Journal of Zhejiang University SCIENCE
B
ISSN 1673-1581
http://www.zju.edu.cn/jzus

Joint Rate Control and Scheduling for Wireless Uplink Video Streaming

[1]Jianwei Huang, [2]Zhu Li, [1]Mung Chiang, [3]Aggelos K. Katsaggelos

1 *Department of Electrical Engineering, Princeton University, Princeton, New Jersey, USA*
2 *Multimedia Research Lab (MRL), Motorola Labs, Schaumburg, Illinois, USA*
3 *Department of Electrical Engineering & Computer Science, Northwestern University, Evanston, Illinois, USA*
E-mail: jianweih@princeton.edu, zhu.li@motorola.com, chiangm@princeton.edu, aggk@ece.northwestern.edu

Received ; revision accepted

Abstract: We solve the problem of uplink video streaming in CDMA cellular networks by jointly design the rate control and scheduling algorithms. In the pricing-based distributed rate control algorithm, the base station announces a price for the per unit average rate it can support, and the mobile devices choose their desired average transmission rates by balancing their video quality and cost of transmission. Each mobile device then determines the specific video frames to transmit by a video summarization process. In the scheduling algorithm, we show that a time-division-multiplexing (TDM) based algorithm provides higher rates and lower distortions than the code-division-multiplexing (CDM) based transmission schemes currently deployed for voice and generic data traffic. In the TDM-based algorithm, the base station collects the information of frames to be transmitted from all devices within the current time window, sort them in the increasing order of deadlines, and schedule the transmissions in a TDM fashion. This joint algorithm takes advantage of the multi-user content diversity of the mobile devices, and maximize the network total utility (i.e., minimize the network total distortion), while satisfying all the delivery deadline constraints. Simulations show that the proposed algorithm significantly outperforms the constant rate provision algorithm in terms of minimizing users' distortions.

Key words: video streaming, pricing, uplink communications, CDMA, cross-layer design, video summarization
doi: **Document code:** **CLC number:**

1. INTRODUCTION

Video streaming is becoming one of the major driving forces for the next generation wireless networks. For the currently deployed cellular networks, the practical data rates are not enough to support full rate, high quality video applications. As a result, many research efforts have been devoted to adapting video content to reconcile the conflict between the high demand of video quality and the limited wireless communication resources among users. A large body of literature utilizes the cross-layer approach, which jointly designs the video coding in the application layer and the resource allocation in lower layers, e.g., [Zhang05], [Schaar03], [Zheng03], [Yoo04], [Zhao02], and the references therein.

To ensure the quality of real-time video streaming, smart video coding and adaptation techniques need to be performed at the application layer to meet the stringent resource constraints at lower layers. For very low bit rate channels, simply coding the video sequences at high quantization distortion levels is unpleasant to viewers. A better solution is to perform content-aware video coding, via summarization, which selects a subset of video frames that best represent the sequence, and encodes them at a higher quality. Various summarization techniques have recently been reported such as in [Liu05], [Li05a], and [Li05b]. The actual adaptation can be achieved through either transcoding or bit stream extraction from scalable video.

This paper develops, analyzes and simulates a new joint rate control and scheduling algorithm for uplink video streaming in CDMA cellular networks. The rate control part of the algorithm relies on adaptive content-aware video summarization, and utilizes a pricing-based approach to distribute the computational burden across the network (i.e., base station) and individual mobile users. The scheduling part of the algorithm let users transmit the summary frames in a time-division-multiplexing (TDM) fashion, which avoids excessive mutual interferences, achieve higher rate compared with a simultaneous equal rate transmission scheme, and fully utilizes the multi-user diversity to minimize users' total distortion.

Previously the pricing approach has been successfully used to efficiently allocate communication resources among elastic data applications in wireless networks (e.g., [Sara02], [Zhang01], [Lee02], [Huang04]), and distribute the computation burden among the network and end-users with limited information passing. We have previously shown that a pricing-based approach combined with adaptive video summarization techniques can greatly improve the performance of multi-user wireless downlink video transmissions [Li06]. This paper further extends the pricing framework to the uplink streaming case, which is more complex due to the interference limited nature of the communication channel.

Scheduling with multiple video users has been considered [?], [?], [?] (ZHU: please fill in several references here). Our contribution here is to show that a TDM-based transmission among video users leads to better performances compared with the scheme where video users also transmit simultaneously. Comparing with typical voice users, video users typically demand higher transmission rate, which lead to higher received power at the base station. This motivates the TDM based transmission among video users to avoid excessive interferences and improve the total achievable rate. Furthermore, the video sources are Variable Bit Rate (VBR) in nature. In the case of simultaneous transmission, a time-varying rate could be provisioned by frequent and precise power control among users, which lead to much signaling overhead between the base station and video users. The implementation of TDM transmission is much simpler, and the multi-user content diversity is naturally exploited to provide the best performance possible.

