high bandwidth. The scheduler allocates resources to each client based on the video content, and *Channel State Information* (CSI) available through channel feedback.

The transmitted video packets are received and decoded by the clients, where lost or dropped packets are concealed using a video error concealment technique. In a video streaming system, all the data packets required in order to decode a frame of the video sequence must be received at the decoder buffer prior to the playback time of that frame. Therefore, each group of packets contains a stringent delay constraint, and any packet that has not been transmitted prior to its deadline, is dropped by the scheduler.

2.2 Channel Resources and Constraints

The resource allocation scheme combines TDM and CDMA such that, at a given time-slot, the scheduler can decide on the number of orthogonal spreading codes, n_i, to be used to transmit to a given user, i. Note that $n_i = 0$ implies that user i is not scheduled for transmission at that time-slot (The time-slot index is omitted for clarity throughout this section). The maximum number of spreading codes per user, N_i, is limited by the specific mobile device used by the client. The total number of spreading codes, N, allocated to all users, is limited by the standard ($N = 15$ for HSDPA). The scheduler can decide on the power level, p_i, used to transmit to a given user. The total power, P, that can be used by the base station is also limited in order to restrict interference across neighboring cells. Assuming K total users, these constraints can be written as,

$$\sum_{i=1}^{K} n_i \le N, \quad \sum_{i=1}^{K} p_i \le P, \quad n_i \le N_i. \tag{1}$$

We assume that the constraints of the system will be such that the transmitter may not be able to transmit all the available video packets to every user prior to their transmission deadlines.

3. PROBLEM FORMULATION

3.1 General Problem Definition

We assume that the channel state for user i, denoted e_i, at a given time-slot is known based on CSI available to the scheduler. The value of e_i represents the normalized *Signal to Interference Noise Ratio* (SINR) per unit power and can vary quite rapidly, and in a large dynamic range, over time. Defining $SINR_i = \frac{p_i}{n_i} e_i$ to be the SINR per code for user i at a given time, we can model the achievable rate per code for user i, denoted by r_i, as,

$$r_i = n_i B \log\left(1 + \frac{p_i e_i}{n_i}\right), \tag{2}$$

where B is the symbol rate, and it is constant.

Now the problem becomes one of specifying the n_i and p_i allocated to each user such that a target rate, r_i, can be achieved.

3.2 Gradient-Based Scheduling Framework

We use a technique proposed in [5] to perform the scheduling task. This technique maximizes the projection of the achievable rate vector, $\mathbf{r} = (r_1, r_2, ..., r_K)$ on to the gradient of the system utility function. The utility function is defined as, $\mathbf{U} = \sum_{i=1}^{K} U_i$, where U_i is a concave user utility function. The problem can be written as,

$$\max_{\mathbf{r}} \sum_{i=1}^{K} u_i r_i \tag{3}$$

where u_i is the gradient of U_i.

Now, taking into account the constraints specified in (1), as well as the formula for calculating each user's achievable rate specified in (2), we can formulate the optimization problem as,

$$\begin{aligned} V^* &:= \max_{(\mathbf{n},\mathbf{p}) \in \chi} V(\mathbf{n},\mathbf{p}), \\ &\text{subject to:} \\ \sum_{i=1}^{K} n_i \le N, &\qquad \sum_{i=1}^{K} p_i \le P, \\ &\text{where:} \\ V(\mathbf{n},\mathbf{p}) &:= \sum_{i=1}^{K} u_i n_i \log\left(1 + \frac{p_i e_i}{n_i}\right), \\ \chi &:= \{(\mathbf{n},\mathbf{p}) \ge 0 : n_i \le N_i \ \forall i\}. \end{aligned} \tag{4}$$

In addition to the main constraints specified above, a practical system will also be limited by some "per-user" constraints. Among them are, a peak power per user, a maximum SINR per code for each user, and the maximum and minimum coding rates allowed by the coding scheme.

The above constraints can be grouped into a *per user power constraint* based on the SINR per code for each user [5]. This constraint can be viewed as,

$$SINR_i = \frac{p_i e_i}{n_i} \in \left[\check{s}_i(n_i), s_i(n_i)\right], \quad \forall i, \tag{5}$$

where $\check{s}_i(n_i) \ge 0$. For the purposes of this work, we have conformed to the case where the maximum and minimum SINR constraints are not functions of n_i, i.e., $SINR_i \in [\check{s}_i, s_i]$.

