

Scalable Video Coding and Packet Scheduling for Multiuser Video Transmission Over Wireless Networks

Ehsan Maani^a, Peshala V. Pahalawatta^b, Randall Berry^a, and Aggelos K. Katsaggelos^a

^aEECS Department, Northwestern University, Evanston, IL;

^bImage Technology Group, Dolby Labs, Burbank, CA

ABSTRACT

Wireless video transmission is prone to potentially low data rates and unpredictable degradations due to time-varying channel conditions. Such degradations are difficult to overcome using conventional video coding techniques. Scalable video coding offers a flexible bitstream that can be dynamically adapted to fit the prevailing channel conditions. Advances in scalable video compression techniques, such as the newly adopted scalable extension of H.264/AVC, as well as recent advances in wireless access technologies offer possibilities for tackling this challenge. In this paper, a content-aware scheduling and resource allocation scheme is proposed, that uses a gradient-based scheduling framework in conjunction with scalable video coding techniques to provide multiple high quality video streams over a range of operating conditions to multiple users. Simulation results show that the proposed scheme performs better than conventional content-independent scheduling techniques.

Keywords: Scalable video coding, wireless video streaming, cross-layer design

1. INTRODUCTION

Video transmission over wireless networks to multiple mobile users has remained a challenging problem due to potential limitations on bandwidth and the time-varying nature of wireless channels. Recent advances in wireless access technologies, such as, HSDPA, IEEE 802.16 (WiMAX) are targeted at achieving higher throughputs over wireless networks. At the same time, advances in video compression, and especially, scalable video coding, aim to provide content that can adapt to time-varying network conditions. In this work, we investigate an approach that combines the advanced resource allocation strategies allowed in wireless technologies such as HSDPA with the flexibility offered by the scalable extension of the H.264/AVC standard,¹ referred to as SVC in this work, to facilitate the transmission of high quality video to multiple users over wireless networks.

For multiuser data transmission, advanced wireless technologies can make use of timely and frequent channel feedback to dynamically determine per-user resource allocation strategies. This enables such systems to exploit the available multiuser diversity by assigning more of the available resources to users with the best channel states.² While achieving high overall throughputs, fairness across users must also be maintained in order to ensure that each user receives a reasonable quality of service (QoS). A number of scheduling and resource allocation policies that aim to achieve fairness across users, while exploiting multiuser diversity, can be categorized as gradient-based scheduling policies.^{3,4} Gradient-based policies define a user utility as a function of some quality of service measure, such as throughput, and then maximize a weighted sum of the users' data rates where the weights are determined by the gradient of the utility function. For example, choosing the weights to be the reciprocals of the long term average throughputs of each user, leads to a proportionally fair scheduling scheme.⁵ In,⁶ we presented a scheme for combining a gradient-based scheduling policy with application-layer information to improve the quality of service for multiuser video transmission. The video content, however, was assumed to be coded using a conventional hybrid motion-compensated technique without any scalable extensions. In this work, we explore the use of scalable video coding techniques for application and channel dependent packet scheduling and resource allocation. We present some of the natural advantages to be had, and some pitfalls to be avoided, when using scalable coded video in conjunction with a content-dependent gradient-based scheduling policy. Furthermore, in contrast to the two-step suboptimal solution in⁶ for determining the allocated user resources and the allowed transmission rates (or equivalently the loss probabilities), here we propose an optimal approach to simultaneously determine all parameters in the optimization.