In the previous work on multi-user uplink video streaming at very low bit rate [Li05c], we try to control the admissible rate profile by iteratively adjusting peak rates and average rates among video users. The drawback is that the convergence is not guaranteed in general. The algorithm proposed here, however, has theoretical provable and practically very fast convergence.

The paper is organized as follows. We first introduce the system model in Section 2, and then describe the two-stage resource allocation algorithm in Sections 3 and 4, which includes price-based rate control and Time-division-multiplexing (TDM) based greedy scheduling. Simulation results are shown in Section 5, and we conclude in Section 6.

2. SYSTEM MODEL

The uplink capacity for the wideband CDMA system is interference limited [Tse05]. In the case of mixed voice and streaming video uplink transmissions, the objective is to provide the best possible Quality of Service to the video users, without interrupting the transmissions of voice users. This could be translated into a total received power constraint of the video users at the base station as explained next.

Consider a single CDMA cell with a set of M voice users and N video users. Assuming perfect power control, each voice user adjusts its transmission power to achieve the same received P_{voice} and the same SINR target γ_{voice} at the base station, thus the same transmission rate R_{voice}. A sufficient condition to achieve this is that the total received power of video users at the base station, P_{video}, should satisfy the following condition:

$$\frac{G_{voice}W}{R_{voice}} \frac{P_{voice}}{n_0 W + P_{video} + (M-1)P_{voice}} \geq \gamma_{voice}, \qquad (1)$$

where W is the channel bandwidth, G_{voice} is a constant that depends on voice users' common modulation scheme (e.g., $G_{voice}=1$ for BPSK), and n_0 is the background noise density (including both thermal noise and intra-cell interference). Based on (1), we can solve for the maximum feasiable value of P_{video}, denoted as P_{max}:

$$P_{\max} = \left(\frac{G_{voice}W}{R_{voice}\gamma_{voice}} + 1 - M \right) P_{voice} - n_0 W. \qquad (2)$$

This is the maximum total received power constraint from all video users in the cell.[1]

[1] For simplicity, we assume that each voice could achieve the same received power P_{voice}. For the more general treatment

In practice, the value of P_{max} could change over time to reflect the load change of voice. Here we only focus on resource allocation during a single time segment [0,T], which corresponds to the time window during which users perform one round of video coding and summarization. The typical value of T is around 3 secs**,** which is sufficient short to assume that the voice load and P_{max} do not change much.

Given the value of P_{max}, we want to allocate resource to the video users such that the total network utility is maximized. To be specific, we focus on a sliding time window, [0,T], and determine each video user *j*'s received power at the base station, $P_j(t)$, for transmitting its video contents in this time window. The Network Utility Maximization (NUM) problem is

$$\begin{aligned} &\max_{P_j(t)\geq 0, 1\leq j\leq N} \sum_{j=1}^{N} U_j\left(\int_0^T R_j(t)dt\right), \\ &s.t. \sum_{j=1}^{N} P_j(t) \leq P_{\max}, \forall t \in [0,T] \end{aligned} \tag{3}$$

Here $R_j(t)$ is the rate achieved by user j at time $t\in[0,T]$, and in general is a function of the received power allocation of all users, $\boldsymbol{P}(t)=[P_1(t),\ldots,P_N(t)]$. U_j is the utility function of user j and is increasing and strictly concave in the average transmission rate achieved during time [0,T]. (A example of the utility function will be given in Section 3.) Here we assume that each video user can achieve a peak received power equal to P_{max}[2].

Due to the time-varying nature of the video streaming contents, it is clear that the optimal solution of (3) includes time-varying power functions as well. Together with the fact that the utility functions U_j usually do not have analytical forms, finding the optimal solutions of (3) is quite difficult. To simplify the analysis, we consider a TDM-based transmission scheme, where the video frames of different users are transmitted one at a time without overlapping.

There are two major motivations of using the TDM-based transmission for video users. First, TDM transmission could avoid interferences among video users, who generate large received power at the base station due to the demand of much higher rate than the normal voice users. A more detailed discussion on this point can be found in [Kumaran03], where the authors show that in an uplink CDMA context where the objective is to maximize total weighted rate, it is optimal to schedule "strong" users to transmit one-at-a-time, and "weak" users to transmit simultaneously in larger groups. Here a "strong" user has a high peak received power, and a "weak" user has a low peak received power. In our context, video users are considered "strong" and voice users are considered "weak".