4. SOLUTION

A solution to the optimization problem of the type given in (4) for the case when the maximum and minimum SINR constraints are not functions of n_i is derived in detail in [5]. In this section, we will briefly summarize the solution.

The Lagrangian for the primal problem in (4) can be defined as,

$$\begin{aligned}\mathscr{L}(\mathbf{p},\mathbf{n},\lambda,\mu) = \quad & \sum_i u_i n_i \log\left(1+\frac{p_i e_i}{n_i}\right) + \lambda\left(P-\sum_i p_i\right) \\ & + \mu\left(N-\sum_i n_i\right).\end{aligned} \tag{6}$$

Based on (6), we can define the dual function,

$$\mathscr{L}(\lambda,\mu) = \max_{(\mathbf{n},\mathbf{p})\in\chi} \mathscr{L}(\mathbf{p},\mathbf{n},\lambda,\mu), \tag{7}$$

which can be analytically computed by first keeping $\mathbf{n},\lambda,\mu$ fixed and optimizing (6) over $\mathbf{p}$, and then optimizing over $\mathbf{n}$. Then, the dual problem is given by,

$$\mathscr{L}^* = \min_{(\lambda,\mu)\geq 0} \mathscr{L}(\lambda,\mu), \tag{8}$$

which, by keeping λ fixed and optimizing over μ, we can write as,

$$\mathscr{L}(\lambda) = \min_{\mu\geq 0} \max_{(\mathbf{n},\mathbf{p})\in\chi} \mathscr{L}(\mathbf{p},\mathbf{n},\lambda,\mu). \tag{9}$$

Based on the concavity of V in (4), and the convexity of the domain of optimization, it can be shown that a solution to the dual problem exists, and that there is no duality gap.

Since, $\mathscr{L}(\lambda)$ is also convex (proof given in [5]), we can find the optimal λ using a one-dimensional convex search procedure that has a geometric convergence rate.

5. DISTORTION-BASED UTILITY FUNCTION

The main contribution of our work is to devise a utility function that relates closely to the quality of the received video and fits well with the gradient based scheduling framework. We propose a distortion-based utility function that takes into account the relative contribution of each data packet to the overall quality of video for each user. The key idea in this method is to prioritize the video data packets by the reduction in distortion that they would provide to the user.

We assume that the current video frame to be transmitted to user i consists of M_i slices, and that each slice, m, consists of $b_{i,m}$ bits. Now, let $\Pi_i = \{\pi_{i,1},\pi_{i,2},...,\pi_{i,M_i}\}$ be the re-ordered set of slices such that $\pi_{i,1}$ will be the first slice of the frame to be transmitted, and given enough time is available to transmit the entire frame, π_{i,M_i} will be the last slice to be transmitted. Let $D_i[\{\pi_{i,1},\pi_{i,2},...,\pi_{i,k_i}\}]$, denote the distortion given the first k_i prioritized slices are transmitted to user i and the remaining $(M_i - k_i)$ slices are dropped. Then, we define the user utility for user i after k_i slice transmissions as,

$$U_i[k_i] = (D_i[\Pi_i] - D_i[\{\pi_{i,1},\pi_{i,2},...,\pi_{i,k_i}\}]), \tag{10}$$

where $D_i[\Pi_i]$ would be the minimum distortion for the frame, which will occur when all slices are received. In our simulations, we define the distortion to be the sum absolute pixel difference between the decoded and error-free frames. For ease of notation, let $\Pi_i(k_i) = \{\pi_{i,1},...,\pi_{i,k_i}\}$. Then, we can guarantee that the user utility function is concave and increasing by iteratively choosing each additional slice π_{i,k_i+1} such that,

$$\pi_{i,k_i+1} = \arg\max_{m\notin\Pi_i(k_i)} u_{i,m}[k_i], \tag{11}$$

where,

$$u_{i,m}[k_i] = \frac{D_i[\Pi_i(k_i)] - D_i[\{\Pi_i(k_i),m\}\,|\Pi_i(k_i)]}{b_{i,m}}. \tag{12}$$

In (12), $D_i[\{\Pi_i(k_i),m\}\,|\Pi_i(k_i)]$ indicates that the distortion after adding slice m is dependent on the currently ordered set of slices $\Pi_i(k_i)$, which will be true if a complex error concealment technique is used at the decoder (See Sec. 6).

