JOINT SOURCE CODING AND DATA RATE ADAPTATION FOR MULTI-USER WIRELESS VIDEO TRANSMISSION

Fan Zhai[1], *Zhu Li*[2], *and Aggelos K. Katsaggelos*[3]
[1]Digital Entertainment Products, Texas Instruments, Dallas, TX 75243, USA
[2]Multimedia Research Lab (MRL), Motorola Labs, Schaumburg, IL 60198, USA
[3]Department of EECS, Northwestern University, Evanston, IL 60208, USA

ABSTRACT

Much attention has been paid to the problem of optimally utilizing resources such as spectrum, power and time in order to achieve the best video delivery quality in wireless communications system, due to the fueling demand for such applications. In this work, we present a joint source coding and data adaptation scheme for downlink video transmission in a multi-user wireless network. We formulate a rate-distortion optimization problem, where the source coding and data rate are jointly designed according to the changing channel conditions. In addition, transmissions of video packets are optimally scheduled through exploiting the multi-user diversity. We solve the problem using a backward stochastic dynamical programming approach, and the simulation results have shown the advantage of the joint selection of source coding parameter and transmission rate coupled with optimal packet scheduling.

1. INTRODUCTION

With the proliferation of camera equipped cell phones and the deployment of 3G and 4G cellular infrastructure systems with higher data rate, video transmission over wireless networks has gained increased popularity. How to achieve a better video transmission quality, however, is still very challenging due to several factors, including the limited battery energy for mobile devices, time-varying distribution of the video source, and the adverse conditions in wireless channel due to fading, multi-path, and shadowing effects. Thus how to optimally utilize resources such as spectrum, power, and time in wireless communications particularly for video transmission has recently received a great amount of attention [1].

Optimal resource allocation across link-layer and physical layer for wireless data communication has been extensively studied, where the goal usually is to improve the transmission rate or spectral efficiency for a given channel through improved detection, modulation, and coding [2]. The fundamental performance limits for cross-layer resource allocation have been studied from information-theoretic perspective in [3], taking into account higher-layer quality of service (QoS) such as delay.

Different from traditional data communications, video applications usually impose a strict end-to-end delay constraint. In addition, video packets are generally of different importance. These special features typically require application layer to be considered in cross-layer resource allocation to achieve better performance. For point-to-point transmission, efficient transmission can be achieved through jointly considering source coding and channel adaptations such as channel coding, power control, data rate adaptation, etc. Reviews of this topic can be found in [1]. One example is the work in [4], where source coding and transmission rate adaptation have been jointly designed for single-user uplink video transmissions with the objective of minimizing the total expected transmission energy required to transmit the video frame subject to both a distortion and delay constraint.

In this work, we extend the problem formulation in our previous work in [4] for the single-user case to the multi-user case, aiming at achieving the best video transmission quality from the base station to multiple mobile users. Thus, in addition to the joint design of source coding and data rate for a single user according to the changing channel conditions as in [4], a key idea in this work is that the system performance can be further improved through multiuser diversity whereby transmissions of video packets for each user are optimally scheduled at each time slot according to the current channel condition, as long as the delay constraints are not violated. Another key difference in this formulation is that unlike the uplink transmission in [4], transmission energy is no longer an imposed constraint for downlink transmission.

With regard to other related work, a cross-layer optimization scheme for downlink video streaming in a multiuser wireless system is studied in [5,6], which considers pre-encoded video stream so that source coding parameter adaptation is not included in the optimization framework. In this work, however, we focus on interactive video streaming applications that require video being encoded on the fly. In addition, different from [5,6], delay constraint is enforced per packet in this work, since this type of applications typically have a short end-to-end delay.

Stochastic dynamic programming (DP) techniques are used in this paper to efficiently find an optimal source coding and transmission policy. The complexity of solving the optimization problem for the multiuser's case is linear with the number of users.

