

RESOURCE ALLOCATION FOR DOWNLINK MULTIUSER VIDEO TRANSMISSION OVER WIRELESS LOSSY NETWORKS

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

Northwestern University
Electrical Engineering and Computer Science Department
2145 Sheridan Rd, Evanston, IL 60208

ABSTRACT

The emergence of 3G and 4G wireless networks brings with it the possibility of streaming high quality video content on-demand to mobile users. Wireless video applications require appropriate scheduling techniques that make use of the specific characteristics of video content, as well as the well known gains from multiuser diversity. While fast and frequent channel feedback is available in the new generation of wireless networks, the channel estimates cannot be perfect, and channel losses should be taken into account in the packet scheduling and resource allocation. The proposed scheme is formulated as a joint optimization over the resource allocation and channel loss protection, in order to minimize the distortion of the received video sequences. The distortion is a function of the packets deliberately dropped at the transmission queue due to congestion, as well as of random channel losses. The scheme makes use of a packet prioritization strategy that orders video packets based on their contribution to reducing the expected distortion of the received video sequence. Simulation results show that the proposed technique significantly outperforms content-independent packet scheduling schemes.

Index Terms— Resource management, Multimedia communication, Multiuser channels, Mobile communication

1. INTRODUCTION

The technology available for on-demand video streaming to mobile clients over wireless networks is rapidly improving with the emergence of third generation and newer standards such as HSDPA and IEEE 802.16 [1, 2]. In this paper, we consider the problem of streaming multiple pre-encoded video sequences to multiple users over a wireless network. To best use the available technology and serve a large number of clients with a high quality of service, smart scheduling techniques need to be employed at the wireless server.

It is well accepted that multiuser diversity can be exploited to maximize the data throughput in wireless networks [3]. 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 [4]. Another important feature 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 [4, 5].

This research was funded in part by the Motorola Center for Seamless Communications at Northwestern University

In this context, we have shown in [4] that gradient-based scheduling techniques [6] can be used, in which a weighted sum of per-user data rates is maximized. The weights are determined by the gradients of a content-aware utility function. The approach in [4], 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. Therefore, the losses considered in [4] occur only when packets are not transmitted on time due to the scheduling priority. In the current work, we consider a realistic scenario in which only an imperfect estimate of the channel state is available. 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 [7]. 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 can use the expected quality at the receiver in order to determine its scheduling decisions. Efficient methods exist for recursively calculating the expected distortion at the receiver [8, 9]. 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 overall expected distortion over all users in the system.

The rest of the paper is organized as follows. In Sec. 2, we provide a brief overview of the system. In Sec. 3, 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 Sec. 4 and a simplified solution to the problem is proposed. Some simulation results are shown in Sec. 5 and conclusions are drawn in Sec. 6.

2. SYSTEM OVERVIEW

Figure 1 highlights some of the major conceptual components of the system under consideration. The 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, 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 sched-

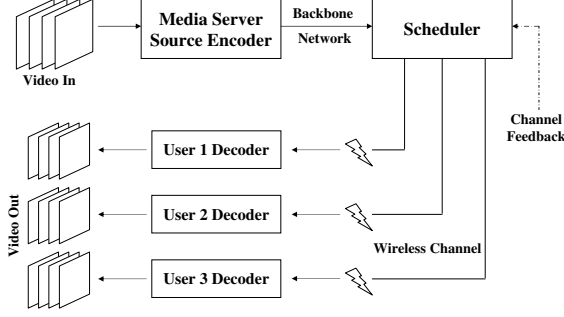


Fig. 1. Block diagram of multiuser downlink video streaming system.

uler 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. We can assume that all the packets in the frame have the same decoding deadline. Note that any packet left in the transmission queue after its decoding deadline has expired is discarded since it has lost its value to the decoder.

After transmission, we assume that any packet that does not arrive error free, and in time for playback, is lost in the channel. Multiple retransmissions of a packet based on feedback are allowed as long as the retransmissions occur prior to the decoding deadline. 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.