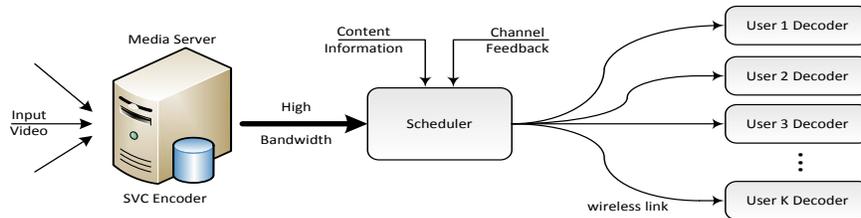


Figure 1. Overview of multiuser downlink video streaming system

Multiuser streaming of scalable video has been considered in.⁷ In,⁷ temporal scalability, in the form of hierarchical Bi-prediction,⁸ and SNR scalability, in the form of progressive refinement through fine granularity scalability (FGS),⁹ is considered. A simple packet dropping strategy is used for buffer management and a maximum throughput scheduling strategy is used at the air interface. In this paper, we consider an optimized, content-dependent packet dropping strategy that prioritizes the available packets based on their contribution to the overall quality of the video signal. Optimized packet dropping techniques can provide significant performance gain especially when combined with an application-aware resource allocation scheme under a variety of channel conditions. Nevertheless, optimized packet prioritization in SVC is considerably more challenging than that of MPEG-4 Visual. The primary reason for this is that the process of motion-compensated prediction (MCP) in SVC is designed such that the highest available picture quality is employed for frame prediction in a GOP, except for the *key frames*, i.e., the lowest temporal layer. Therefore, dropping quality refinement packets of a picture results in propagation of *drift* to all pictures predicted from it. In other words, the distortion of a picture (except for the key frames) depends on the enhancement layers of the pictures from which it has been predicted. This concept of key frames when combined with the hierarchical Bi-prediction structure of SVC results in highly distinctive packet importance and make the optimal extraction challenging. The true impact of a refinement packet on the global sequence quality can only be found by performing the computationally expensive motion-compensation operations for all pictures that depend on the picture being refined. In our recent work we proposed a model to efficiently and accurately approximate the distortion and expected distortion of an SVC bit stream for inclusion of an arbitrary subset of the available refinement packets.¹⁰ Here, we employ this distortion model to prioritize source packets in the queue based on their estimated impact on the global video quality.

The remainder of this paper is organized as follows. In Sec. 2, we provide an overview of the wireless multiuser video transmission application under consideration. We also provide a brief description of the scalable video coding techniques used, and the gradient-based resource allocation scheme employed at the scheduler. In Sec. 3, we develop the packet prioritization strategies that can be used with scalable coded video content. Problem formulation and solution is presented in Sec. 4. Section 5 details some of the simulation results that show the improvement in performance of the content-aware scheme over a content-independent scheme, and also show a significant improvement in performance from the use of scalable video coding over conventional non-scalable video coding. The main conclusions of this work are presented in Sec. 6.

2. WIRELESS MULTIUSER VIDEO TRANSMISSION SYSTEM

The packet scheduling scheme and resource allocation metric developed in this work are for a wireless multiuser downlink video transmission system such as the one shown in Fig. 1. The system consists of a media server that contains multiple pre-encoded video sequences in packetized scalable bitstreams.

2.1 Media Encoding with Scalable Video Coding

An overview of the techniques and applications of scalable video coding, especially as it pertains to SVC, the scalable extension to H.264/AVC, is provided in.¹¹ In general, a scalable video bitstream offers three primary types of scalability that may be used individually, or in combination. They are: spatial scalability, which allows the transmission of the same video sequence at different resolutions depending on the user requirements or bandwidth constraints, temporal scalability, which allows the transmission of the video sequence at different

frame rates without error propagation due to the skipped frames, and quality (SNR) scalability, which allows the transmission of progressively refined bitstreams depending on the available data rates. The design of the SVC allows for these three scalabilities. The video bit stream generated by the SVC is commonly structured in layers, consisting of a base layer (BL) and one or more enhancement layers (ELs). Each enhancement layer either improves the resolution (spatial or temporal) or the quality of the video sequence. A layer representing a specific spatial or temporal resolution is identified with a dependency identifier d or temporal identifier t . Moreover, quality refinement layers inside each dependency layer are identified by a quality identifier q . Similarly to H.264/AVC, the coded video data of the SVC are organized into packets with an integer number of bytes, called Network Abstraction Layer (NAL) units. Each NAL unit, belongs to a specific spatial, temporal, and quality layer. This work, focuses on optimizing over temporal and SNR scalability levels and excludes spatial scalability since it is reasonable to assume that the spatial resolution will remain static within one video streaming session.

This work assumes that each application layer packet contained in the media server is independently decodable as specified in the NAL unit structure. For transport from the media server to the base station, each coded NAL unit is packetized into one or more RTP packets but it is reasonable to assume that no two NAL units will be contained in one RTP packet. Therefore, each video packet will contain information about its own decoding deadline, in addition to the number of bits contained in the packet. The decoding deadline of a packet stems from the video streaming requirement that all the packets needed for the decoding of a frame of the video sequence must be received at the decoder buffer prior to the playback time of that frame. Multiple packets (e.g., all the packets belonging to one picture or GOP) can be assigned the same decoding deadline. Any packet that is left in the transmission queue after its decoding deadline has expired must be dropped since it has lost its value to the decoder.

2.2 Gradient-Based Resource Allocation and User Scheduling

The media server is connected to one or more wireless base stations over a high bandwidth network. Our work is focused on the scheduling and resource allocation performed at each base station, which as shown in Fig. 1, services multiple mobile clients over the air interface. As in,¹² we assume a wireless network where a combination of TDM and CDMA or OFDM is used to simultaneously transmit data to multiple users during each transmission time slot. We assume that channel feedback is available to the scheduler at each time slot, which is also a reasonable assumption based on recent wireless access technologies. The achievable data rate for each user during a particular time slot is determined by the number of spreading codes, the fraction of the transmission power, and the adaptive modulation and coding scheme assigned to the user. Given a particular spreading code (n_i), and transmission power (p_i) assignment for user i , the achievable information rate for the user cannot reliably exceed the channel state dependent capacity of the user's channel. Nevertheless, when the exact channel state is not known 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 (ϵ_i) can be calculated, using an outage probability formulation,¹³ as a function of the assigned transmission power, bandwidth, and rate. As discussed in,⁶ this loss probability can be obtained by

