Resource Allocation for Downlink Multiuser Video Transmission Over Wireless Lossy Networks

Ehsan Maani, *Student Member, IEEE*, Peshala V. Pahalawatta, *Member, IEEE*, Randall Berry, *Member, IEEE*, Thrasyvoulos N. Pappas, *Fellow, IEEE*, and Aggelos K. Katsaggelos, *Fellow, IEEE*

Abstract—Demand for multimedia services, such as video streaming over wireless networks, has grown dramatically in recent years. The downlink transmission of multiple video sequences to multiple users over a shared resource-limited wireless channel, however, is a daunting task. Among the many challenges in this area are the time-varying channel conditions, limited available resources, such as bandwidth and power, and the different transmission requirements of different video content. This work takes into account the time-varying nature of the wireless channels, as well as the importance of individual video packets, to develop a cross-layer resource allocation and packet scheduling scheme for multiuser video streaming over lossy wireless packet access networks. Assuming that accurate channel feedback is not available at the scheduler, random channel losses combined with complex error concealment at the receiver make it impossible for the scheduler to determine the actual distortion of the sequence at the receiver. Therefore, the objective of the optimization is to minimize the expected distortion of the received sequence, where the expectation is calculated at the scheduler with respect to the packet loss probability in the channel. The expected distortion is used to order the packets in the transmission queue of each user, and then gradients of the expected distortion are used to efficiently allocate resources across users. Simulations show that the proposed scheme performs significantly better than a conventional content-independent scheme for video transmission.

Index Terms—Cross-layer design, error concealment, H.264/AVC, HSDPA, video streaming, wireless packet scheduling.

I. INTRODUCTION

RECENT years have witnessed a drastic growth in demand for multimedia services such as video streaming on mobile terminals. The technology available for high data-rate multimedia services to mobile clients over wireless networks is rapidly improving with the emergence of third generation and newer wireless standards such as HSDPA and IEEE 802.16 [1], [2]. In most scenarios, multiple video sequences are transmitted to multiple users simultaneously by sharing a resource-limited wireless network. Transmitting multiple compressed video programs over a wireless network in real time is considered a challenging task due to several reasons. First, wireless channels are impaired by deleterious effects such as fading and co-channel interference (CCI). Second, resources such as bandwidth and power are limited in wireless networks and have to be shared among multiple users. Furthermore, enormous fluctuations in rates of compressed video programs due to the differences in video content and intra/intercoding modes, can complicate resource allocation.

Wireless resource allocation and scheduling approaches can be categorized into two classes: i) time-division multiplexed (TDM) systems, where a single user is transmitted to in each time-slot, as in CDMA 1xEVDO, and ii) systems in which the transmitter can simultaneously transmit to multiple users in each time-slot. These systems employ a combination of TDM and another multiplexing technique such as CDMA or OFDM. In this article, we focus on the second class of systems where in addition to deciding which users to schedule, the available physical layer resources (bandwidth and power) are optimally allocated to the users.

Traditionally, cross-layer scheduling and resource allocation methods exploit the time varying nature of the wireless channel to maximize the throughput of the network while maintaining fairness across multiple users [3]–[5]. These methods rely on the multiuser diversity gain achieved by selectively allocating a majority of the available resources to users with good channel quality who can support higher data rates. In [6], the authors discuss the implementation of gradient-based scheduling schemes. Optimization over the available resources is performed at each time-slot while taking into account the fading state of each user at that time. The utility function used in [6] is defined as either a function of each user's current average throughput, or of each user's queue length or delay of the head-ofline packet. A queue-length based utility can be employed for video streaming applications where the delay constraints are stringent.

Video quality, however, is not simply a function of the data throughput but is also determined by the video content because of inefficiencies in video compression, as well as the potential for spatial and temporal error concealment of lost/missing data [7]. Furthermore, an important requirement in video streaming is that the video will be played back in real-time at the decoder, and, therefore, the appropriate video packets need to be available at the decoder in time for playback. Therefore, any packet that remains in the transmission queue after its decoding time has expired will be discarded prior to transmission [7], [8]. Consequently, in order to efficiently utilize the limited resources of the wireless networks for video delivery, a content-aware scheduling technique must be employed. Methods that have been specifically designed for video applications have conventionally focused on satisfying the delay constraint requirements inherent to the system [9]–[11]. Received video quality

in these approaches is only measured as a function of delay or packet loss rate. In [8], a concept of incrementally additive distortion among video packets, introduced in [12], is used to determine the importance of video packets. Scheduling across users, however, is performed using conventional, content-independent techniques. In [7] and [13], we have developed a content-based utility function that can be integrated into the utility-based framework of [6] to provide a content-aware scheduling technique. The key to deriving this utility function is to appropriately prioritize the video packets according to their "importance." This importance is measured in terms of the distortion of the received video signal, taking into account the concealment of the lost packets. The approach in [7], however, assumes that perfect channel information is available at the scheduler and, therefore, uses a zero-outage capacity model to determine the achievable data rates. Hence, the losses considered in [7] and [13] occur only when packets are not transmitted on time due to their scheduling priority, and as a result all losses are known to the transmitter. In the current work, we consider a realistic scenario in which only an imperfect estimate of the channel state is available at the transmitter. In this case, an outage capacity model must be used to determine a probability of channel loss based on the estimated channel state, the allocated resources, and the transmission rate [14]. We still assume that the error concealment strategy utilized by the decoder is also known to the transmitter and, therefore, is employed by the scheduler to achieve better performance. Random channel losses combined with complex error concealment at the decoder make it impossible for the scheduler to determine the actual distortion of the sequence at the receiver. Instead, the scheduler employs a per-pixel decoder distortion estimation to determine its scheduling decisions. Efficient methods exist for recursively calculating the expected distortion at the receiver [15], [16]. The main contribution of this paper is to provide a method for calculating a prioritized set of video packets in which the packets are ordered by their contribution towards reducing the expected distortion of the received video. Using this scheme, we jointly optimize the resource allocation (power and bandwidth) and transmission rates assigned to each user to reduce the end-to-end distortion estimate over all users in the system.