2. PRELIMINARY

2.1. Delay Components

In a video streaming system, the end-to-end delay, defined as the time between when a frame is captured at the encoder and displayed at the decoder, should be constant, if the encoder and decoder operate at the same frame rate of F frames per second [7]. The end-to-end delay of each packet can be decomposed into

$$T = \Delta T_e + \Delta T_{eb} + \Delta T_n + \Delta T_{db} + \Delta T_d, \qquad (1)$$

where ΔT_e, ΔT_{eb}, ΔT_n, ΔT_{db}, and ΔT_d, are the encoder delay, encoder buffer delay, network delay, decoder buffer delay and decoder delay for each packet, respectively. Let M be the number of packets in a video frame and k the packet index. Without loss of generality, we assume that the processing times for both encoding and decoding a packet are constants and equal to $T_p = 1/(MF)$.

Based on this assumption, the maximum encoder buffer and network delay is $T_{\max} = T - (M+1)T_p$ [7].

The encoder buffer delay can be written as $\Delta T_{eb}(k) = w_k + \frac{B_k}{R_k}$, where w_k is the waiting time in the encoder buffer, B_k is the packet length in bits, and R_k is the transmission rate in bits/sec for packet k, respectively. From [4], w_k is recursively calculated by

$$w_k = w_{k-1} + \frac{B_{k-1}}{R_{k-1}} - T_p. \tag{2}$$

This relationship will be used in the following text when deriving the per packet delay constraint.

2.2. Channel model

We consider a slow Rayleigh fading wireless channel using Gilbert-Elliot channel model, which has been shown to model Rayleigh fading channel with sufficient accuracy [8]. During each packet's transmission, the channel is modeled as a band-limited AWGN channel with fixed gain h_k. Different user's channel is assumed to be independent from each other as in [3,5,6]. The state space of the channel is $\mathcal{H}$ and the channel is governed by the state transition matrix A. For simplicity, we consider a two-state Markov model with channel state $\mathcal{H} = \{h^0, h^1\}$ and $A = \begin{vmatrix} 1-p & p \\ q & 1-q \end{vmatrix}$, where p and q are the probability of channel state transition from state h^0 to h^1 and from h^1 to h^0, respectively.

In the base station, the transmission power is subject to a constraint. Given the transmission power P, the channel capacity at channel gain h_k can be given by

$$C_k = W \log_2(1 + \frac{h_k P}{N_0 W}) \tag{3}$$

where W is the channel bandwidth and N_0 is the power spectral density of noise. From Shannon's coding theorem, (3) gives an upper bound on the transmission rate for reliable transmission given the channel gain and the transmission power.

3. PROBLEM FORMULATION

In this work, we consider a TDMA-based downlink system (for example, in IS-856) and we assume the transmission rate at each time slot can be ideally adapted to the detected channel gain, as given by (3). As streaming video applications typically enforce strict end-to-end delay constraint for each video packet as discussed in the previous section, the selection of source bit rate of each video packet needs to be balanced against transmission rate as shown in (2). In addition, source bit rates need to be optimally traded off among packets to achieve the best overall quality. This can be achieved through optimal mode selection [4], which is the technique employed in this work.

Unlike the uplink transmission application as in [4], where the goal is to optimize the tradeoff between transmission energy and video quality, in the downlink, the base station does not impose a transmission energy constraint, even though a transmission power constraint still needs to be enforced. In addition, in a multi-user system, unlike in a single user system, resources (time slots in TDMA system, for example) need to be optimally allocated among users as well in order to maximize the overall system performance.

For video transmission systems in a single user system, the metrics such as the mean squared error (MSE) between the original video and the reconstructed video at the receiver can be used as the objective function for system performance evaluation. As for a multi-user system, the video transmission quality matrix could be the extension of MSE, such as sum of the distortion of all users $\mathbb{E}_H\{\sum_{i\in\Phi}[\sum_k D_{i,k}]\}$ (the overall video transmission quality) or the max distortion of all users $\max_{i\in\Phi}\mathbb{E}_H\{\sum_k D_{i,k}\}$ (to optimize the video transmission quality of the worst user), where Φ is the set of users and the subscript $i \in \Phi$ denotes the index of users. In this work, we employ the MSE as the distortion metrics for a single user and the sum distortion as the objective function.