Now, we can use the utility gradients, $u_{i,\pi_{k_i+1}}[k_i]$ in (12) with the gradient based scheduling framework in Sec. 3.2 in order to ensure that the resource allocation will explicitly consider the improvements in quality of service for each user.

6. ERROR CONCEALMENT

Most video codec implementations employ some form of error concealment to alleviate the distortion due to lost packets. Therefore, any method that chooses to drop video packets must consider the role of error concealment at the receiver.

Error concealment techniques exploit either the temporal correlations between neighboring frames, or spatial correlations across pixels in the same frame, to estimate lost pixel values. A simple temporal concealment strategy is to replace the MBs (MacroBlocks) of a lost slice of the image with MBs from the same position in the previous frame. This technique does not introduce any dependencies between slices in the same frame. Therefore, if simple concealment is used, then the prioritization scheme in (11), is optimal and the dependency term in (12) is not necessary.

Complex temporal concealment techniques, however, do introduce dependencies across slices. For example, in one such method, the decoder uses the motion vector (MV) from a neighboring received MB in order to find the motion compensated concealed block. The MVs from each neighboring block are tested based on a boundary matching criterion, which calculates the distortion between the concealed block and correctly received blocks at their bordering edges. The MV that provides the minimum edge distortion is chosen for the final error concealment [13]. With complex concealment, our slice ordering scheme is *myopic* in that the contribution of each new slice depends on the set of slices of the same frame that have already been chosen to be transmitted. Unlike simple concealment, complex concealment with myopic slice ordering does not always guarantee a concave utility function. For example, Fig. 2 shows a plot of the distortion gradients after slice ordering for slices in a particular frame of the Foreman sequence. It can be seen that the addition of the third ordered slice gives a greater reduction in distortion per bit than the first two slices, given that the first two slices have already been transmitted. Therefore, in our implementation, after the slice ordering, Π_i is found, we re-estimate the utility gradients at each slice by averaging the distortion reduction over a window of L successive ordered slices such

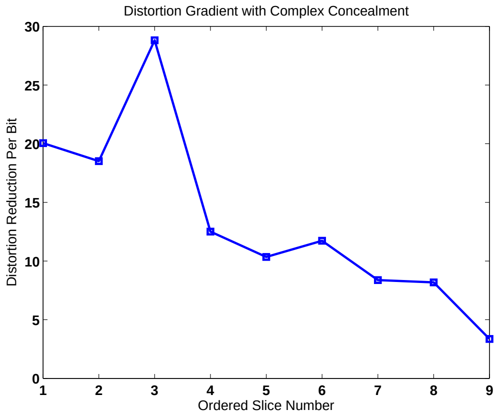


Figure 2: Non-monotonic gradient after complex concealment

that,

$$u_i[k_i] = \frac{D_i[\Pi_i(k_i)] - D_i[\Pi_i(k_i+L)]}{\sum_{m=k_i+1}^{k_i+L} b_{i,\pi_{i,m}}}, \qquad (13)$$

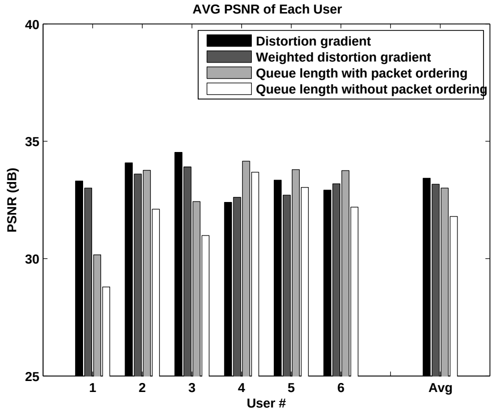
where $u_i[k_i]$ would be the utility gradient used in the gradient-based scheduling scheme for user i when the first k_i slices have already been transmitted. Empirical results show that $L = 3$ provides a sufficient degree of smoothing.