Second, the video sources are Variable Bit Rate (VBR) sources by nature. Provisioning a constant rate pipe for each video user without taking the actual video content into consideration would be a huge waste of resource. If users transmit simultaneously with rates changing on a frame-by-frame basis, then calculating the optimal transmission rate for each frame considering its unique size and deadline constraint, and jointly find the optimal power levels that achieve such rates for all users would be quite computational expensive and lead to much signaling overhead. On the other hand, the TDM-based transmission simplifies the scheduling problem by letting video frameworks to be transmitted at the highest rate possible, one by one, in the order of deadlines. This is very easy to implement since the frame sizes and deadlines for a given time segment are typically available at the beginning of the frame due to the buffering by the video users.

In the next two sections, we will explain the two-stage resource allocation algorithm based on the assumption of TDM-based transmissions.

3. STAGE I – PRICING BASED RATE CONTROL

In the first stage, we aim at allocating averaged transmission rate among users to maximum total utility. The delivery deadlines of video frames will be considered in the second stage. Based on the discussion above, we can rewrite (3) in the following form

$$\max_{t_j\geq 0, 1\leq j\leq N} \sum_{j=1}^{N} \tilde{U}_j(t_j), s.t. \sum_{j=1}^{N} t_j \leq T, \tag{4}$$

where

$$\tilde{U}_j(t_j) = U_j(R_{TDM} t_j). \tag{5}$$

Here R_{TDM} is the transmission achieved by letting

of different peak power constraints, see [Sampath95].

[2] For a user that can not achieve a received power equal to P_{max}, it might need to handoff to a closer base station with better channel gain.

only one user transmitting with a received power P_{max}. A user j's average transmission rate during time [0,T] is determined by the product of R_{TDM} and the active transmission time allocated to him, denoted by t_j.

Problem (4) could be solved by the standard dual decomposition technique. First relax the total time constraint by associating it with a dual price, λ, and then solving problem (4) is equivalent to maximizing the following Lagrangian, i.e.,

$$\max_{\boldsymbol{t}} J(\boldsymbol{t}, \lambda) = \sum_{j=1}^{N} \tilde{U}_j(t_j) - \lambda\left(\sum_{j=1}^{N} t_j - T\right), \qquad (6)$$

for some optimal nonnegative value λ. Here $\boldsymbol{t}=[t_1,\ldots,t_N]$.

Now let us first consider how to solve (6) for a fixed λ. We observe that the objective function in (6) can be decomposed into several sub-objective functions, one for each user. Thus (6) can be solved letting each user find an optimal value of $t_j(\lambda)$ such that

$$t_j(\lambda) = \arg\max_{t_j}\left(\tilde{U}_j(t_j) - \lambda t_j\right). \qquad (7)$$

The next question is how to find the value $t_j(\lambda)$ in (7). This requires more detailed information on how the utility function is defined. Although any formulation that makes the utility functions increasing and concave would work, here we focus on the case where the utility function is defined on the adapted video quality in terms of summarization distortion level.

Consider a sequence of n video frames $V = \{f_0, f_1, \ldots, f_{n-1}\}$, encoded at a pre-determined PSNR quality levels from some video clip. The SNR distortions introduced by the encoding preexist before transmission and thus are not considered in the optimization here. Further consider a summary of m frames $S = \{f_{l_0}, f_{l_1}, \ldots, f_{l_{m-1}}\}$ of the sequence V, where $m \le n$. The sequence S is then transmitted through the wireless channel. After receiving S (assuming error free), the receiver reconstructs the original sequence V as $V_S' = \{f_0', f_1', \ldots, f_{n-1}'\}$ by substituting the missing frames with the most recent frame that is in the summary S. The video summary quality, which is defined as the average distortion caused by the missing frames, is given as,

$$D(S) = \frac{1}{n}\sum_{k=0}^{n-1} d(f_k, f_k') \qquad (8)$$

where $d(f_k,f_k')$ is the distance between the kth original frame in V and the corresponding constructed frame in V'. Therefore, the optimization problem in (7) can be transformed into the problem of summarization with a price on total transmission time,

$$S_j^*(\lambda) = \arg\min_{S_j} D(S_j) + \lambda t_j(S_j). \qquad (9)$$

where the total transmission time l_j is a function of the resulting video summary bit rate. Eq. (9) can be solved with a Dynamic Programming approach at the video sources, more detail can be found in [Li05b].

It is known from information theory [Cover91] that a variety of practical signal sources have convex rate-distortion function. This is also the case empirically for the rate-summarization distortion function, $D(S_j)$, as shown in [Li05a]. . As a result, the utility U_j is an increasing and strictly concave function in the rate, so is $\tilde{U}_j(t_j)$ in transmission time t_j. An example of the rate-distortion tradeoff curve is plotted in Figure **1**, where the video sequence corresponds to frames 150-239 from the "foreman" sequence. The per frame distortion is averaged over all 90 frames.