$$\epsilon_i = F_{h_i|e_i}\left(\frac{n_i}{p_i}\left(2^{\frac{r_i}{n_i B}} - 1\right)|e_i\right), \quad (1)$$

where B denotes the maximum symbol rate per code, h_i the instantaneous channel fading state (SINR per unit power) in the time-slot, r_i is the assigned transmission rate, and $F_{h_i|e_i}$ the cumulative probability density function of the instantaneous channel fading state conditioned on the observed channel estimate, e_i .

This work assumes that only partial (imperfect) channel state information is available at the scheduler. Errors in the channel estimate can arise from the delay in the feedback channel combined with Doppler spread and quantization errors. Here, we employ a Nakagami- m channel model, thus, the cumulative probability density function can be written as,

$$F_{h_i|e_i}(x) = \frac{\gamma\left(m, \frac{mx}{e_i}\right)}{\Gamma(m)}, \quad (2)$$

where m is the Nakagami shape parameter, $\gamma()$ denotes the incomplete gamma function, and $\Gamma(m)$ denotes the gamma function of m .

3. PACKET ORDERING FOR SCALABLE VIDEO

The most important aspect of this work is that of choosing a packet scheduling strategy and a content-aware utility metric to be used within the gradient-based scheduling framework. The key idea in a content-aware gradient-based scheduling technique is to sort the packets in the transmission buffer for each user based on the contribution of each packet to the overall video quality, and then, to construct a utility function so that the gradient of the utility reflects the contribution of the next packet in the ordered queue.

As mentioned earlier, packet prioritization in SVC is a nontrivial process especially in a packet lossy environment. Missing quality NAL units in one picture will result in drift propagations in the following dependent pictures. The exact amount of the propagated drift and its impact on the video quality can only be computed by decoding. More importantly, missing the base layer of a picture forces the decoder to drop all the dependent pictures since they are predictively coded from the lost information. Yet, due to the unreliable nature of the wireless link, the intact arrival of the any of the transmitted packets at the receiver remains a stochastic process as seen by the transmitter. Therefore, loss probability of the transmitted packets has to be taken into account in the prioritization of the remaining packets. This is especially important in SVC due to the strong dependency of the NAL units among each other in quality, spatial, and temporal dimensions.

The computational burden of evaluating the expected sequence distortion for a given set of NAL units \mathcal{Q} and their associated loss probabilities using direct computation is far from being manageable. We addressed this problem in great detail in¹⁰ and proposed a method to accurately approximate the expected distortion. This method decodes the bit stream with various picture qualities and analyzes the results in order to build a model for the estimation of the distortion caused by drift at the picture level. This model consists of five floating point numbers per frame that determine the drift distortion in the frame based on the distortion of its parents (reference frames). Using this distortion model, the expected distortion can be estimated by considering various packet loss scenarios with their associated probabilities as discussed in.¹⁰

Armed with a technique for fast evaluation of the expected distortion, a content-aware packet ordering scheme can be envisaged. Since the base layer of the key picture is required for the decoding of all pictures of the GOP, it is given the highest priority and thus is the first packet to be added in the queue for transmission. Subsequent packets are ordered such that the next highest priority is given to the decodable (decodable given that the higher priority packets are received) packet, π_i that maximizes $\{\frac{\partial ED}{\partial r_i}\}_{\pi_i}$ where ED represents the expected distortion of the entire GOP. Note that the loss probability of the packets already transmitted is known and therefore the expected distortion of the GOP can be evaluated. An example is shown in Fig. 2. The prediction hierarchy of a GOP of size 4 is illustrated in Fig. 2(a). A possible packet prioritization in this GOP is given in Fig. 2(b). At the first step, only the base layer packet of the key frame, S_4 is decodable (assuming the key frame of the previous GOP, S_0 is received). If S_4 is decoded, then both the base layer of S_2 , i.e., (S_2, Q_0) , and the first quality refinement of S_4 i.e., (S_4, Q_1) , are decodable and the packet with the largest utility gradient is picked. Once that packet, in this case (S_4, Q_1) , is added to the queue, the set of new decodable packets is (S_4, Q_2) , (S_2, Q_1) , (S_1, Q_0) , and (S_3, Q_0) .

A simple alternative to the maximum-gradient packet ordering scheme is the basic packet ordering method implemented in the H.264 scalable reference software JSVM.¹⁴ This technique utilizes the high-level syntax elements dependency identifier d , temporal identifier t , and quality identifier q for prioritization. A target spatial resolution is first determined by the application. Then, the base layer of each spatial and temporal resolution lower or equal to the target spatial and temporal resolutions have to be included first. Next, for each lower spatial resolution, NAL units of higher quality levels are ordered in increasing order of their temporal level. Finally, for the target spatial resolution, NAL units are ordered based on their quality level and are included until the target bit rate is reached.