Let $\mathcal{S}$ and $\mathcal{C}$ be the set of source coding and channel rate parameters, respectively. Let $\boldsymbol{s} = \{s_1, ..., s_M\} \in \mathcal{S}^M$ and $\boldsymbol{c} = \{c_1, ..., c_M\} \in \mathcal{C}^M$ denote, respectively, the vector of source coding parameters and channel rate parameters for the M packets in one video frame or a group of frames. The target of optimal resource allocation is then to maximize the overall video transmission quality as well as meeting the delay constraint for all the users:

$$\begin{aligned} &\min_{s_{i,k}\in\mathcal{S}^M, c_{i,k}\in\mathcal{C}^M} \mathbb{E}_H\left\{\sum_{i\in\Phi}[\sum_{k=1}^{M} D_{i,k}]\right\} \\ &\text{s.t.} \quad \sum_{k=1}^{M} B_{i,k}/R_{i,k} \le T_{i,0}, \forall i \in \Phi \end{aligned} \tag{4}$$

where $T_{i,0}$ is the transmission delay constraint, which is usually determined by applications based on the estimated channel throughput and assigned by a higher-level rate controller. In this formulation, the source coding and transmission rate are to be adapted to the changing channel conditions. In addition, the channel resource, which is time in this case, are optimally allocated to each user, as long as the delay constraints for each packet enforced by the rate controller are not violated. This leads to multiuser diversity gain by making use of independently fading channels.

4. SOLUTION ALGORITHM

4.1. Dynamic Programming

The proposed optimization problem (4) is solved based on stochastic DP techniques. Let us first consider the single user case. Considering the transmission delay constraint in (4), it is natural to regard the state of rate controller (or the encoder buffer state) together with the channel state as the *system state*. Thus the system state at encoding packet k can be defined as $x_k = [w_k, h_k]$. The *decision* u_k at each stage is the source coding parameter s_k and the transmission rate c_k for packet k, i.e., $u_k = [s_k, c_k]$.

Next we need to convert the transmission delay control in (4) into what is characterized by the defined system state. As derived in [4], the delay constraint can be expressed in terms of w_k as

$$w_{k+1} - w_1 + kT_p \le T_0, \quad \text{for} \quad k = 1, ..., M. \tag{5}$$

Thus, the feasible choices of decisions for the k-th packet at state x_k are given by

$$\begin{aligned} \mathcal{U}(x_k) = \Big\{ &u_k \in \mathcal{S}\times\mathcal{C} : 0 \le \frac{B_k(s_k)}{R_k(c_k)} + w_k \\ &-T_p \le \min\left(T_{max}, T_0 + w_1 - kT_p\right)\Big\}, \end{aligned} \tag{6}$$

and the *cost* at each stage is the resulting distortion $D_k(u_k, x_k)$.

Since $w_k \in [0, T_{max}]$ is real-valued, the resulting state space is infinite. For computational reasons, we quantize w_k into a set of N values, $\mathcal{W} = \{w^0, w^1, ..., w^{N-1}\}$, with $w^j = (jT_{max})/(N-$

1) as in [4]. Finer quantization of w_k leads to a better approximation of the optimal solution, at the cost of more computations. The effect of this approximation is to restrict the set of feasible choices for each system state. Therefore, the resulting solution will be a conservative approximation to the optimal solution.

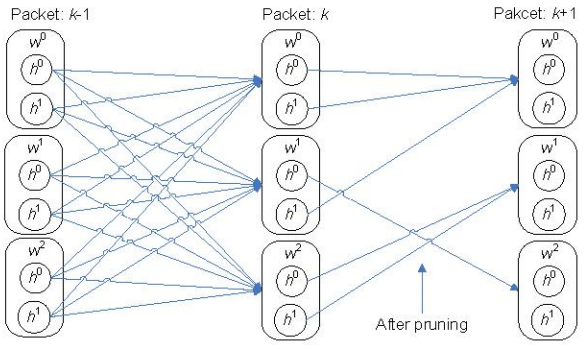

Fig. 1. Illustration of DAG formulation. Each branch leads to a deterministic value of w_k. The channel state is determined by the channel statistics.