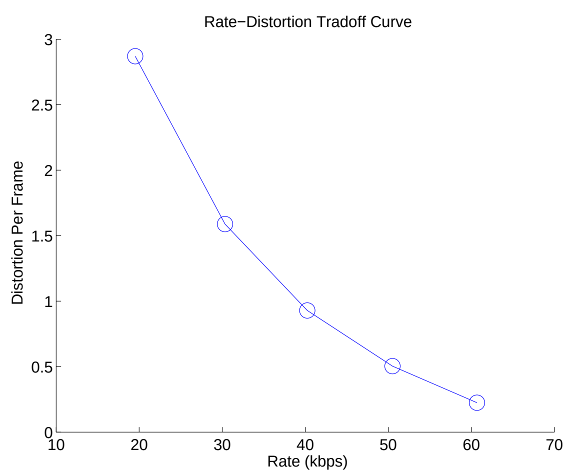

Figure 1: Rate-distortion tradeoff curve of frames 150-239 from the "foreman" sequence in a 3 sec time segment

Once $S_j^*(\lambda)$ is found, the corresponding transmission time $t_j(S_j(\lambda))$ can be computed assuming a rate of R_{TDM}. Each user j sends the value of $t_j(S_j(\lambda))$ to the base station, which wants to solve the following dual problem

$$\max_{\lambda \geq 0} J(\boldsymbol{t}(\lambda), \lambda), \tag{10}$$

This can be solved by a subgradient method, where the subgradient is determined by the degree of violation of the resource constraint. To be specific, the price λ could be updated according to the following,

$$\lambda^{i+1} = \max\left\{0, \lambda^i + \alpha^i\left[\sum_{j=1}^{N} t_j\left(S_j\left(\lambda^i\right)\right) - T\right]\right\}. \tag{11}$$

where α^j is the step-size at price iteration *i*. In (11), if the requested total transmission time is larger than *T*, the price is increased in the next iteration and vice versa for the case when requested total time is below T.

Proposition 1*: If the step-sizes satisfy* $\lim_{i\to\infty} \alpha^i = 0$ *and* $\sum_i \alpha^i \to \infty$, *then updates (7) to (11) converge to the optimal solution of problem (4).*

The proposition can be shown by similar techniques [Srikant04] and is omitted here.

During stage I, we vertically decompose the NUM problem in (3) into video summarization problem (6) that can be solved in the application layer, and rate control problem (10) that can be solved in the transport layer. Furthermore, horizontal decomposition is used such that problem (6) can be solved in a distributed fashion by letting each video user solve a subproblem (7).

Let's denote the price that corresponds to the optimal solution of (4) as λ^*, then the resulting $\{S_j(\lambda^*)\}$ or $\{t_j(\lambda^*)\}$ are just indication of the resource consumption levels for delivering certain level of utility for each user. The actual transmission schedule of individual frames is computed by the scheduling algorithm in stage II.

4. STAGE II – TDM GREEDY SCHEDULING

To ensure the satisfying reception of the video streaming application, each video summary frame has to be delivered to the receiver before a certain deadline. The pricing-based rate control algorithm leads to an "optimal" averaged rate allocation without considering the deadlines of video summary frames. The GREEDY scheduling algorithm in this section targets at meeting all the deadline requirements.

Since here we only consider the uplink transmission, we would like to upper-bound the delay of the frames received at the base station. The deadline constrained downlink video streaming has been considered in [Li06], which can be jointly used with the approach in this paper to provide bounded end-to-end delay guarantees.

The delivery deadline of each summary frame is determined by three components: the initial delay of all frames F_{ini}, its position delay $F_{j,position}$, and the total length of the current time segment *T*.

The initial delay F_{ini} is typically determined by the sizes of users' first frames. Since each user's first frame in a time segment is intra-coded, and typically has large size and has to be included in the summary frames. For these frames, their deadlines equal to the values of initial delay, since the values of $F_{j,position}$ for these frames are zero. For other summary frames, $F_{j,position}$ is determined as follows. If frame *j* is the 40th frame in its original source sequence *V*, and the sampling rate of the frames is 30Hz, then $F_{j,position}$=(45-1)/30sec. The minus one here is due to the fact we count time 0 at the position of the first frame. Finally, the entire summary frames need to be delivered within the current time segment, i.e., not interfering with the transmission of the next time segment. Thus the deadline of packet *j* is

$$T^j = \min\left(F_{ini} + F_{j,position}, T\right). \tag{12}$$

Assume all video frames from different users to be transmitted in time segment [0,T] are available at time 0, and the video users communicate the individual frame sizes and deadlines to the base station. The GREEDY algorithm works as follows. The base station first sorts the frames in the increasing order of the delivery deadline. The base station transmits the frames one-at-a-time, with constant rate R_{TDM}, such that the video SINRs meet the target value

$$\frac{G_{video} W}{R_{TDM}} \frac{P_{\max}}{n_0 W + M P_{voice}} \geq \gamma_{video}, \tag{13}$$

Here G_{video} is a constant that depends on the modulation scheme deployed for the voice traffic (e.g., $G_{video}=2$ for QPSK). Assume the *k*-th frame in the sorted sequence (containing summary frames from all users) has a frame size and delivery deadline $\{B^k, T^k\}$, it then takes a time equal to B^k/R_{TDM} to transmit this frame.