The rest of the paper is organized as follows. In Section II, we provide a brief overview of the system. In Section III, we formulate the problem by describing the error concealment technique, the proposed packet ordering technique, and the wireless channel model. The resource allocation problem is formalized in Section IV and a simplified solution to the problem is proposed. Experimental results are shown in Section V and finally conclusions are drawn in Section VI.

II. SYSTEM OVERVIEW

Fig. 1 depicts a generic framework for multiuser video transmission over wireless networks, which consists of the video server/encoder, backbone network, scheduler, and the receivers. Captured video sequences are first compressed by the video encoder and recorded in a media server. We assume that each sequence is packetized into multiple data units. Each data unit/packet is independently decodable and represents a slice of the video. Note that, although in terms of decoder operation,
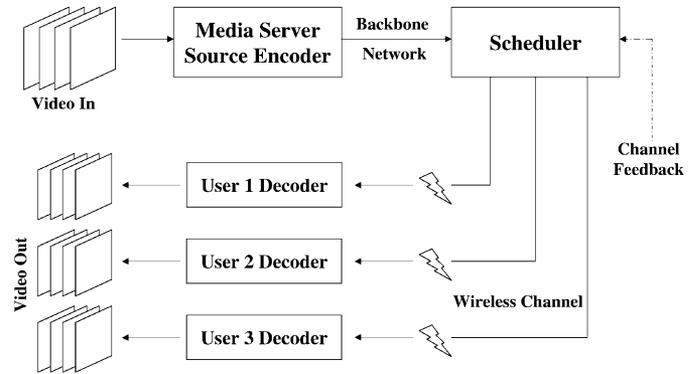
Fig. 1. Block diagram of multiuser downlink video streaming system.

each slice is independently decodable, in reality, most frames of a compressed sequence are inter frames, in which MBs may be dependent on macroblocks of previous frames through motion prediction. Once a video stream is requested by a client, the packets are transmitted over a backbone network (assumed lossless) to the scheduler at a base station servicing multiple clients. In addition to Channel State Information (CSI) available through channel feedback, the scheduler uses three features of each packet to allocate resources across users. These features, for each packet m of each client i, are the utility gained due to the transmission of the packet, the size of the packet in bits, $R_{i,m}$, and the decoding deadline for the packet, $\tau_{i,m}$, which represents the delay constraint in order to reach the receiver in time for playback. This decoding deadline is determined by the frame rate of the video being streamed. We assume that all the packets in a frame have the same decoding deadline. Any packet left in the transmission queue after its decoding deadline has expired is discarded since it has lost its value to the decoder. In other words, there are a specific number of time slots available for transmission of each frame depending on the streaming frame rate, and after those time slots have elapsed, no further packet from the current frame is transmissible. Hence, only packets that arrive intact and on time at the receiver are decoded by the decoder. Multiple retransmissions of a packet based on feedback are allowed as long as the retransmissions occur prior to the decoding deadline. In this scenario, the lost packet is reinserted into the transmission queue and re-ordered based on the current state of the queue. Errors introduced to the decoded image are due to the loss of packets in the wireless channel, or due to the discarding of packets from the transmission queue. These errors are concealed using an error concealment technique.

III. PACKET ORDERING WITH EXPECTED DISTORTION

Careful packetization of the video data is necessary to ensure the optimal tradeoff between channel utilization and error robustness. In addition, since we require each packet to be decodable by itself, small packet sizes will degrade the source compression efficiency due to limited prediction. On the other hand, large packet sizes result in greater packet loss probability and ineffective concealment in case of a packet loss. Note that error concealment in this work not only helps error hiding at the decoder, but it also plays an important role in packet ordering and

resource allocation. Based on the above discussion, it is assumed in this work that a slice consists of a row of macroblocks and is directly packetized into a transport packet.

The video data packets, then, are ordered in the scheduler buffer such that the most important packets are first served and, therefore, have a greater likelihood of being received at the decoder. Packet prioritization and resource allocation in this work is performed one frame at a time. Nonetheless, this scheme can potentially be improved by optimizing the scheduling and resource allocation over multiple buffered frames. Such a scheme, however, would lead to a considerably higher computational complexity. Using an outage capacity model, the probability of loss of each transmitted packet can be estimated based on the imperfect channel state information available at the scheduler. The error concealment technique employed at the decoder is assumed to be known by the scheduler. The rest of this section describes a method for ordering each user's video packets within each user based on the contribution of the packets towards reducing the expected distortion of the sequence. Since the same technique is used for all users, the user index, i, is omitted during this discussion.

A. Error Concealment and the Calculation of Expected Distortion

Due to channel losses, we use the expected end-to-end distortion to evaluate video quality. Three factors can be identified as affecting the end-to-end distortion: the source behavior (quantization and packetization), the channel characteristics, and the receiver behavior (error concealment) [17], [18].

A robust error concealment technique helps avoid significant visible errors in the reconstructed frames at the decoder. Currently, there does not exist a standardized error concealment scheme for wireless communication. This work, however, assumes that the error concealment scheme is known at both the transmitter and the decoder. Given the importance of error concealment in determining the final decoded quality of the transmitted video, a protocol in which the error concealment scheme is known to both the scheduler and decoder can potentially be highly beneficial in providing significant performance improvements through content-aware packet scheduling schemes. In our previous work [7], we have examined this issue in greater detail. In this work, we consider a simple but efficient temporal concealment scheme: a lost macroblock (MB) is concealed using the median motion vector candidate of its received neighboring MBs (the top-left, top, and top-right). The candidate motion vector of a MB is defined as the median motion vector of all 4×4 blocks in the MB. If the preceding row of MBs is also lost, then the MB in the same spatial location in the previously reconstructed frame is used to conceal the current loss. Note that this concealment strategy is employed both in the scheduler optimization framework and at the decoder.