Based on the establishment of system state, (4) can be solved by a backward DP:

$$\begin{cases} J_M(x_M) = \min\limits_{\mathcal{U}(x_M)} \{D_M(x_M, u_M)\} \\ J_k(x_k) = \min\limits_{\mathcal{U}(x_k)} \mathrm{E}_H \{D_k(x_k, u_k) + J_{k+1}(x_{k+1})\} \\ \text{for} \quad 1 \le k \le M-1 \end{cases} \tag{7}$$

Given the state of the initial stage x_1, we can obtain $J_1(x_1) = \mathbb{E}_H\{\min \sum_{k=1}^{M} D_k(s_k, c_k)\}$, which is the solution to (4) for the single user case, by recursively solving (7).

Figure 1 depicts the directed acyclic graph (DAG) of the state diagram. In this diagram, three stages corresponding to packets $k-1$ to k+1 are shown, with $N = 3$ and $\mathcal{H} = \{h^0, h^1\}$. As shown in the figure, each branch in the graph corresponds to a choice of $u_k \in \mathcal{U}(x_k)$, leading the branch to end at a specific waiting time for the next packet but not a specific channel gain, since the channel gain is a random process. The resulting cost is then $\mathbb{E}_H\{D_k(x_k, u_k) + J_{k+1}(x_{k+1})\} = D_k(u_k, w_k, h_k) + \Pr(h_k \to h^0) J_{k+1}(w_{k+1}, h^0) + \Pr(h_k \to h^1) J_{k+1}(w_{k+1}, h^1)$.

For all feasible branches $u_k \in \mathcal{U}(x_k)$ emanating from the same state, we can prune out all branches except the one with the minimum cost associated with it. By doing so starting from the last packet and moving backward to the first packet in a frame, we always keep at most $|\mathcal{H} \times \mathcal{W}|$ paths at each stage of the DP algorithm, where $|\cdot|$ is the cardinality of the set[1]. Thus the time complexity of this approach is $O(M \cdot |\mathcal{H} \times \mathcal{W} \times \mathcal{S} \times \mathcal{C}|)$. The optimal solution, denoted by $\Gamma(\boldsymbol{s}, \boldsymbol{c})$, is achieved through backtracking by choosing the path through the trellis with the minimum cost among all feasible paths. The optimal path is then selected from $\Gamma(\boldsymbol{s}, \boldsymbol{c})$ on the fly with the specific channel gain realization.

4.2. Multiple-User Case

Now we consider the case of multiple users, where the transmitter is able to selectively allocate time slots to different users. Without loss of generality, let us consider two users. The system state and decision at each stage will be $x_k = [w_{1,k}, h_{1,k}, w_{2,k}, h_{2,k}]$ and $u_k = [s_{1,k}, c_{1,k}, s_{2,k}, c_{2,k}]$, respectively. The solution process is the same as the single user case, but the complexity will be rather significant because of the large sets of system states and decisions.

As discussed regarding Gaussian source in [2], multiuser diversity improves system performance by exploiting channel fading: channel fluctuations due to fading ensure that with high probability there is a user with a channel strength much higher than the mean level; by allocating all the system resources to that user, the benefit of this strong channel is fully capitalized. Here we assume this rule generally holds for video source as well, so that the transmitter in our system always selects a packet from the user with stronger channel gain at that time slot to transmit, as long as the delay requirement is satisfied.

Because different users' channels are assumed to be independent, the trellises of each user's optimal decisions, $\Gamma_i(\boldsymbol{s}, \boldsymbol{c})$, are independent of each other. In addition, because the objective function used to evaluate the overall performance (the sum distortion) is additive, the trellis of the set for all the users' optimal decisions, denoted by $\Gamma(\boldsymbol{s}, \boldsymbol{c})$, should cover and only cover each user's trellis, i.e. $\Gamma(\boldsymbol{s}, \boldsymbol{c}) = \Gamma_1(\boldsymbol{s}, \boldsymbol{c}) \bigcup \Gamma_2(\boldsymbol{s}, \boldsymbol{c})$. This means that each user independently maintains one trellis, which is obtained the same way as in the single-user case. Thus the complexity of solving the multiuser case is merely linear with the number of users.