Although the GREEDY algorithm is simple, it is optimal among all TDM-based algorithms:

Proposition 2*: If any TDM-based scheduling algorithm can meet the deadlines of all video frames, the GREEDY scheduling algorithm also can.*

Proposition 2 can be proved as follows: pick any TDM-based scheduling algorithm where all deadlines are met and one or more packets are transmitted out of the deadline order. Then by rearranging the corresponding out of order packets by the deadline as in the GREEDY algorithm, all the deadline constraints are still satisfied.

It is also not difficult to show that if no TDM-based scheduling algorithm can meet all deadline constraints, then the GREEDY algorithm incurs the least deadline violation. To formally state the result, let us define

$$\Delta^{\Pi} = \max_k \left(T_{\Pi}^k - T^k\right) \qquad (14)$$

as the maximum delay violation under TDM-based scheduling policy Π, where T_{Π}^k denotes the actual delivery time of the *k*th packet under TDM-based algorithm Π. If $\Delta^{\Pi} \leq 0$, then all deadline constraints are met. We have

Proposition 3*: Among all TDM-based scheduling algorithms, the GREEDY algorithm yields the smallest value of* Δ^{Π}.

In fact, Proposition 2 is just a special case of Proposition 3, and the same proof technique can be generalized to prove the latter.

In the case of $\Delta^{GREEDY} > 0$, the base station needs to increase price so that video users request less rate. One way of adjusting price is the following

$$\lambda^{i+1} = \max\left\{0, \lambda^i + \beta \max\{\Delta^{GREEDY}\left(\lambda^i\right), 0\}\right\}, \qquad (15)$$

where β is a small step-size. In other words, the price is increased until the resulting frame sequences are schedulable (i.e., all deadline constraints can be met under GREEDY algorithm). There is a tradeoff between the value of β and convergence speed. If β is large, then the schedulability will be achieved by one or two adjustments; however, a significant portion of the time segment [0,T] might be wasted. If β is small, then it might take a longer time to achieve schedulability, but the resource utilization will be high. In either case, since users' average transmission rates are decreasing with λ, the price adjustment process in (15) always converges.

5. SIMULATION RESULTS

We choose four different video clips with different content activity levels, similar as in [Li06]. Clips 1 and 2 are segments from the "foreman" sequence, frames 150-239 and frames 240-329, respectively. Clips 3 and 4 are frames 50-139 and 140-229 from the "mother-daughter" sequence, respectively. There are 90 frames within each video clip at a sampling frequency of 30Hz, which corresponds to a time segment of $T=3$secs. Besides the GREEDY scheduling algorithm, we also simulate a simultaneous transmission scheme with equal constant rate (SIMCONST), where all four video users are allowed to transmit simultaneously with equal rates. In other words, the received power from each of the video user is the same at the base station in the SIMTRANS scheme, and no scheduling is needed due to simultaneous transmission.

In Table 1, we list the simulation parameters that kept constant throughout this section. These values are just chosen for illustration purpose instead of from any current standard, and our proposed algorithm is applicable to any interference limited CDMA communication systems.

Item	Symbol	Value
Bandwidth	W	1.228MHz
Noise density	n_0	$8.3*10^{-7}$ mW/Hz
Voice target SINR	γ_{voice}	6dB
Voice modulation		BPSK
Voice received power	P_{voice}	1mW
Voice spreading gain	G_{voice}	128
Voice transmission rate	R_{voice}	9.6kbps
Video target SINR	γ_{video}	6dB
Video modulation		QPSK

Table 1 Simulation parameters

We first compare the video users' total achievable rates under GREEDY and SIMCONST algorithms for different voice user load. Under GREEDY, we plot the maximum rate achieved by allowing only one user transmitting. Under SIMCONST, we plot the total rate achieved by all four users. Figure **2** shows that video users' total achievable rate decreases with the number of voice users, and becomes zero when there are more than 31 voice users in the system. In other words, the system's ability of supporting video users depends heavily on the current voice load in the cell. It is also clear that GREEDY algorithm always outperforms SIMCONST, with a rate increase more than 200% when the voice load is low. In that case, the mutual interference among video users becomes the bottleneck in achieving high rate under SIMCONST.