4. PROBLEM FORMULATION

Given the packet ordering scheme and the relationship between the loss probability ϵ_i and the rate r_i , at each time slot, a gradient-based scheduling policy will allocate resources to solve

$$\max_{(\mathbf{n}, \mathbf{p}, \boldsymbol{\epsilon})} \nabla U \cdot \mathbf{r}(\mathbf{n}, \mathbf{p}, \boldsymbol{\epsilon}), \quad (3)$$

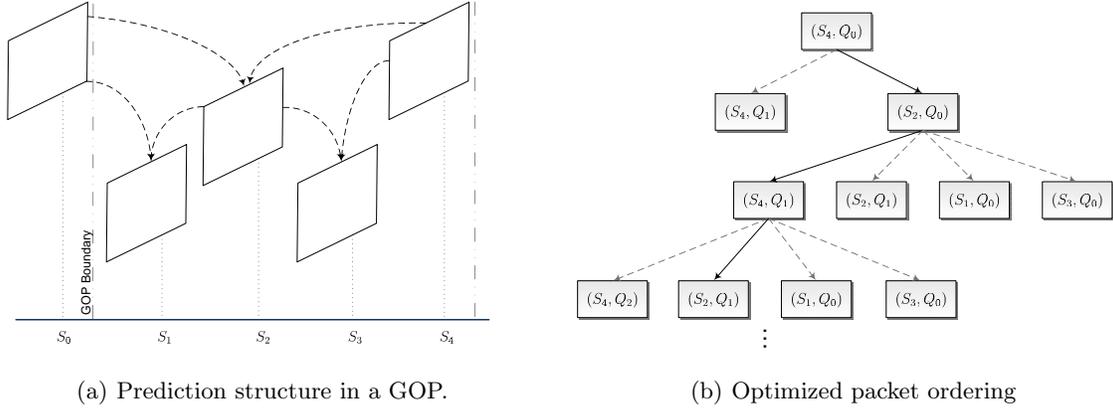


Figure 2. An example of prediction structure and packet ordering based on expected distortion.

where $\mathbf{p} = (p_1, p_2, \dots, p_K)$, $\mathbf{n} = (n_1, n_2, \dots, n_K)$, $\boldsymbol{\epsilon} = (\epsilon_1, \epsilon_2, \dots, \epsilon_K)$, $\mathbf{e} = (e_1, e_2, \dots, e_K)$ and K is the number of users. U denotes the system utility function which is a sum of users' utility functions U_i , i.e., $U = \sum_i^K U_i$. Note that the time slot index, t , is omitted, since it remains the same throughout this discussion. The available resources are constrained by,

$$0 \leq \sum_{i=1}^K n_i \leq N, \quad 0 \leq n_i \leq N_i, \quad n_i \in \mathbb{Z}^+, \quad \forall i, \quad (4)$$

and

$$0 \leq \sum_{i=1}^K p_i \leq P, \quad 0 \leq \frac{r_i}{n_i} \leq \hat{S}_i, \quad 0 \leq \epsilon_i < 1, \quad \forall i. \quad (5)$$

The constraint on the maximum number of spreading codes, N , depends on the specific wireless standard, and N_i depends on the mobile user's equipment. Furthermore, n_i 's must have integer values. A constraint on the total power, P , is introduced to decrease the potential for interference across neighboring cells. In (3), $\mathbf{r}(\mathbf{n}, \mathbf{p}, \boldsymbol{\epsilon}, \mathbf{e})$ denotes the achievable rate vector where each of its elements $r_i(n_i, p_i, \epsilon_i, e_i)$ is the achievable rate for user i given the resources (n_i, p_i) , loss probability ϵ_i , and user i 's estimated channel state e_i . Furthermore, a constraint on the rate per code is imposed by the maximum rate of the available modulation and coding schemes. The utility function we consider in this work is the expected distortion of a GOP. It should be noted that the values of $\frac{\partial U}{\partial r_i}$, and e_i are also time dependent and change at each time slot based on the current state of the transmission queue of user i , the current state of the channel, and the end-to-end channel feedback if available.