Given the dependencies introduced by the error concealment scheme, and assuming dependent packet cases, the expected distortion of the mth slice $E\{D_m\}$, can be calculated at the encoder as

$$E\{D_m\} = (1 - \epsilon_m)E\{D_{R,m}\} + \epsilon_m(1 - \epsilon_{m-1})E\{D_{LR,m}\}$$
$$+ \epsilon_m \epsilon_{m-1} E\{D_{LL,m}\} \quad (1)$$

where ϵ_m is the loss probability of the mth packet, $E\{D_{R,m}\}$ is the expected distortion of the mth packet if received, and $E\{D_{LR,m}\}$ and $E\{D_{LL,m}\}$ are respectively the expected distortion of the lost mth packet after concealment when packet $(m-1)$ is received or lost. Note that in this equation ϵ_{-1} is always equal to 1.0 since there is no packet before the first packet $(m = 0)$. Assuming an additive distortion measure, the expected distortion of a frame of M packets, denoted by $E\{D\}$, can be written as

$$E\{D\} = \sum_{m=1}^{M} E\{D_m\}. \quad (2)$$

This distortion measurement is based on a per pixel recursive algorithm called ROPE, which was originally proposed in [15] as an efficient means to accurately estimate end-to-end distortion at the encoder.

The accuracy of ROPE in end-to-end distortion estimation is attributed to its ability to calculate the first and second moments of the decoder reconstructed pixels. Sub-pixel prediction employed in H.264/AVC, however, involves interpolation of neighboring pixels [19], which gives rise to cross-correlation terms in the second moment calculation. To deal with the cross correlation terms in our experiments, the cross correlation approximation method introduced in [16] is used to calculate the end-to-end expected distortion. In this model, the correlation coefficient between two points X and Y is assumed to be approximated by

$$\rho_{XY} = \exp(-\alpha \cdot d_{XY}) \quad (3)$$

where d_{XY} is the Euclidean distance between two decoder reconstructed pixels X and Y, and α is a constant, whose value is experimentally obtained from training data (typically 0.04 to 0.06).

In addition to pixel cross-correlations, an important, often neglected, issue in per pixel distortion estimation, is that of rounding errors. A rounding operation is usually employed whenever a filtering or averaging operation results in a floating-point pixel value. In H.264/AVC, rounding operations are encountered in sub-pixel prediction, weighted prediction, in-loop filtering, etc. Rounding can be viewed as a special case of uniform quantization with a quantization step size of one unit, and in which the quantized value is the nearest integer. In [16], a rounding error compensation (REC) technique based on quantization theory (QT-based) is proposed to deal with rounding errors in distortion estimations.

B. Packet Ordering

In this section, we present a rate-constrained scheme to order the packets in the transmission buffer of each user based on the contribution of each packet to the end-to-end expected distortion. There are multiple challenges in ordering packets in a lossy environment in conjunction with a complex error concealment strategy. First, because error concealment (EC) introduces inter packet dependencies, the ordering process cannot be done greedily and, therefore, all possible packet loss combinations have to be taken into account. This is because the selection of

the first packet causes the existing symmetry in the packet locations to break since lost packets can be better concealed if they are close to a received packet. Note that the "concealability" of a slice, strongly depends on its motion correlation with the neighboring slice as well as the reliability of the neighboring slice, i.e., the loss probability of the packet in which it belongs. In addition, the expected distortion of a frame, as discussed earlier, depends on the loss probability of the consisting packets, which in general, could be different for each packet due to the channel fading process.

To overcome the aforementioned challenges, we propose a rate-constrained scheme to order the video data packets. Let $\mu_m \in \{0,1\}$ denote whether packet m is transmitted ($\mu_m = 1$) or not ($\mu_m = 0$), during the current transmission timeslot. In order to determine the transmission policy vector $\boldsymbol{\mu} = (\mu_1, \mu_2, \ldots, \mu_M)$, a Lagrangian cost function is introduced. The Lagrangian expresses the problem of minimizing end-to-end expected distortion of the frame given a rate constraint as

$$\mathcal{L}(\boldsymbol{\mu}, \boldsymbol{\epsilon}; \lambda) = \sum_{m=1}^{M} E\{D_m(\mu_m, \epsilon_m, \mu_{m-1}, \epsilon_{m-1})\} + \lambda R_m(\mu_m) \quad (4)$$

where $\boldsymbol{\epsilon} = (\epsilon_1, \epsilon_2, \ldots, \epsilon_M)$ denotes the vector of packet loss probabilities, ϵ_m, of each packet m, and $\lambda \geq 0$ is a real parameter determining the transmission cost. $R_m(\mu_m)$ denotes the number of bits transmitted for packet m, which will be 0, if $\mu_m = 0$, and the length of the packet, if $\mu_m = 1$.

For a fixed $\boldsymbol{\epsilon}$, let the mode vector $\boldsymbol{\mu}^*$ be the one that minimizes the cost function, i.e.,

$$\boldsymbol{\mu}^*(\lambda, \boldsymbol{\epsilon}) = \arg \min_{\boldsymbol{\mu} \in \{0,1\}^M} \mathcal{L}(\boldsymbol{\mu}, \boldsymbol{\epsilon}; \lambda). \quad (5)$$

Given the error concealment technique discussed above which limits the dependencies between packets, the above optimization can be performed efficiently using a dynamic programming (DP) technique. The DP can be viewed as a shortest path problem in a trellis, where each stage corresponds to the mode (SEND or SKIP) selection for a given packet with the complexity equal to $2 \times 2 \times M$.

The frame rate $R(\boldsymbol{\mu})$ is obtained by

$$R(\boldsymbol{\mu}) = \sum_{m=1}^{M} R_m(\mu_m). \quad (6)$$

Note that the solution in (5) is optimal in the sense that, if a rate constraint R_c corresponds to λ, then the total expected distortion $E\{D\}$ is minimum for all combinations of transmission options with bit rate less than or equal to R_c.

An increase in the value of the Lagrange parameter, λ, resembles a rise in the cost of the transmission bits. Consequently, the frame rate $R(\boldsymbol{\mu}^*)$ monotonically decreases as λ increases. In other words, fewer packets will be selected in $\boldsymbol{\mu}^*$ for transmission. Furthermore, there exists some $\lambda_{\max} \geq 0$ such that $\mu_m^* = 0$ for all $m \in [1, 2, \ldots, M]$, and assuming that all packets have some contribution towards reducing the expected distortion, there exists some $\lambda_{\min} \geq 0$ such that $\mu_m^* = 1$ for all m.