As we argued in Section 2, a TDM-based transmission scheme among “strong” uplink users not only could achieve higher rate but also leads to efficient contents multiplexing. To illustrate this point, let us consider a cell with 28 voice users, and the GREEDY algorithm offers almost the same total rate as the SIMCONST. For the rest of the simulation, we will further assume the following parameters as in Table 2.

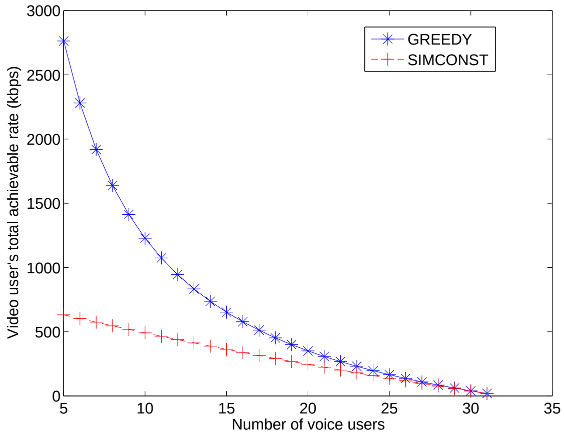

Figure 2: Total achievable rate comparison between GREEDY and SIMCONST

Item	Symbol	Value
Number of voice users	M	28
Maximum video received power	P_{max}	4mW
Video transmission rate under GREEDY algorithm	R_{TDM}	84.7kbps
Video transmission rate under constant rate transmission	R_{CR}	19.2kbps

Table 2 Further simulation parameters

First consider the pricing-based rate control algorithm. Based on the assumption of TDM scheduling, pricing on transmission time is equivalent to pricing on the achievable rate. Starting from an initial price λ=0.1, with diminishing step-sizes $\alpha^i=0.05/i$ that satisfy the conditions in Proposition 1. We terminate the iteration when the total transmission time of four video users achieves more than 99% of the time segment length (3sec here). Figure **3** shows the convergence of price with iterations, where the iterations converge in 6 iterations with a final optimal price $\lambda^*= 8.674*10^{-3}$.

Figure **4** shows how the summary distortion per frame of each individual user decreases (or increases) as the price decreases (or increases). Depending on the video contents that determine the specific rate-distortion functions, users experiences different levels of distortions under the same price. Among the four users, user 2 experiences the largest distortion due to the large temporal variations of its contents. Users 3 and 4 achieve similar distortions that are much smaller than users 1 and 2, due to the small time variations in the contents.

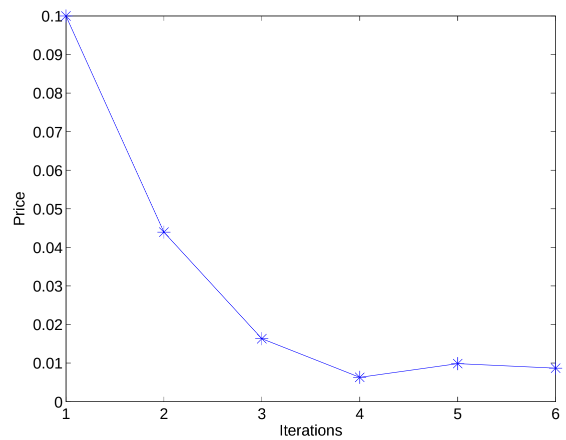

Figure 3: Pricing iteration

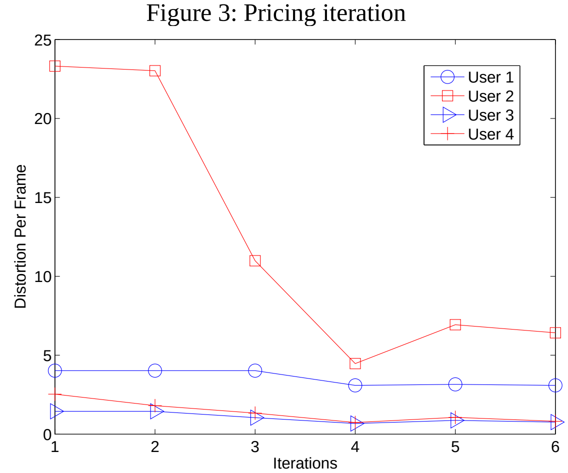

Figure 4: Distortion iteration

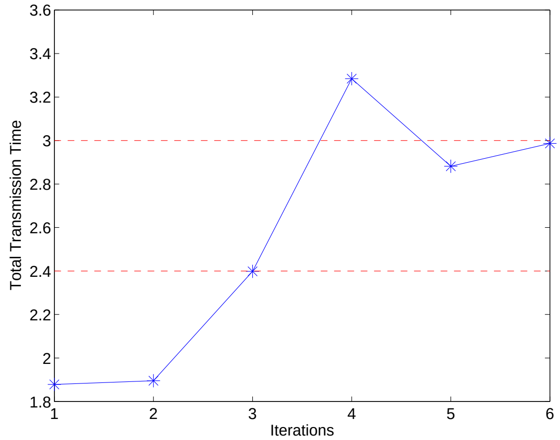

Figure 5: Transmission time iteration

Figure **5** shows how the total transmission time of the summarized packets changes during the price iteration. If we relax the convergence criterion from 99% to 80% (i.e., the price converges when it first enters the region bounded by the two dashed lines in Figure **5**), then the convergence time can be shortened by half. This reflects a trade-off between the computational complexity and resource utilization efficiency