5. EXPERIMENTAL RESULTS

We use an H.263+ codec to perform source coding, and each row of macroblocks or slice is coded as one source packet and every packet is independently decoded. Rate control is not implemented in this work. Thus, every frame has the same transmission delay constraint of one frame's duration, i.e., $T_0 = 1/F$, and we set $F = 15$ fps. The source coding adaptation parameters considered in this work include the prediction mode (Intra, Inter, or Skip) and the quantizer used for each packet.

For simplicity, we consider two users and the settings for the two users are the same, including the test sequences (QCIF size Foreman sequence is used in simulations) and the channel parameters. T_{max} is set as 300 ms and $N = 300$ for the two users. As for the channel, we set channel bandwidth $W = 500$ kHz and AWGN with variance $N_0 W = 0.39$ W [2]. We consider a symmetric transition channel, i.e., $p = q$. The image quality measure used is the average peak signal-to-noise-ratio (PSNR), defined as PSNR$= 10 \log \frac{255^2}{\mathrm{MSE}}$ dB. Experiments based on asymmetric user settings (e.g., different users may have different channel parameters and may request different video streams) have also been performed, which are not covered here due to space limitation.

Three systems are compared. In system 1, transmission rate adaptation is disabled (so the lower rate corresponding to the lower channel gain is always used) and only the source coding parameter can be optimally selected. Packet scheduling is fixed so that time slots are allocated to the two users alternatively. In system 2, channel transmission rate is optimally selected together with source coding parameter according to the channel conditions, while packet scheduling is still fixed. In system 3, packet scheduling is enabled in additional to the optimization in system 2.

[1] Actually at each stage the maximum number of possible branches is far less than $|\mathcal{H} \times \mathcal{W}|$ because the initial system state x_1 already eliminates a lot of invalid branches according to (6).

[2] Note that the chosen parameters are the same as in [4], which are not compatible to the currently used cellular systems, so that the resulting channel SNRs, $\mathbb{E}(H)P/(N_0 W)$, do not reflect the numbers used there.

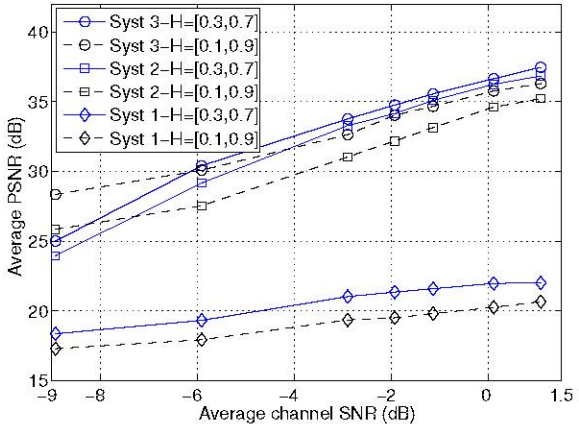

Fig. 2. Average PSNR vs. average channel SNR ($p = 0.3$).

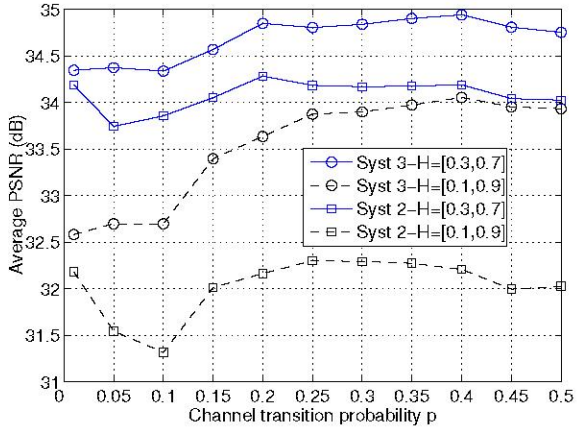

Fig. 3. Average PSNR vs. channel state transition probability p (average channel SNR is -1.93 dB).

In experiment 1, we set $p = 0.3$. In Fig. 2, we plot the average PSNR (averaged over 50 channel realizations) against different average channel SNR for two settings of channel gain set $\mathcal{H}$. It can be seen that system 2 outperforms system 1 by 5-15 dB at different channel SNRs. The gain of system 2 over system 1 comes from the flexibility of channel adaptation. When the transmitter is able to adaptively schedule packet transmission as in system 3, that is, time slots can be intelligently allocated to each user based on the detected channel gains for each user at that time slot, system 3 achieves 1.2-2.5 dB gain compared to system 2. The gain is more significant when channel gets worse, which means packet scheduling is more desirable when channel is poor.