Therefore, the threshold, λ_m at which the mode of a packet m switches from $\mu_m^* = 0$ to $\mu_m^* = 1$ can be obtained by sweeping λ from $\lambda_{\max}$ to $\lambda_{\min}$. Finally, the order in which each packet is added to the transmission queue is efficiently determined by the threshold λ_m, i.e., packets with larger values of λ_m correspond to more important packets in terms of reducing the expected distortion and, therefore, are transmitted first. Note that the thresholds depend on the probability of loss, ϵ, as well, and cannot be known *a priori*. In practice, a $\lambda \geq \lambda_{\max}$ is found first based on a rough estimate of $\lambda_{\max}$, then the thresholds λ_m are obtained utilizing a bilinear search.

IV. RESOURCE ALLOCATION

A. Introduction

In the previous section, we have described the proposed scheme for reordering packets within the transmission queue of each user. The current section discusses the resource allocation across users that will determine the transmission rates assigned to each user, and thereby the number of transmissible packets from each user's transmission queue.

As in [7], we consider a scheme, such as HSDPA, where a combination of TDM and CDMA is used for resource allocation. In this scheme, at each transmission timeslot, t, the scheduler can decide on the number of spreading codes, n_i, (assumed to be orthogonal) and the transmission power, p_i, that can 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 remains the same throughout this section and is omitted for simplicity). The maximum number of spreading codes that can be handled by each user is determined by the user's mobile device. However, the total number of spreading codes, N, that can be allocated to all users, is limited by the specific standard (15 for HSDPA). The total power, P, that can be used by the base station is also limited in order to restrict the possibility of interference across neighboring cells. In the case that the exact channel state at each time-slot is known to the scheduler, the achievable error-free transmission rate, r_i, for each user can be precisely calculated given the allocated resources, n_i and p_i [7]. In the case, when the exact channel state is not known, however, and only an estimate of the channel state is available, it is also necessary to consider the probability of loss in the channel due to random channel fading that may occur during the transmission. Depending on the assumed wireless channel model, the probability of loss can be calculated, using an outage probability formulation [14], as a function of the assigned transmission power, bandwidth, and transmission rate.

B. Outage Probability

Since the concept of outage probability is discussed in detail in [14], this section will simply summarize its application to the current work. Again, the time index, t will be omitted during this discussion as the outage probability will be calculated at each transmission time-slot. Also, note that ε_i refers to the probability of loss of the transmission to user i in the current time-slot. All packets, m_i, transmitted to user i during the current time-slot will have a packet loss probability, ϵ_{m_i}, equal to ε_i. Using the
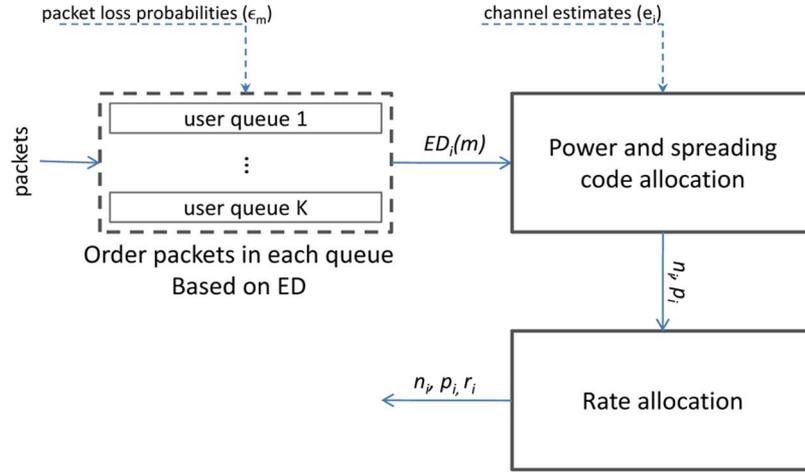
Fig. 2. Block diagram of the proposed scheduler.

model derived in [14], the probability of loss of a transmission to user i can be written as

$$
\begin{aligned}
\varepsilon_i &= \text{Prob}\left[n_i B \log\left(1 + \frac{p_i h_i}{n_i}\right) \leq r_i | e_i\right] \\
&= \text{Prob}\left[h_i \leq \frac{n_i}{p_i}\left(2^{\frac{r_i}{n_i B}} - 1\right)|e_i\right] \\
&= F_{x|e_i}(h_i | e_i)
\end{aligned}
\tag{7}
$$

where B denotes the maximum symbol rate per code, h_i denotes the instantaneous channel fading state (SINR per unit power) at that time-slot, and $F_{x|e_i}$ denotes the cumulative probability density function of the instantaneous channel fading state conditioned on the observed channel estimate, e_i. It is clear from (7) that the probability of loss, ε_i, depends on four factors: the allocated resources (n_i, p_i), the estimated channel SINR (e_i), the assigned transmission rate (r_i), and the conditional cumulative density function (cdf) given by the wireless channel model $(F_{x|e_i})$.

C. Wireless Channel Model

This work assumes that only partial (imperfect) channel state information is available at the scheduler/transmitter. Errors in the channel estimate can arise from the delay in the feedback channel combined with Doppler spread and quantization errors. It is possible to empirically determine the conditional cdf of the channel SINR conditioned on the channel estimate and the feedback delay using channel measurements. For the purposes of this work, we employ a Nakagami-m channel model which exhibits similar patterns to HSDPA RF channel traces obtained from Motorola, Inc. In this model, the channel SINR can be modeled as a gamma distribution with mean at the channel estimate, e_i. The cumulative probability density function can be written as

$$
F_{x|e_i}(h_i) = \frac{\gamma\left(m, \frac{mh_i}{e_i}\right)}{\Gamma(m)}
\tag{8}
$$

where m is a shape parameter determined by the order, m, of the distribution, $\gamma()$ denotes the incomplete gamma function,

and $\Gamma(m)$ denotes the gamma function of order m. Note that for a fixed order, m, the variance of the Nakagami-m distribution increases with increasing mean (i.e., channel estimate).