The pricing algorithm will introduce additional delay into the system due to its iterative nature, and this delay can be leveraged by the having sufficient buffer at the transmitter of each mobile user. Assume that each video transmitter has enough buffers to hold all the packets for two time segments. During the transmissions of the frames for the first time segment, the price iteration and summarizations for the frames to be transmitted in the second time windows can take place in a distributed fashion at the base station and the mobile users. The iterative pricing process will not lead to any additional reception delay at the base station if the convergence can be achieved within the time segment length (e.g., 3 sec).

If there are not enough buffers available at the transmitter side, then a system designer needs to carefully choose the iteration parameters to tradeoff the convergence speed and performance. In general, the convergence speed of the pricing algorithm depends on the video contents, the initial price, the choice of step-size and the stopping criterion. Except the video contents that can be not adjusted by the system, all other factors can be continuously tuned based on practice to offer the best tradeoff of convergence and performance. Typically the requirement faster convergence inevitably leads to degraded performance since the resource (transmission time) can not be fully unutilized (e.g., reducing the stopping criterion from 99% to 85% as explained before). This tradeoff becomes more important as the number of video users increases. We want to emphasize that the convergence time does not increase linearly with the number of users, since most time consuming operation is the summarization process, which is performed by users in parallel.

The resulting video summary distortions based on the optimal price λ^* are plotted in Figure **6**. The vertical arrows indicate video summary frame locations in the sequence. Notice that the distortion is zero at summary frame locations, since the received frames are exactly the same as the original frames before summarization. The optimal pricing gives a good tradeoff between total transmitting power and total video summary distortion. Clips 1 and 2 are coded at an average PSNR of 27.8dB, and clips 3 and 4 at 31.0dB. The resulting average bit rates for 4 clips are *18.26*, *47.79*, *8.04* and *10.22*kbps, respectively.

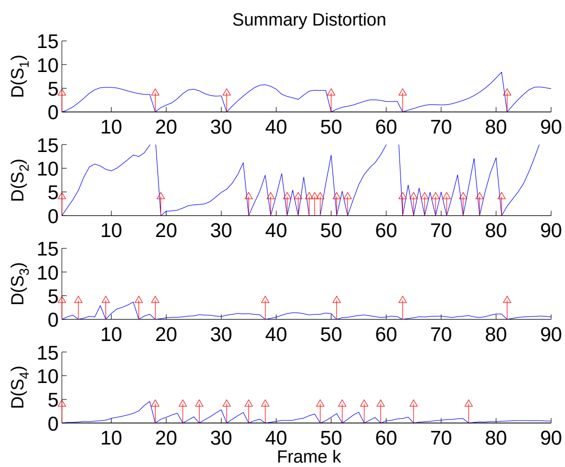

Figure 6: Resulting video summary distortion at the optimal pricing

Given the summarization results, the GREEDY algorithm performs scheduling based on sorted packet deadlines. The received power from each user over time as the result of the GREEDY algorithm is plotted in Figure **7**, and the corresponding delivery deadlines are plotted in Figure **8**. Under an initial delay of 30 frames (1sec), the GREEDY algorithm successfully transmits all packets within 3secs and meets all deadline requirements.

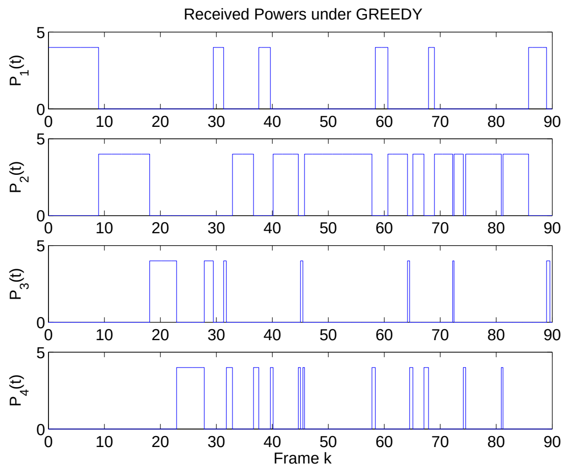

Figure 7: Video users' received powers at the base station under GREEDY scheduling algorithm (power is measured in mW)