3. PROBLEM FORMULATION

3.1. Error Concealment

A robust error concealment technique helps avoid significant visible errors in the reconstructed frames at the decoder. To perform error concealment we use an algorithm similar to the one implemented in the H.26L test model [10]. In this algorithm, all correctly received slices of a picture are decoded first, and then the lost slices are concealed. A record of the status of each macroblock is kept, where each MB can be one of the following, Correctly received, Lost or Concealed. The concealment starts by processing MB columns at the frame boundaries and then moves inwards MB column by MB column. The algorithm for inter-coded frames tries to estimate the motion in the missing MB by prediction schemes from available motion information of spatial neighbors. In this work, since causality plays an important role in our optimization algorithm, only the data from the correctly received MBs above the Lost MB are used for concealment. Note that this concealment strategy is employed both in the scheduler optimization framework and at the decoder.

3.2. End-to-End Expected Distortion

Due to channel losses, we use the expected end-to-end distortion to evaluate video quality. Three factors can be identified as effect-

ing the end-to-end distortion: the source behavior (quantization and packetization), the channel characteristics, and the receiver behavior (error concealment) [11, 12]. The expected distortion of the m^{th} packet 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 for the m^{th} packet, $E[D_{R,m}]$ is the expected distortion if the m^{th} packet is received, and $E[D_{LR,m}]$ and $E[D_{LL,m}]$ are respectively the expected distortion of the m^{th} packet after concealment when packet $(m-1)$ is received or lost. The distortion measurement is based on a per pixel recursive algorithm called ROPE, which was originally proposed in [8] 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 decoder reconstructed pixels; however, sub-pixel prediction employed in H.264 involves interpolation of neighboring pixels, 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 [9] is used to calculate end-to-end expected distortion.

3.3. Packet Ordering

The main contribution of this paper is the development of a method to order the packets in the transmission buffer for each user based on the importance of each packet in reducing the end-to-end expected distortion. Note that due to the dependencies among the packets introduced by error concealment, the ordering packets is a challenging task. In addition, the expected distortion of a frame depends on the loss probability of the contained packets that can be different for each packet due to channel fading.

Let us assume that each packet $m \in [1, \dots, M]$ has a mode $\mu_m \in \{0, 1\}$ where 0 denotes *skip* and 1 denotes *send*. A skipped packet m is concealed using the motion information of its top macroblock row contained in packet $m-1$ if the packet $m-1$ is available ($\mu_{m-1} = 1$). Otherwise, the sample values of the previous frame at the same location are copied. Let $R_m(\mu_m)$ be the rate of the m^{th} packet which is equal to 0 when $\mu_m = 0$. Furthermore, we define the Lagrangian function $\mathcal{L}(\boldsymbol{\mu}, \boldsymbol{\epsilon}; \lambda)$ as,

$$\mathcal{L}(\boldsymbol{\mu}, \boldsymbol{\epsilon}; \lambda) = \sum_{m=1}^M E[D_m(\mu_m, \epsilon_m, \mu_{m-1}, \epsilon_{m-1})] + \lambda R(\boldsymbol{\mu}_m), \quad (2)$$

where $\boldsymbol{\mu}$ is the vector of all packet modes, i.e., $\boldsymbol{\mu} = (\mu_1, \dots, \mu_M)$, $\boldsymbol{\epsilon}$ is the vector of packet loss probabilities, i.e., $\boldsymbol{\epsilon} = (\epsilon_1, \dots, \epsilon_M)$, and λ is a real parameter. The mode vector $\boldsymbol{\mu}^*$ is defined to be the minimum of the Lagrangian function, i.e.,

$$\boldsymbol{\mu}^*(\lambda, \boldsymbol{\epsilon}) = \arg \min_{\boldsymbol{\mu} \in \{0,1\}^M} \mathcal{L}(\boldsymbol{\mu}, \boldsymbol{\epsilon}; \lambda). \quad (3)$$