D. Problem Formulation

Given the packet ordering scheme and method for calculating the loss probability described above, the scheduler jointly optimizes the rate assignment, $\boldsymbol{r} = (r_1, r_2, \ldots, r_K)$, where K is the number of users, the power assignment, $\boldsymbol{p} = (p_1, p_2, \ldots, p_K)$, and the spreading code assignment, $\boldsymbol{n} = (n_1, n_2, \ldots, n_K)$, in order to minimize the total expected distortion in the system at each time slot. For a given rate and packet loss probability, let the expected distortion of the frame currently being transmitted to user i given the packet ordering specified in Section III-B be $E\{D_i(r_i, \epsilon_i)\}$, obtained as in (2). Then, the optimization problem can be written as

$$
\min_{\boldsymbol{n}, \boldsymbol{p}, \boldsymbol{r}} \sum_{i=1}^{K} E\{D_i[r_i, \varepsilon_i(n_i, p_i, r_i, e_i)]\}
\tag{9}
$$

such that

$$
0 \leq \sum_{i=1}^{K} n_i \leq N, 0 \leq n_i \leq N_i, \quad \forall i
\tag{10}
$$

$$
0 \leq \sum_{i=1}^{K} p_i \leq P, \quad \forall i
\tag{11}
$$

and

$$
0 \leq \frac{p_i e_i}{n_i} \leq \hat{S}_i, \quad \forall i
\tag{12}
$$

where $\hat{S}_i$ is the maximum SINR constraint [6] and all other parameters here are previously defined. In principle, a nonlinear optimization scheme can be used to find the solution to (9). In practice, however, the solution can be highly complex, as an analytical form for $E\{D_i\}$, which will satisfy different video content and channel conditions, cannot be easily derived. Therefore, this paper uses a two-step approach to simplify the solution to the problem.

Our solution is based on a few observations. One observation is that the packet ordering arrived at by the technique described

TABLE I
SYSTEM PARAMETERS USED IN SIMULATIONS

N	N_i	P	$\hat{S}_i$
15	5	25W	1.8dB

in Section III.B is not overly sensitive to the probability of loss of each packet. Another observation is that, due to the constraints on transmission power and bandwidth imposed by the system as well as the limited length of a time-slot, the number of bits that can be transmitted to a user at any given transmission opportunity is limited. Therefore, as a first step, we fix the probability of loss of each packet in the transmission queue of each user to a reasonable value, denoted by $\tilde{\varepsilon}_i$. Then, we use a linear approximation to $E\{D_i[r_i, \tilde{\varepsilon}_i]\}$ over the limited number of bits that might be transmitted. Now, given, the probability of loss, $\tilde{\varepsilon}_i$, and channel estimate, e_i, the rate assignment, r_i, must be a function of n_i and p_i as specified in (7). Therefore, for the fixed $\tilde{\varepsilon}_i$, the problem of determining

$$V(\boldsymbol{n}^*, \boldsymbol{p}^*) = \max_{\boldsymbol{n},\boldsymbol{p}} \sum_{i=1}^{K} -\frac{\partial E\{D_i[r_i(n_i, p_i, \tilde{\varepsilon}_i), \tilde{\varepsilon}_i]\}}{\partial r_i}$$
$$\cdot r_i(n_i, p_i, \tilde{\varepsilon}_i) \quad (13)$$

can be solved subject to the constraints in (10), (11), and (12) where, $\partial/\partial r_i$ denotes the partial derivative with respect to r_i. Note that the gradient of $E\{D_i\}$ with respect to r_i for a fixed probability of loss can be numerically calculated using the methods described in Section III-B, and the formulation described in [7]. The solution to the type of problem in (13) can be found in [6].

For the second step, it can be observed from (7) that when n_i and p_i are fixed, then ε_i is a function of only r_i (i.e., ε_i increases as r_i increases). As a consequence of this relationship, when r_i is small, increasing r_i leads to a lower $E\{D_i\}$ as it results in a larger number of transmitted bits. As r_i increases, however, ε_i also increases leading to a higher $E\{D_i\}$ as the transmitted bits are no longer reliable. Therefore, $E\{D_i\}$ is typically a convex function of r_i. Since there is no multiuser constraint on the r_i assignment for a given user, the following convex optimization problem can then be solved separately for each user i with a simple 1-D line search to find the optimal value of r_i which leads to the minimum expected distortion for user i

$$\min_{r_i} E\{D_i[r_i, \epsilon_i(n_i^*, p_i^*, r_i, e_i)]\} \quad (14)$$

where n_i^* and p_i^* are the values of n_i and p_i found by solving (13). Fig. 2 demonstrates these procedures.

V. EXPERIMENTAL RESULTS

Seven video sequences with varied content (foreman, carphone, mother and daughter, news, hall monitor, silent, and stefan), in QCIF (176×144) format were used at a rate of 30 fps for the simulations. Therefore, the packets of each frame had a total of 33 ms to be transmitted on time to be played back at the decoder. The video sequences were encoded in H.264 (JVT reference software, JM 10.2 [20]) at variable bit rates (VBR) to obtain a decoded PSNR of 35 dB at each frame. Nonetheless,
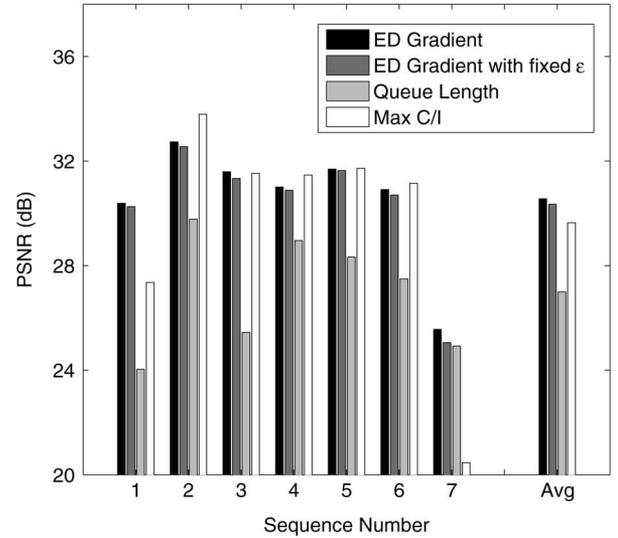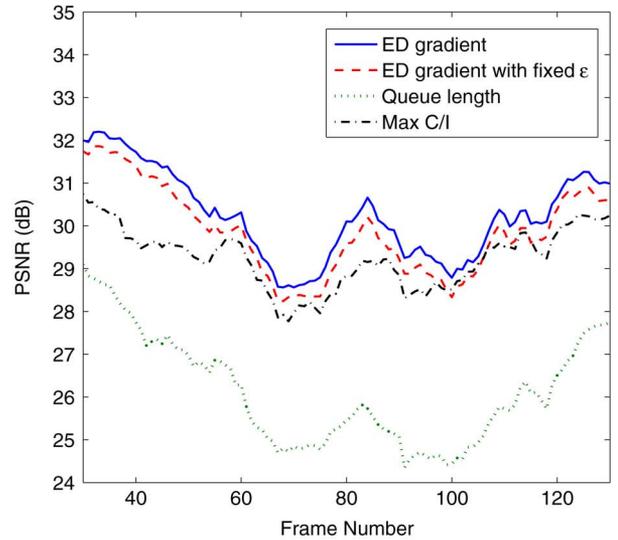
Fig. 3. Average received PSNR. (a) average quality for each user, (b) average quality over all users.