4.1 Solution Algorithm

A problem similar to (3) in the case of a TDM/CDMA, or TDM/OFDMA type system is considered in.³ However, it is assumed that the loss probabilities $\epsilon_i = 0$ if the user's rate in a time slot is less than the channel state dependent capacity of the users channel. Here, we expand the solution in³ to jointly determine optimal loss probabilities ϵ_i^* (and hence the optimal user rate r_i^*) as well as the optimal resource parameters $(\mathbf{n}^*, \mathbf{p}^*)$. Based on the assumed channel model (see Sec. 2.2), the user rate r_i can be computed according to

$$r_i = n_i B \log_2 \left(1 + \frac{e_i p_i}{n_i} \frac{Q^{-1}(m, \epsilon_i)}{m} \right), \quad (6)$$

where $Q^{-1}(m, z)$ represents the inverse regularized incomplete gamma function. By substituting for r_i in equation (3), the optimization of (3) becomes

$$V(\mathbf{n}^*, \mathbf{p}^*, \boldsymbol{\epsilon}^*) = \max_{\mathbf{n}, \mathbf{p}, \boldsymbol{\epsilon}} \sum_{i=1}^K u_i(\epsilon_i) n_i B \log \left(1 + \frac{e_i p_i}{n_i} \frac{Q^{-1}(m, \epsilon_i)}{m} \right), \quad (7)$$

in which $(\mathbf{n}, \mathbf{p}, \boldsymbol{\epsilon})$ are subject to the constraints of (4), and (5). $u_i(\epsilon_i)$ denotes the derivative of the user's utility function at a given loss probability of ϵ_i . This optimization can be carried out by considering the dual formulation of the problem. Numerical evaluation of the hessian matrix, shows that this is a convex optimization and thus has no duality gap. To further simplify notation, we define the function $f(\epsilon_i) = Q^{-1}(m, \epsilon_i)/m$, in which the dependance on m is implicit since m is a fixed model parameter. Therefore, the Lagrangian $L(\mathbf{p}, \mathbf{n}, \boldsymbol{\epsilon}, \mu, \lambda)$ with two multipliers λ and μ can be expressed as

$$L(\mathbf{p}, \mathbf{n}, \boldsymbol{\epsilon}, \mu, \lambda) = \sum_{i=1}^K u_i(\epsilon_i) n_i B \log\left(1 + \frac{e_i p_i}{n_i} f(\epsilon_i)\right) + \lambda\left(P - \sum_i p_i\right) + \mu\left(N - \sum_i n_i\right). \quad (8)$$

Based on this we can define the dual function

$$L(\lambda, \mu) = \max_{(\mathbf{p}, \mathbf{n}, \boldsymbol{\epsilon}) \in \mathcal{X}} L(\mathbf{p}, \mathbf{n}, \boldsymbol{\epsilon}, \mu, \lambda), \quad (9)$$

where the set \mathcal{X} defines the per-user constraints, that is,

$$\mathcal{X} := X\{(\mathbf{n}, \mathbf{p}, \boldsymbol{\epsilon}) \geq \mathbf{0} : n_i \leq N_i, \epsilon_i < 1, 0 \leq p_i \leq \frac{\hat{s}_i n_i}{e_i f(\epsilon_i)}\}, \quad (10)$$

and \hat{s}_i is determined by the maximum rate per code constraint in (5). To evaluate the dual function (9), we proceed in three steps. First, we optimize the Lagrangian (8) over \mathbf{p} , for a fixed λ, μ, \mathbf{n} and $\boldsymbol{\epsilon}$. We then optimize over $\boldsymbol{\epsilon}$ to obtain the value of the dual function for a fixed \mathbf{n} . Last, we perform the optimization over \mathbf{n} to compute the dual function. Computation of the optimal power p_i^* directly follows from the Kuhn-Tucker conditions. For a fixed $\mathbf{n}, \mathbf{0} \leq \boldsymbol{\epsilon} \leq \mathbf{1}$ and any $\lambda, \mu \geq 0$, the power optimal allocation \mathbf{p}^* with respect to the constraints in (4) and (5) is given by

$$p_i^* = \frac{n_i}{e_i f(\epsilon_i)} C\left(\left(\frac{e_i u_i(\epsilon_i) f(\epsilon_i)}{\lambda} - 1\right), 0, \hat{s}_i\right), \quad (11)$$

where the clipping function $C(x, a, b) := \max\{\min\{x, b\}, a\}$. Substituting \mathbf{p}^* in the Lagrangian (8), we obtain

$$L(\mathbf{p}^*, \mathbf{n}, \boldsymbol{\epsilon}, \mu, \lambda) = \sum_i (n_i h_i(\epsilon_i, \lambda) - \mu n_i) + \lambda P + \mu N, \quad (12)$$

where

$$h_i(\epsilon_i, \lambda) = \begin{cases} 0, & \lambda \geq e_i u_i(\epsilon_i) f(\epsilon_i), \\ \frac{\lambda}{e_i f(\epsilon_i)} - u_i(\epsilon_i) - u_i(\epsilon_i) \log\left(\frac{\lambda}{e_i u_i(\epsilon_i) f(\epsilon_i)}\right), & \frac{e_i u_i(\epsilon_i) f(\epsilon_i)}{\hat{s}_i + 1} \leq \lambda < e_i u_i(\epsilon_i) f(\epsilon_i), \\ u_i(\epsilon_i) \log(1 + \hat{s}_i) - \lambda \frac{\hat{s}_i}{e_i f(\epsilon_i)}, & \lambda < \frac{e_i u_i(\epsilon_i) f(\epsilon_i)}{\hat{s}_i + 1}. \end{cases} \quad (13)$$