Assuming that the Lagrangian function \mathcal{L} is convex for $\lambda \geq 0$, for all $m \in [1, \dots, M]$, there exists a threshold $\lambda_m > 0$ such that $\mu_m^* = 1$ for $\lambda \geq \lambda_m$ and $\mu_m^* = 0$ otherwise. Since a packet with a larger threshold makes a larger contribution to the expected video quality per bit, the thresholds $\lambda_1, \dots, \lambda_M$ are used to determine the order of the packets in the transmitter buffer. Note that due to the convexity assumption there is a $\lambda_{max} > \lambda_m$ such that $\mu_m^* = 0$ for all m . Also note that the thresholds depend on $\boldsymbol{\epsilon}$, and cannot be known *a priori*. To find the thresholds, given a fixed $\boldsymbol{\epsilon}$, a simple bisection search can be effectively employed.

3.4. Wireless Channel

We consider a scheme where a combination of TDM and CDMA is used, in which at a given time interval, t , the scheduler can decide on the number of spreading codes, n_i , that can be used to transmit to a specific user, i . Note that $n_i = 0$ indicates that the user is not scheduled for transmission in that particular time interval. The maximum number of spreading codes that can be handled by each user N_i is determined by the users mobile device. However, the total number of spreading codes, N , that can be allocated to users, is limited by the specific standard. In addition to the number of spreading codes, the transmission power level for a user p_i is determined by the scheduler. In order to limit the interference across neighboring cells, the total power used by the base station is limited to some maximum P .

In this work, it is assumed that the channel state e_i (for user i) is estimated at each time interval based on channel quality feedback available in the system. The value of e_i represents an estimate of the normalized *Signal to Interference Noise Ratio* (SINR) per unit power and can vary quite significantly over time.

For the purposes of this paper, we assume that the channel can be modeled as a Nakagami- m fading channel with a mean of e_i . If h_i is a random variable representing the channel fading for user i at some time interval, then, the probability ϵ_i that a packet of rate r_i is lost in the channel is obtained by,

$$\epsilon_i = \Pr\{C_i(h_i p_i, n_i) \leq r_i\}, \quad (4)$$

where $C_i(x_i, n_i)$ is the Shannon capacity of the channel for user i with a received power of x_i per code and a spreading code number of n_i , i.e.,

$$C_i(x_i, n_i) = n_i B \log_2(1 + \zeta_u \frac{x_i}{n_i}), \quad (5)$$

where $\zeta_u \in [0, 1)$ represents a scaling factor and determines the *gap from capacity* for a realistic system, and B is the maximum symbol rate which is constant. Consequently, from equation (4) and the cumulative distribution function (cdf) of the Nakagami- m distribution with a mean of e_i , the probability of packet loss ϵ_i is,

$$\epsilon_i(n_i, p_i, r_i, e_i) = \frac{\gamma(m, \frac{m y}{e_i})}{\Gamma(m)}, \quad (6)$$

where γ and Γ are the lower incomplete gamma function and gamma function with parameter m , respectively, and y is a function of n_i, p_i, r_i , such that,

$$C_i(y p_i, n_i) = r_i. \quad (7)$$

4. RESOURCE ALLOCATION

Given the formulation described above, the scheduler jointly optimizes the rate assignment, $\mathbf{r} = (r_1, r_2, \dots, r_K)$, where K is the number of users, the power assignment, $\mathbf{p} = (p_1, p_2, \dots, p_K)$, and the spreading code assignment, $\mathbf{n} = (n_1, n_2, \dots, n_K)$, in order to minimize the expected distortion in the system at each time slot. Let the expected distortion of the frame currently being transmitted to user i given the packet ordering specified in Sec. 3.3 be ED_i , where, from (1), $ED_i = \sum_{m=1}^M E[D_m]$. Then, the optimization problem can be written as,