our optimization framework does not depend on the encoding configuration, and can equally apply to CBR encoding. All frames except the first one were encoded as P frames. To reduce error propagation due to packet losses, 15 random I MBs were inserted into each frame, and constrained intra prediction was used at the encoder. The frames were packetized such that each slice contained one row of MBs, which enabled a good balance between error robustness, and compression efficiency.

The wireless network was modeled as an HSDPA system. The system parameters used in the simulations are shown in Table I. HSDPA provides 2 ms transmission time slots. A Nakagami channel with shaping parameter $m = 10$ is considered for the channel model.

We allowed larger application layer packets to be fragmented into smaller packets at the MAC layer prior to scheduling. It was assumed that all fragments of an application layer packet must be received at the decoder in order for it to be correctly
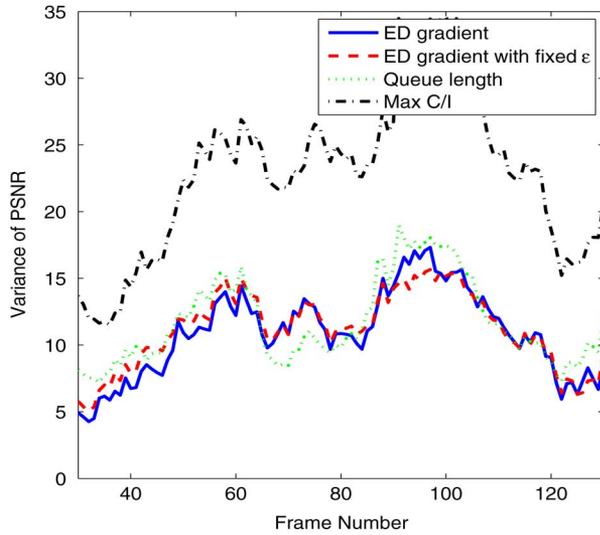
Fig. 4. Variance of received PSNR.

decoded. An ACK/NACK feedback for transmitted packet fragments was assumed to be available with a feedback delay of 10 ms. Therefore, if a NACK is received for a fragment of a transmitted application layer packet whose decoding deadline has not yet expired, then the packet will be reinserted into the transmission queue and re-ordered based on the current state of the queue.

The simulations compare four different methods for determining the resource allocation. They are as follows.

1) Expected Distortion Gradient—This is the proposed content-aware method as described in Section IV.
2) Expected Distortion Gradient with Fixed Loss—In this method, packet ordering is performed using the expected distortion as specified in Section III.B, but in the resource allocation, the probability of loss, ε_i is fixed for all users. Essentially, this method eliminates the second step of the solution in Section IV and, thus, is less computationally complex than the first.
3) Queue Length—This method is not content-aware and uses the queue lengths at each user's transmission buffer [21] to determine the resource allocation. As in the second method, this also assumes a fixed ε_i for all users. The main difference between this method and the second is that in this method, the packets are not ordered according to their expected distortion gradients.
4) Max C/I—This method takes advantage of the variations in the radio channel conditions and always chooses to serve the user experiencing the best channel condition, maximizing the system throughput. The same loss probability of method 2 and 3 is also used here.

Fig. 3 shows the average quality of the received video after scheduling and transmission over a packet lossy network, using the four different schemes. In Fig. 3(a), the results are averaged over each video sequence and also 50 channel realizations. Fig. 3(b) shows the average quality at each frame averaged over all the users and channel realizations. For the fixed loss schemes, ε_i is fixed at the average loss probability obtained from the first scheme which is about 0.1 in our simulation setup. The figures
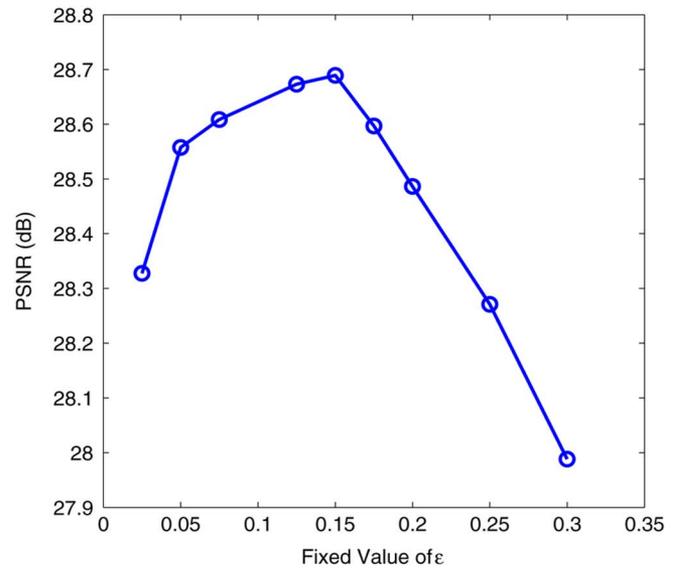
(a)
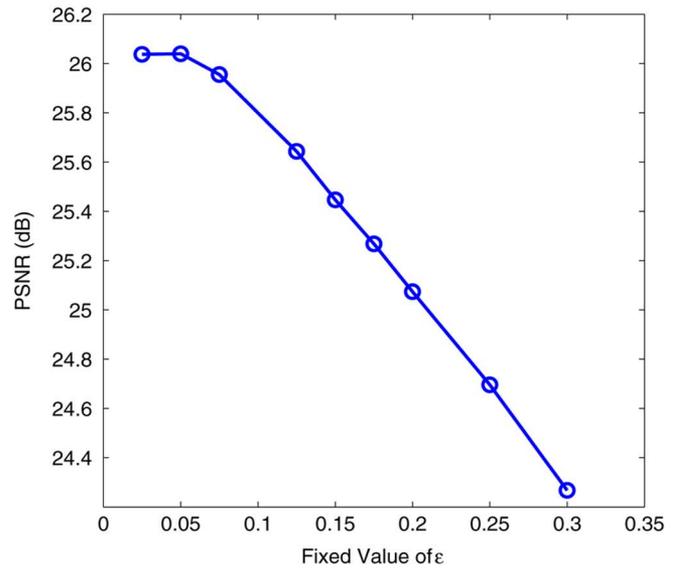
(b)