It is apparent from equation (12), that the Lagrangian only depends on ϵ_i through the function $h_i(\epsilon_i, \lambda)$. Since there is no constraint on $\boldsymbol{\epsilon}$, for each λ , the maximum of the Lagrangian is achieved when $h_i(\epsilon_i, \lambda)$ is maximized for each user i . Consequently, the optimization over loss probability ϵ_i simplifies to

$$\epsilon_i^* = \arg \max_{\epsilon_i \in [0, 1]} h_i(\epsilon_i, \lambda). \quad (14)$$

The one-dimensional optimization of (14) can be carried out by setting $\partial h_i(\epsilon_i, \lambda)/\partial \epsilon_i = 0$. Note that since the utility gradient $u_i(\epsilon_i)$ is linear in ϵ_i^{10} and the optimal loss probability ϵ_i^* is expected to be in the range of $[0, 0.5]$, a solution can be efficiently found using a simple one-dimensional root-finding algorithm such as bisection search. We further notice that the Lagrangian (8) is linear with respect to n_i . Therefore, the optimization over \mathbf{n} is straightforward:

$$n_i^* = \begin{cases} 0, & h_i(\epsilon_i^*, \lambda) < \mu, \\ N_i, & h_i(\epsilon_i^*, \lambda) > \mu. \end{cases} \quad (15)$$

In cases where $h_i(\epsilon_i^*, \lambda) = \mu$, the choice of n_i is arbitrary. Therefore, the dual function (9) can be evaluated as

$$L(\lambda, \mu) = \sum_i [h_i(\epsilon_i^*, \lambda) - \mu]^+ N_i + \mu N + \lambda P, \quad (16)$$

where $[x]^+$ denotes the positive part of x . The optimization of the dual over λ and μ is the same as in³ and is not covered here. A closed form solution exists for μ^* , however, λ^* is obtained by using a convex search technique, such as the bisection method or a Fibonacci search.³ Also note that, from (11), if $\lambda > e_i u_i(\epsilon_i) f(\epsilon_i)$, then user i will be allocated zero power. Therefore the optimal λ^* , must satisfy

$$0 \leq \lambda^* \leq \max_i \{ \max_{\epsilon_i} \{ e_i u_i(\epsilon_i) f(\epsilon_i) \} \}. \quad (17)$$

Notice that $u_i(\epsilon_i) f(\epsilon_i) \propto (1 - \epsilon_i) Q^{-1}(m, \epsilon_i)$ and thus its maximum over ϵ_i is known *a priori*.

5. SIMULATION STUDY

Simulations were performed to determine the gains to be expected by using the gradient-based scheduling framework for content and channel dependent scheduling of scalable video. Eight video sequences with varied content: “Mobile”, “Stefan”, “Soccer”, “City”, “Foreman”, “Coastguard”, “News”, and “Silent”, in QCIF (176x144) format were used for the simulations. The sequences were encoded with JSVM 9.15 reference software¹⁴ into 2 layers: a base layer and an enhancement layer (with the same spatial resolution). The basis quantization parameter for the enhancement layer was chosen such that an average maximum PSNR of 38 ± 0.5 (dB) is achieved. The quantization parameter of the base layer was selected to be that of the enhancement layer minus 8. The enhancement layer is further divided in to 5 MGS layers to provide more flexibility in rate adaptation. The GOP size is 8 pictures and the transmission time per GOP was set at 256 msec. A sufficient buffer time was assumed to be available for the reliable transmission of the first frame (I frame) of each user. If a video packet could not be completely transmitted within a given transmission opportunity, it was fragmented at the MAC layer. All fragments generated from an application layer packet were given the same priority as the original packet. Furthermore, It was assumed that all fragments of an application layer packet must be received at the decoder in order for it to be correctly decoded. An ACK/NACK feedback for transmitted packet fragments was assumed to be available with a feedback delay of 4 msec. Therefore, if a NACK is received for a fragment of a transmitted application layer packet whose decoding deadline has not yet expired, then that fragment will be reinserted in the transmission queue with its original priority. In case the application layer packet that belongs to a base layer is not available at the decoder by the decoding deadline, a simple decoder error concealment technique mentioned in Sec. 3 is used to conceal losses.