It can also be seen from Fig. 2 that the gain of system 3 over system 2 when $\mathcal{H} = [0.1, 0.9]$ is more significant than that when $\mathcal{H} = [0.3, 0.7]$. This is because the channel in the former setting fluctuates more dramatically than the latter. As mentioned above, multiuser diversity gain is achieved through making use of the channel fluctuations. Systems 2 and 3 generally perform better at more stable channels except at very poor channel (e.g., when the average channel SNR=-9 dB). All video packets suffer from very poor channel if the channel is stable. However, as video packets are generally of different importance, more resources (the time slots with higher channel gain in this case) can be allocated to more important packets when channel fluctuates more.

Next we consider the performance of system 2 and 3 versus the change of channel state transition probability p at average channel SNR=-1.93 dB. It can be seen from Fig. 3 that, as p increases from 0.1 to 0.4, the average PSNRs of system 2 and 3 increase by around 1.2 dB and 0.5 dB, respectively. However, when $p > 0.4$, both PSNRs gradually decrease as p increases. This generally implies that video streaming favors relatively uniformly distributed channel given the same amount of average channel gain under the current settings. When p decreases from 0.1 to 0.01, the PSNR of system 2 becomes larger. This is because at such small transition probability, the channel is more likely to stay fixed for longer. Hence the initial channel state gives a better estimate of the channel, leading to better system performance. However, system 3 does not benefit from the relatively stable channel (in the range of $p < 0.1$) as shown in Fig. 3, since more stable channel means packet scheduling is less useful.

6. CONCLUSIONS

In this paper, we studied video transmission over a downlink channel in a multi-user wireless system. Specifically, we presented an optimal problem formulation, where source coding parameter can be jointly selected with data rate adaptation based on the changing channel conditions. The system performance can be further improved by allowing packet scheduling through exploiting multiuser diversity. The formulated optimization problem has been efficiently solved using a stochastic DP approach, and the solution complexity is linear with the number of users considered. Simulation results demonstrated the gain of the proposed scheme.

7. REFERENCES

[1] A. K. Katsaggelos, Y. Eisenberg, F. Zhai, R. Berry, and T. N. Pappas, "Advances in efficient resource allocation for packet-based real-time video transmission," *Proceedings of the IEEE*, vol. 93, pp. 135–147, Jan. 2005.

[2] D. Tse and P. Viswanath, *Fundamentals of wireless communications*, Cambridge University Press, 2005.

[3] R. Berry and E. Yeh, "Cross-layer wireless resource allocation," *IEEE Signal Proc. Mag.*, pp. 59–68, Sept. 2004.

[4] C. E. Luna, Y. Eisenberg, R. Berry, T. N. Pappas, and A. K. Katsaggelos, "Joint source coding and data rate adaptation for energy efficient wireless video streaming," *IEEE J. Select. Areas Commun.*, vol. 21, pp. 1710–1720, Dec. 2003.

[5] L.U. Choi, W. Kellerer, and E. Steinbach, "Cross layer optimization for wireless multi-user video streaming," in *Proc. IEEE Int. Conf. Image Processing*, Singapore, Oct. 2004.

[6] Y. Peng, S. Khan, E. Steinbach, M. Sgroi, and W. Kellerer, "Adaptive resource allocation and frame scheduling for wireless multi-user video streaming," in *Proc. IEEE Int. Conf. Image Processing*, Genova, Sept. 2005.

[7] C. Y. Hsu, A. Ortega, and M. Khansari, "Rate control for robust video transmission over burst-error wireless channels," *IEEE J. Select. Areas Commun.*, vol. 17, pp. 756–773, May 1999.

[8] H. S. Wang and P. Chang, "On verifying the first-order Markovian assumption for a Rayleigh fading channel model," *IEEE Trans. Vehicular Technol*, vol. 45, pp. 353–357, May 1996.