$$\min_{\mathbf{n}, \mathbf{p}, \mathbf{r}} \sum_{i=1}^K ED_i[r_i, \epsilon_i(n_i, p_i, r_i, e_i)], \quad (8)$$

such that,

$$0 \leq \sum_{i=1}^K n_i \leq N, \quad 0 \leq n_i \leq N_i, \quad \forall i, \quad (9)$$

$$0 \leq \sum_{i=1}^K p_i \leq P, \quad \forall i, \quad (10)$$

and,

$$0 \leq \frac{p_i e_i}{n_i} \leq \hat{S}_i, \quad \forall i. \quad (11)$$

In (11), \hat{S}_i is a maximum SINR constraint [6]. The solution to (8), however, is not trivial, as an analytical form for ED_i , which will satisfy different video content and channel conditions, cannot be easily derived. Therefore, we propose a two-step approach to tackling the problem.

As a first step, we observe that for a given probability of loss, $\tilde{\epsilon}_i$, and channel estimate, e_i , the rate assignment, r_i , must be a function of n_i and p_i as specified in (4). To further simplify the implementation of this step, we linearize ED_i , and then, for the fixed value of $\tilde{\epsilon}_i$, we solve,

$$\max_{\mathbf{n}, \mathbf{p}} \sum_{i=1}^K -\frac{\partial}{\partial r_i} ED_i[r_i(n_i, p_i, \tilde{\epsilon}_i, \tilde{\epsilon}_i) \cdot r_i(n_i, p_i, \tilde{\epsilon}_i)], \quad (12)$$

subject to the constraints in (9), (10), and (11). Here, $\partial/\partial r_i$ denotes the partial derivative with respect to r_i . Note that the gradient of ED_i with respect to r_i for a fixed probability of loss can be numerically calculated using the methods described in Sec. 3.3 and our previous work [4]. A solution to the type of problem in (12) can be found in [6].

For the second step, we observe that when n_i and p_i are fixed, then ϵ_i is a function of only r_i , and that ED_i becomes a convex function of r_i . Since there is no multiuser constraint on the r_i assignment for a given user, we can solve the following convex optimization problem separately for each user i with a simple one-dimensional line search.

$$\min_{r_i} ED_i[r_i, \epsilon_i(\hat{n}_i, \hat{p}_i, r_i, e_i)], \quad (13)$$

where \hat{n}_i and \hat{p}_i are the values of n_i and p_i found in the solution to (12).

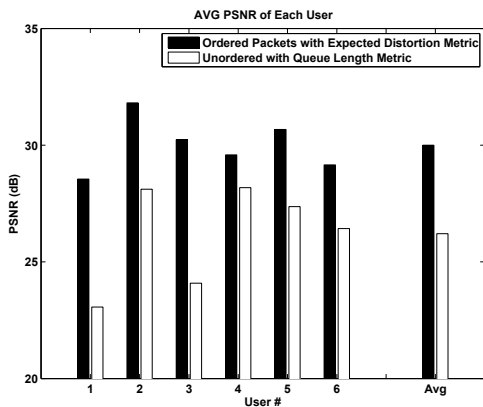
5. SIMULATION RESULTS

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 video sequences were encoded in H.264 (JVT reference software, JM 10.2) at variable bit rates to obtain a decoded PSNR of 35dB at each frame. All frames except the first were encoded as P frames. To reduce error propagation due to packet losses, 15 random INTRA 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.