7. SIMULATION RESULTS

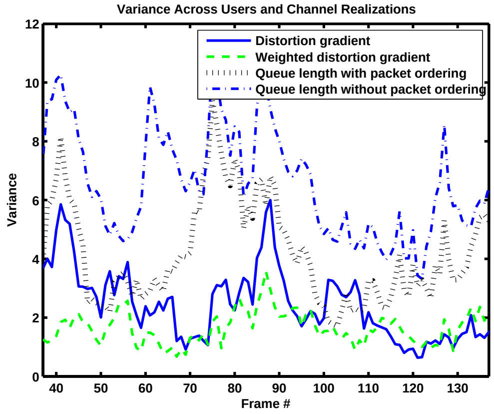
Six video sequences with varied content: "foreman", "carphone", "mother and daughter", "news", "hall monitor", and "silent", in QCIF (176x144) format were used for the simulations. The sequences were encoded in H.264 (JVT reference software, JM 9.3 [14]) at variable bit rates to obtain a specified average PSNR of 35dB. All frames except the first were encoded as P frames. To reduce error propagation due to packet losses, random I MBs were inserted into each frame. The frames were packetized such that each slice contained one row of MBs, which enabled a good balance between error robustness and compression efficiency. Constrained intra prediction was used at the encoder for further error robustness. The wireless network was modeled as an HSDPA system with $N = 15$, which is the actual total number of spreading codes available in HSDPA, and $N_i = 5$ for each user. HSDPA provides 2 msec transmission timeslots. Realistic channel traces for an HSDPA system were obtained using a proprietary channel simulator developed at Motorola Inc.

Figure 3 compares the average quality of the received video, with simple error concealment at the decoder, using 4 different methods for calculating the utilities in (3). The first, is the proposed technique, which uses the utility functions described in Sec. 5. The second, is a minor modification of the first, where the utility gradients are also weighted by the distortion of the currently transmitted sequence to ensure fairness across users. The third method is only partially content-aware, in that it orders the video packets of each user according to their importance. The resource allocation across users, however, is performed assuming that the utility gradients in (3) are proportional to the current queue lengths at each user's transmission buffer. In terms of computational complexity, the first three methods are very similar and they use the proposed packet ordering scheme. The final method, is similar to the conventional content-independent scheduling techniques, where no packet ordering is performed at the scheduler.

Figure 3(a) shows that the content-aware schemes can significantly out-perform the conventional queue length based scheduling scheme. Also, Fig. 3(b) shows that the two content-aware schemes for resource allocation, tend to provide similar quality across all the users (lower variance), while queue-dependent resource allocation schemes tend to favor some users over others. Between the two content-aware schemes, we can see that a small sacrifice in average PSNR yields significant improvement in terms of the variance across users.



(a) Average PSNR over 100 frames, and 5 channel realizations.



(b) Variance of PSNR across users

Figure 3: Average and variance of PSNR. User #'s in (a) represent sequences: 1- Foreman, 2- Mother and Daughter, 3- Carphone, 4- News, 5- Hall Monitor, 6- Silent. P = 10W, Max SINR = 1.5

Figure 4 shows a comparison with weighted distortion gradients using simple and complex concealment at the decoder. Note that the algorithm provides the optimal slice ordering with the simple concealment scheme, while a suboptimal myopic solution is used for the complex concealment scheme. Overall, the complex concealment scheme tends to perform better for all the users despite the suboptimal packet ordering. The third scheme in Fig. 4 is when the complex concealment used at the decoder is not accounted for in the distortion utility function. We can see that the content-aware scheme is sensitive to mismatch between error concealment techniques used at the scheduler and receiver.

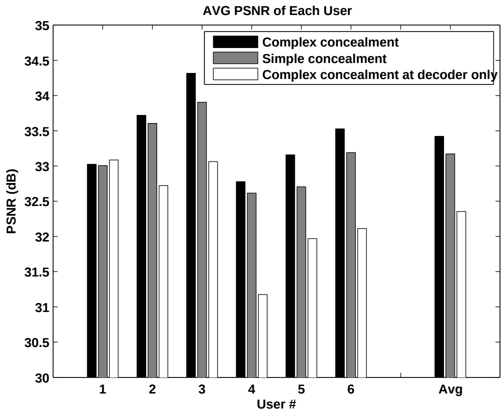


Figure 4: Average PSNR with simple and complex error concealment. P = 10W, Max SINR = 1.5

The previous simulations assume a realistic scenario according to which the original video frames are not available at the scheduler in order to compare with the decoded frames after packet losses. Therefore, the distortion calculations at the scheduler are performed using the decoded video frames assuming no packet losses. Figure 5 shows the average performance loss when the original video frames are not used. While the performance loss is not significant, it can be further reduced by transmitting a higher quality video stream to the scheduler through the high bandwidth backbone, which can then be used for estimation of the distortion metric.