Fig. 5. Sensitivity of received quality to choice of ε_i. (a) Content-aware scheme (method 2), (b) Queue length scheme.

show that the proposed content-dependent schemes significantly outperform the queue-length dependent and max C/I scheme in terms of average received quality.

Fig. 4 shows the variance of quality at each video frame across all users and channel realizations. The max C/I scheme shows a significantly larger variance across users than the others. Our proposed methods on the other hand, exhibit the lowest variance across users. These results can be attributed partly to the packet ordering and also to the fact that the queue length dependent scheme does not consider the concealability of video packets when allocating resources across users. Therefore, assuming two users have equal queue lengths, the user whose video packets are difficult to conceal if lost, will not be given priority over the other user.

In the previous figures we chose the fixed loss probability for schemes 2, 3, and 4 to match the average loss probability

in scheme 1. Next, we study how sensitive these schemes are to the value of ε_i. Fig. 5 shows the variation in the average received PSNR as the value of ε_i is varied for the two schemes that use a fixed probability of loss. Fig. 5(a) shows the results for the content-aware scheme, and it is apparent that the overall video quality remains within a 0.5 dB range over a large range of ε. This result shows that the choice of ε does not significantly affect the performance of the system for the content-aware case. Fig. 5(b) shows the results for the queue length scheme. In this case, the choice of ε has a greater impact on the average received PSNR.

VI. CONCLUSION

This work introduces a content-aware multiuser resource allocation and packet scheduling scheme that can be used in wireless networks where only imperfect channel state information is available at the scheduler. The scheme works by jointly optimizing the resource allocation and transmission rate allocation in a content-aware manner while also prioritizing video packets in the transmission queue. The content dependent techniques shown in this paper significantly outperform a conventional content-independent scheduling scheme. While results comparing CBR encoded content are not shown in this paper, with CBR encoded content, potentially greater improvements can be expected with a content dependent scheme because, unless an ideal rate control scheme is used, the bit allocations for the video sequences will be less correlated with the final decoded video quality. A simplified content dependent technique that fixes the probability of loss is also shown. Although the scheme with fixed probability of loss can achieve similar results to one that optimizes the probability of loss, it requires tuning of the probability of loss parameter, whose optimal value cannot be known without knowledge of the video content and channel conditions.

REFERENCES

[1] *High Speed Downlink Packet Access; Overall Description*, 3GPP Std. TS 25.308 v7.0.0, 2006.

[2] *IEEE Standard for Local and Metropolitan Area Networks; Part 16: Air Interface for Fixed and Mobile Broadband Wireless Access Systems*, IEEE Std. 802.16e, 2005.

[3] S. Lu, V. Bharghavan, and R. Srikant, "Fair scheduling in wireless packet networks," *ACM SIGCOMM Comput. Commun. Rev.*, vol. 27, no. 3, Oct. 1997.

[4] R. Agrawal, R. L. A. Bedekar, and V. Subramanian, "A class and channel-condition based weighted proportionally fair scheduler," presented at the ITC, Sep. 2001.

[5] S. Shakkottai, R. Srikant, and A. Stolyar, "Pathwise optimality and state space collapse for the exponential rule," presented at the IEEE Int. Symp. Information Theory, Jul. 2002.

[6] R. Agrawal, V. Subramanian, and R. Berry, "Joint scheduling and resource allocation in CDMA systems," *IEEE Trans. Inf. Theory*, to be published.

[7] P. Pahalawatta, R. Berry, T. Pappas, and A. Katsaggelos, "A content-aware scheduling scheme for video streaming to multiple users over wireless networks," presented at the Eur. Signal Processing Conf., Florence, Italy, Sep. 2006.

[8] G. Liebl, T. Stockhammer, C. Buchner, and A. Klein, "Radio link buffer management and scheduling for video streaming over wireless shared channels," presented at the Packet Video Workshop, 2004.

[9] Y. Ofuji, S. Abeta, and M. Sawahashi, "Unified packet scheduling method considering delay requirement in forward link broadband wireless access," in *Proc. Vehicular Technology Conf.*, Oct. 2003, pp. 1603–1607.

[10] 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. Personal, Indoor and Mobile Radio Communications PIMRC*, Sep. 2004, vol. 3, pp. 2202–2206.

[11] D. Kim, B. Ryu, and C. Kang, "Packet scheduling scheme for real time video traffic in wcdma downlink," in *Proc. 7th CDMA Int. Conf.*, Oct. 2002, vol. 2524, pp. 285–293.

[12] P. Chou and Z. Miao, "Rate distortion optimized streaming of packetized media," *IEEE Trans. Multimedia*, vol. 8, no. 2, pp. 390–404, Apr. 2006.

[13] P. Pahalawatta, R. Berry, T. Pappas, and A. Katsaggelos, "Content-aware resource allocation and packet scheduling for video transmission over wireles networks," *IEEE J. Sel. Areas Commun.*, pp. 749–759, May 2007.

[14] L. Ozarow, S. Shamai, and A. Wyner, "Information theoretic considerations for cellular mobile radio," *IEEE Trans. Veh. Technol.*, vol. 43, no. 2, pp. 359–378, Feb. 1994.

[15] R. Zhang, S. L. Regunathan, and K. Rose, "Video coding with optimal inter/intra-mode switching for packet loss resilience," *IEEE Trans. Commun.*, vol. 18, no. 6, pp. 966–976, Jun. 2000.