The wireless network was modeled as an HSDPA system with $N = 15$, which is the total number of spreading codes available in HSDPA, and $N_i = 5$ for each user. HSDPA provides 2 msec transmission time slots (i.e., 128 time slots per GOP for the specified GOP size), with channel feedback also available every 2 msecs. Realistic channel traces for an HSDPA system were obtained using a proprietary channel simulator developed at Motorola Inc. The simulator accounts for correlated shadowing and multipath fading effects with 6 multipath components. For the channel traces, users were located within a 0.8km radius from the base station and user speeds were set at 30km/h. The maximum SINR per code constraint was set at 2dB for each user. The average PSNR results are the average Y-PSNR over 145 frames of each sequence under 50 different channel realizations.

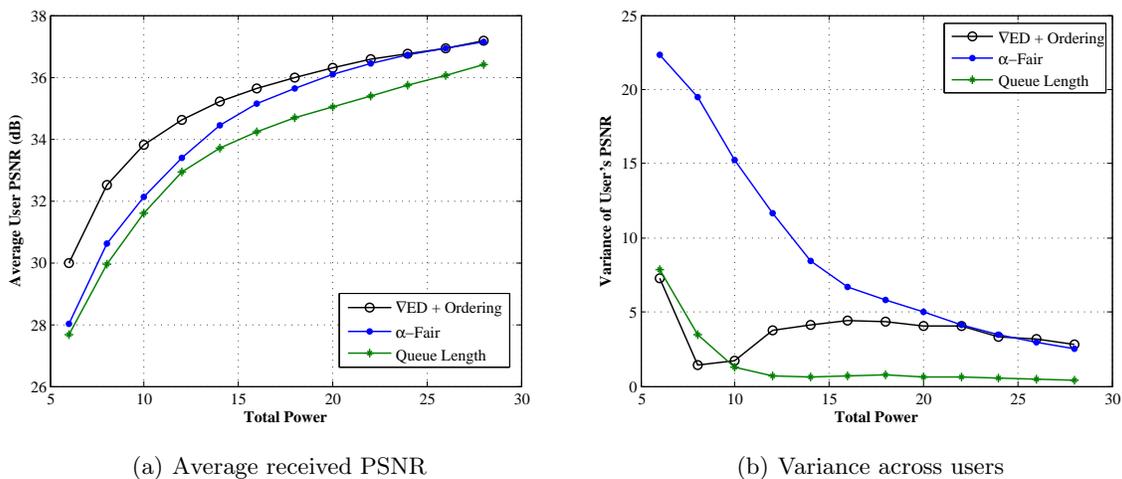
We compare three different scheduling and packet ordering strategies: 1) Our proposed content-aware distortion based scheduling and optimized packet ordering; 2) A throughput-based scheduling scheme as in¹⁵ with a fairness parameter α in which the user’s utility function is defined by

$$U_i(W_{i,t}) = \frac{1}{\alpha} (W_{i,t})^\alpha, \quad \alpha \leq 1, \quad \alpha \neq 0, \quad (18)$$

where $W_{i,t}$ is the average user throughput; 3) Gradient-based scheduling framework but with a queue length dependent metric as in the M-LWDF (Modified-Largest Weighted Delay First) algorithm proposed in.¹⁶ In this case, the utility gradient, $\frac{\partial U}{\partial r_i}$, in (3), is replaced by the total length in bits of the remaining packets in user i ’s

transmission queue. Systems 2 and 3 will perform very poorly especially at low power if they are implemented as described above since they may not schedule the key picture of a sequence within a GOP duration. In that case, all packets received at the decoder for the future GOPs are dropped by the decoder since they depend on the key picture of a preceding GOP (key pictures are inter-coded) which will result in extensive inefficiency. To avoid this problem, we modify the utility gradients of the key pictures in these two systems according to System 1 to guarantee that they will be scheduled for transmission. Thus, the performance of the content-independent systems shown is the upper bound of what we expect in reality.

Figure 3 shows a comparison between the distortion-based scheduling metric and α -fair scheduling over varying total power, and therefore, varying network operating conditions. α in this experiment is set to 0.80. The basic ordering scheme is used for packet prioritization in both maximum throughput and queue length methods. Figure 3(a) shows that the distortion-based metric performs better in terms of average received PSNR over the tested range of operating conditions. As should be expected, Fig. 3(b) shows that in an α -fair scheduling scheme, although users receive similar overall throughputs, the average user video quality is quite different. This is also expected since for a particular quality the rate of the video signal may vary substantially. The queue length method, on the other hand, achieves low variance in signal quality across users; however the average quality is also low. This is especially more visible in Figure 4 where we plot the average user's PSNR for each sequence.



(a) Average received PSNR

(b) Variance across users

Figure 3. Comparison with content-independent scheduling techniques.

6. CONCLUSIONS

This work presents a content-aware packet scheduling and resource allocation scheme for use in a scalable video coding framework that achieves a significant improvement in performance over content independent schemes. Simulation results show that the proposed content-aware metric provides a more efficient and robust method to allocate resources in a downlink video transmission system. In our simulations, we assumed all users had equal distances from the base station to be able to compare our proposed scheme against the throughput-based schemes at the peak of their performance.