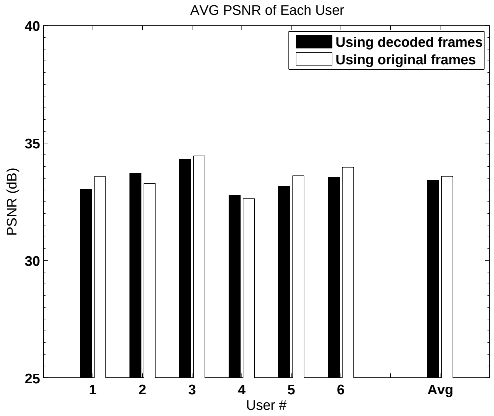


Figure 5: Average PSNR using decoded vs original frames at scheduler

8. CONCLUSIONS

We introduced a new content-aware multi-user resource allocation scheme for downlink video streaming in a wireless network. The approach, which is discussed in the context of a TDM/CDMA system, is also equally applicable to other systems such as TDM/OFDM, or even pure TDM. We have shown that the scheme significantly out-performs current, content-independent, scheduling techniques. We have also shown that content-independent buffer management techniques tend to favor some users over others based on their video content. In the future, we plan to extend the work to cover scalable coded bit streams, and also consider fast hybrid ARQ schemes that can be used with the emerging wireless standards.

REFERENCES

[1] 3GPP, *High Speed Downlink Packet Access*, Overall description, December 2004, TS 25.308 v6.3.0.

[2] *Part 16: Air interface for fixed broadband wireless access systems*, October 2004, IEEE Std 802.16-2004.

[3] R. Knopp and P.A. Humblet, “Information Capacity and Power Control in Single-Cell Multiuser Communications,” in *Proc. of IEEE Int. Conference on Communications*, 1995.

[4] S. Shakkottai, R. Srikant, and A. Stolyar, “Pathwise Optimality and State Space Collapse for the Exponential Rule,” in *Proc. of the IEEE International Symposium on Information Theory*, 2002.

[5] R. Agrawal, V. Subramanian, and R. Berry, “Joint Scheduling and Resource Allocation in CDMA Systems,” in *2nd Workshop on Modeling and Optimization in Mobile, Ad Hoc, and Wireless Networks (WiOpt ’04)*, March 2004.

[6] Y. Ofuji, S. Abeta, and M. Sawahashi, “Unified Packet Scheduling Method Considering Delay Requirement in Forward Link Broadband Wireless Access,” in *Proc. of Vehicular Technology Conference*, Fall 2003.

[7] P. Falconio and P. Dini, “Design and Performance Evaluation of Packet Scheduler Algorithms for Video Traffic in the High Speed Downlink Packet Access,” in *Proc. of PIMRC 2004*, September 2004.

[8] D.H. Kim, B.H. Ryu, and C.G. Kang, “Packet Scheduling Scheme for real time Video Traffic in WCDMA Downlink,” in *Proc. of 7th CDMA International Conference*, October 2002.

[9] R.S. Tupelly, J. Zhang, and E.K.P. Chong, “Opportunistic Scheduling for Streaming Video in Wireless Networks,” in *Proc. of Conference on Information Sciences and Systems*, 2003.

[10] G. Liebl, T. Stockhammer, C. Buchner, and A. Klein, “Radio Link Buffer Management and Scheduling for Video Streaming Over Wireless Shared Channels,” in *Proc. of the Packet Video Workshop*, 2004.

[11] P.A. Chou and Z. Miao, “Rate Distortion Optimized Streaming of Packetized Media,” *IEEE Trans. Multimedia*, To appear.

[12] G. Liebl, M. Kalman, and B. Girod, “Deadline-Aware Scheduling for Wireless Video Streaming,” in *Proc. IEEE Int. Conf. on Multimedia and Expo*, July 2005.

[13] Y-K. Wang, M.M. Hannuksela, V. Varsa, A. Hourunranta, and M. Gabbouj, “The Error Concealment Feature in the H.26L Test Model,” in *Proc. IEEE Int. Conference on Image Processing*, 2002.

[14] *JVT Reference Software*, `http://iphome.hhi.de/suehring/tml/download`, JM 9.3.