[16] H. Yang and K. Rose, "Advances in recursive per-pixel estimation of end-to-end distortion for application in H.264," in *Proc. Int. Conf. Image Processing*, Genova, Italy, Sep. 2005.

[17] D. Wu, Y. T. Hou, B. Li, W. Zhu, Y. Q. Zhang, and H. J. Chao, "An end-to-end approach for optimal mode selection in internet video communication: Theory and application," *IEEE Trans. Commun.*, vol. 18, no. 6, pp. 977–995, Jun. 2002.

[18] Y. Wang, G. Wen, S. Wenger, and A. K. Katsaggelos, "Review of error resilient techniques for video communications," *IEEE Signal Process. Mag.*, vol. 17, pp. 61–82, July. 2000.

[19] T. Wiegand, G. J. Sullivan, G. Bjntegaard, and A. Luthra, "Overview of the H.264/AVC video coding standard," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 13, no. 7, pp. 560–576, Jul. 2003.

[20] H.264 JM 10.2 reference software [Online]. Available: http://iphome.hhi.de/suehring/tml/download/

[21] M. Andrews, K. Kumaran, K. Ramanan, A. Stolyar, P. Whiting, and R. Vijayakumar, "Providing quality of service over a shared wireless link," *IEEE Commun. Mag.*, vol. 39, no. 2, pp. 150–154, Feb. 2001.

Ehsan Maani (S'05) received the B.S. degree in physics from the Sharif University of Technology, Tehran, Iran, in 2004, and the M.Sc. degree in electrical engineering from Northwestern University, Evanston, IL, in 2007. He is currently pursuing the Ph.D. degree at the School of Electrical Engineering and Computer Science, Northwestern University.

His research interests include image/video compression and transmission, indexing, and retrieval.

Peshala V. Pahalawatta (S'97–M'08) received the B.Sc. degree in electrical engineering from Lafayette College, Easton, PA, in 2000, and the M.Sc., and Ph.D. degrees in electrical engineering from Northwestern University, Evanston, IL, in 2002, and 2007, respectively.

He is currently a Senior Research Engineer with the Image Technology group at Dolby Laboratories, Inc., Burbank, CA. His primary research interests include image and video compression and transmission, image and video quality evaluation, wireless communications, and computer vision.

Randall Berry (S'93–M'00) received the B.S. degree in electrical engineering from the University of Missouri, Rolla, in 1993, and the M.S. and Ph.D. degrees in electrical engineering and computer science from the Massachusetts Institute of Technology (MIT), Cambridge, in 1996 and 2000, respectively.

Since 2000, he has been on the faculty of Northwestern University, Evanston, IL, where he is currently an Associate Professor in the Department of Electrical Engineering and Computer Science. He was on the technical staff at MIT Lincoln Laboratory in 1998. His primary research interests include wireless communication, data networks, and information theory.

Dr. Berry is the recipient of a 2003 National Science Foundation CAREER award. He is currently an Associate Editor for the IEEE Transactions on Wireless Communications and the IEEE Transactions on Information Theory.

Thrasyvoulos N. Pappas (M'87–SM'95–F'06) received the B.S., M.S., and Ph.D. degrees in electrical engineering and computer science from the Massachusetts Institute of Technology, Cambridge, in 1979, 1982, and 1987, respectively.

From 1987 to 1999, he was a Member of the Technical Staff at Bell Laboratories, Murray Hill, NJ. In 1999, he joined the Department of Electrical and Computer Engineering (now EECS), Northwestern University, Evanston, IL, as an Associate Professor. His research interests are in image and video quality and compression, perceptual models for image processing, image and video analysis, model-based halftoning, and multimedia signal processing.

Dr. Pappas has served as an elected member of the Board of Governors of the Signal Processing Society of IEEE (2004–2007), Chair of the IEEE Image and Multidimensional Signal Processing Technical Committee, Associate Editor of the IEEE Transactions on Image Processing, and technical program Co-Chair of ICIP 2001 and the Symposium on Information Processing in Sensor Networks (IPSN 2004). He is a Fellow of SPIE. Since 1997, he has been Co-Chair of the SPIE/IS&T Conference on Human Vision and Electronic Imaging. He has also served as Co-Chair of the 2005 SPIE/IS&T Electronic Imaging Symposium.

Aggelos K. Katsaggelos (S'80–M'85–SM'92–F'98) received the Diploma degree in electrical and mechanical engineering from the Aristotelian University of Thessaloniki, Greece, in 1979, and the M.S. and Ph.D. degrees in electrical engineering from the Georgia Institute of Technology, Atlanta, in 1981 and 1985, respectively.

In 1985, he joined the Department of Electrical and Computer Engineering, Northwestern University, Evanston, IL, where he is currently a Professor. He was the holder of the Ameritech Chair of Information Technology (1997–2003). He is also the Director of the Motorola Center for Seamless Communications and a member of the Academic Affiliate Staff, Department of Medicine, Evanston Hospital. He has published extensively in the areas of signal processing, multimedia transmission, and computer vision. He is the editor of *Digital Image Restoration* (Springer-Verlag, 1991), co-author of *Rate-Distortion Based Video Compression* (Kluwer, 1997), co-editor of *Recovery Techniques for Image and Video Compression and Transmission* (Kluwer, 1998), and co-author of *Super-resolution for Images and Video* (Claypool, 2007) and *Joint Source-Channel Video Transmission* (Claypool, 2007). He is the co-inventor of twelve international patents.

Dr. Katsaggelos has served the IEEE and other Professional Societies in many capacities; for example, he was Editor-in-Chief of the *IEEE Signal Processing Magazine* (1997–2002), a member of the Board of Governors of the IEEE Signal Processing Society (1999–2001), and a member of the Publication Board of the IEEE Proceedings (2003–2007). He is the recipient of the IEEE Third Millennium Medal (2000), the IEEE Signal Processing Society Meritorious Service Award (2001), an IEEE Signal Processing Society Best Paper Award (2001), an IEEE International Conference on Multimedia and Expo Paper Award (2006), and an IEEE International Conference on Image Processing Paper Award (2007). He is a Distinguished Lecturer of the IEEE Signal Processing Society (2007–2008).