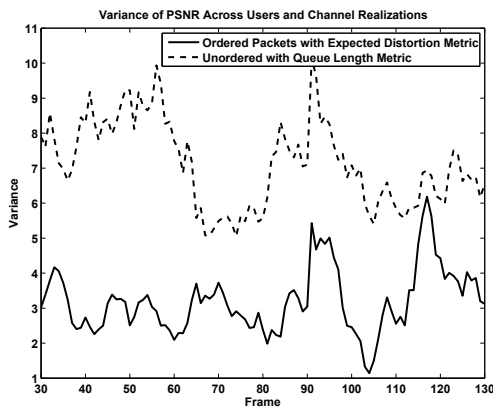
The wireless network was modeled as an HSDPA system. The system parameters were set such that $P = 10W$, $N = 15$, which is the actual total number of spreading codes available in HSDPA, and $N_i = 5$ for each user. \hat{S}_i was set at 1.8dB for each user. A Nakagami- m channel with $m = 10$ was assumed. HSDPA provides 2 msec transmission time slots. We also assumed that an ACK/NACK feedback for transmitted packets was available with a feedback delay of 10 msec. Realistic channel traces for an HSDPA system were obtained using a proprietary channel simulator developed at Motorola Inc.

Figure 2(a) shows the average quality of the received video of each user, after scheduling and transmission over a packet lossy network, using the proposed content-aware technique with the optimal

packet ordering. The results are averaged over each video sequence and 5 channel realizations. The technique is compared to a queue length dependent scheduling technique [13] with the transmission rates, r_u , assigned such that ϵ_u is fixed at 0.1 for all users (Note that the value of 0.1 was used since it showed the best performance). The figure shows that the proposed scheme significantly outperforms the queue-length dependent scheme in terms of average received quality. Figure 2(b) shows the variance of the quality at each frame over the users and channel realizations. Again, the queue length dependent scheme shows a larger variation in quality. 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 who receives video packets, that are difficult to conceal if lost, will not be given priority over the other user.



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



(b) Variance of PSNR across all users calculated over 5 channel realizations

Fig. 2. 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.

6. CONCLUSIONS

In this paper, we introduced a content-aware multi-user resource allocation and packet scheduling scheme that can be used in wireless

networks where imperfect channel state information is available at the scheduler. The scheme works by jointly optimizing the resource allocation and channel error protection in a content-aware manner while also prioritizing video packets in the transmission queue. We have shown that the scheme significantly outperforms a conventional content-independent scheduling scheme. A potential for future work is to consider the fast hybrid ARQ schemes available in the newer wireless communications systems in order to further improve the performance of the system.

7. 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. Shakkottai, T. Rappaport, and P. Karlsson, "Cross-layer design for wireless networks," *IEEE Commun. Mag.*, vol. 41, no. 10, pp. 74–80, Oct. 2003.
- [4] P. Pahalawatta, R. Berry, T. Pappas, and A. Katsaggelos, "A content-aware scheduling scheme for video streaming to multiple users over wireless networks," in *Proc. European Signal Processing Conference*, Florence, Italy, Sept. 2006.
- [5] G. Liebl, T. Stockhammer, C. Buchner, and A. Klein, "Radio link buffer management and scheduling for video streaming over wireless shared channels," in *Proc. Packet Video Workshop*, 2004.
- [6] R. Agrawal, V. Subramanian, and R. Berry, "Joint scheduling and resource allocation in CDMA systems," *IEEE Trans. Inform. Theory*, to appear.
- [7] L. Ozarow, S. Shamai, and A. Wyner, "Information theoretic considerations for cellular mobile radio," *IEEE Trans. on Vehicular Technology*, vol. 43, no. 2, pp. 359–378, 1994.
- [8] 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, pp. 966–976, June 2000.
- [9] H. Yang and K. Rose, "Advances in recursive per-pixel estimation of end-to-end distortion for application in H.264," in *Proc. ICIP*, Genova, Sept. 2005.
- [10] Y. Wang, M. M. Hannuksela, V. Varsa, A. Hourunranta, and M. Gabbouj, "The error concealment feature in the H.26L test model," in *Proc. ICIP*, Rochester, NY, Sept. 2002, pp. 729–732.
- [11] 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, pp. 977–995, June 2002.
- [12] Y. Wang, G. Wen, S. Wenger, and A. K. Katsaggelos, "Review of error resilient techniques for video communications," *IEEE Signal Processing Mag.*, vol. 17, pp. 61–82, July 2000.
- [13] 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.