It is also apparent that scalable video coding offers the possibility of using simple packet prioritization strategies without compromising the performance of the system. The packet prioritization can be performed offline and signaled to the scheduler along with the utility metrics of each packet. Most importantly, significant gains in performance can be seen in using scalable video coding as opposed to conventional non-scalable video coding over the types of time-varying networks studied in this work.

REFERENCES

- [1] "Joint draft ITU-T rec. H.264 — ISO/IEC 14496-10 / amd.3 scalable video coding," (2007).

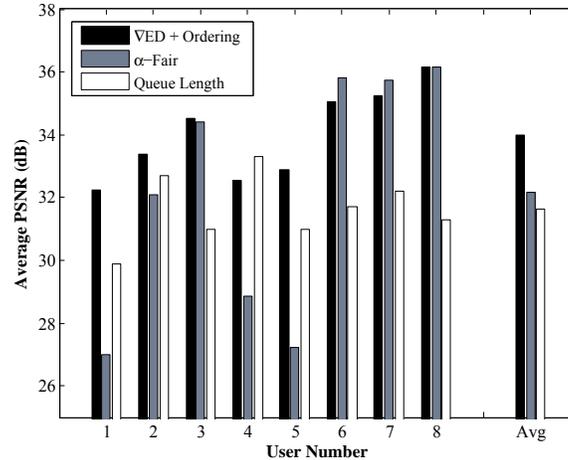


Figure 4. Average received PSNR per user, $P = 10w$.

- [2] Shakkottai, S., Rappaport, T., and Karlsson, P., "Cross-layer design for wireless networks," *IEEE Commun. Mag.* **41**, 74–80 (Oct. 2003).
- [3] Agrawal, R., Subramanian, V., and Berry, R., "Joint scheduling and resource allocation in CDMA systems," *IEEE Trans. Inf. Theory*. to appear.
- [4] Kumaran, K. and Viswanathan, H., "Joint power and bandwidth allocation in downlink transmission," *IEEE Trans. Wireless Commun.* **4**, 1008–1016 (May 2005).
- [5] Jalali, A., Padovani, R., and Pankaj, R., "Data throughput of CDMA-HDR a high efficiency - high data rate personal communication wireless system," in [*Proc. VTC*], (Spring 2000).
- [6] Maani, E., Pahalawatta, P., Berry, R., Pappas, T. N., and Katsaggelos, A. K., "Resource allocation for downlink multiuser video transmission over wireless lossy networks," *IEEE Transactions on Image Processing* **17**, 1663–1671 (September 2008).
- [7] Liebl, G., Schierl, T., Wiegand, T., and Stockhammer, T., "Advanced wireless multiuser video streaming using the scalable video coding extensions of H.264/MPEG4-AVC," in [*Proc. IEEE Int. Conf. on Multimedia and Expo*], (2006).
- [8] Schwarz, H., Marpe, D., and Wiegand, T., "Analysis of hierarchical B pictures and MCTF," in [*Proc. IEEE Int. Conf. on Multimedia and Expo*], (2006).
- [9] Radha, H., van der Schaar, M., and Chen, Y., "The MPEG-4 fine-grained scalable video coding method for multimedia streaming over IP," *IEEE Trans. Multimedia* **3**, 53–68 (Mar. 2001).
- [10] Maani, E. and Katsaggelos, A. K., "Unequal error protection for robust streaming of scalable video over packet lossy networks," *Circuits and Systems for Video Technology, IEEE Transactions on* (2009). to appear.
- [11] Schwarz, H., Marpe, D., and Wiegand, T., "Overview of the scalable video coding extension of the H.264/AVC standard," *Circuits and Systems for Video Technology, IEEE Transactions on* **17**, 1103–1120 (Sept. 2007).
- [12] Pahalawatta, P., Berry, R., Pappas, T., and Katsaggelos, A., "Content-aware resource allocation and packet scheduling for video transmission over wireless networks," *IEEE Journal on Selected Areas in Communications* **25**, 749–759 (May 2007).
- [13] Ozarow, L., Shamai, S., and Wyner, A., "Information theoretic considerations for cellular mobile radio," *IEEE Trans. on Vehicular Technology* **43**(2), 359–378 (1994).
- [14] Reichel, J., Schwarz, H., and Wien, M., "Joint scalable video model 11 (jsvm 11)." Joint Video Team, Doc. JVT-X202 (July 2007).
- [15] Agrawal, R., Bedekar, A., La, R., and Subramanian, V., "A class and channel-condition based weighted proportionally fair scheduler," in [*Proc. of ITC 2001*], (Sept. 2001).
- [16] Andrews, M., Kumaran, K., Ramanan, K., Stolyar, A., Whiting, P., and Vijayakumar, R., "Providing quality of service over a shared wireless link," *IEEE Commun. Mag.* **39**, 150–154 (Feb. 2001).