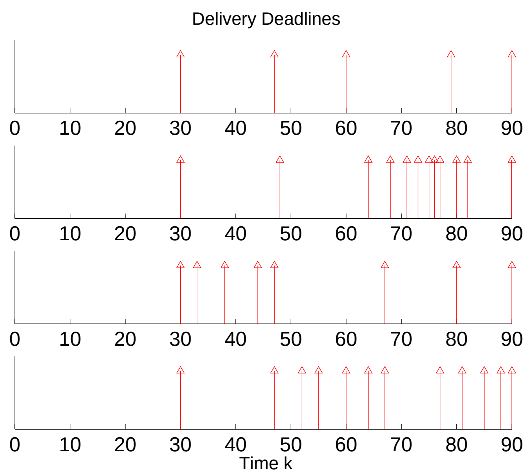

Figure 8: Frame delivery deadlines under GREEDY scheduling algorithm

As we mentioned in Section 4, if the current summary frames can not be scheduled (i.e., deadline violation occurs), then the base station needs to increase the price and let the users re-compute the summarization. However, in all the simulations that we perform, the summarization result from the pricing-based rate control is always schedulable. This is due to the fact that by taking advantage of the multi-user content diversity, the deadline requirements of the summary frames are typically spread out through the time segment, thus is relatively easy to satisfy. This implies that as long as there are enough content differences among the video users, the two stages of the algorithm can actually operated separately in practice. This avoids unnecessary iterations among the two stages and ensures fast convergence of the algorithm.

For comparison purpose, we also simulate the SIMCONST scheme, where all four video users are allowed transmitting simultaneously. Each video user can only generate a received power of 1mW at the base station. The base station can only guarantee a constant rate of *19.2*kbps for each user, so that each user meets the target SINR constraint of $\chi_{video}=6dB$. All other system parameters are the same as in the GREEDY case. The users will perform summarization based on the rate provided, so that all the summary frames can be transmitted within their individual deadline constraints. The resulting summary distortion is shown in Figure **9**.

The averaged distortions per frame for all users are *2.85*, *31.43*, *0.059* and *0.068*, respectively, with a total distortion per frame of *34.4*. As comparison, the averaged distortions per frame for all users achieved under optimal pricing are 3.09, *6.42*, *0.76* and *0.81*, respectively (see Figure **4**), with a total distortion per frame of *11.09*. Under SIMCONST, user 2 encounters a much larger distortion due to its busy contents and limited communication resource. As a result, the total distortion per frame increases more than *200%* from optimal pricing to SIMCONST.

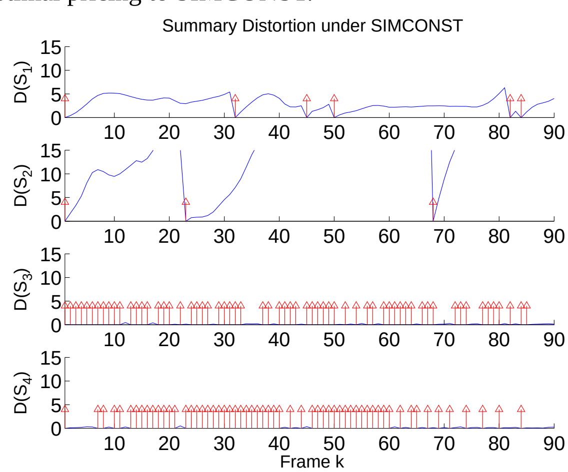

Figure 9: Resulting summary distortion under SIMCONST scheme

7. CONCLUSION AND FUTURE WORK

In this paper, we consider a cross-layer deign approach for uplink video streaming in a single

CDMA cell. The application layer (video coding), transport layer (rate control) and data link layer (scheduling) are jointly optimized. We propose a two-stage resource allocation scheme, which includes a price-based rate control algorithm and TDM-based GREEDY scheduling algorithm. In the rate control algorithm, the base station announces a price for the rate, and the mobile video users independently choose their average rate by performing optimal content-aware video summarization based on both the price and their utility functions. In other words, the operations in the application layer (video coding and summarization) and transport layer (rate control) are coupled only through a single price signal. In the data link layer, the base station performs TDM-based GREEDY scheduling based on the deadlines of the summary frames, and adjusts the price if it is not schedulable. Simulation results show that it significantly improves the network utility compared with the constant rate transmission scheme.

This paper is one further step in designing the network protocols in the framework of "layering as optimization decomposition" [Chiang05a], where the network protocols are analyzed and systematically designed as distributed solutions to some global optimization problems in the form of Network Utility Optimization (NUM). This approach provides insight into what the protocols optimizes and structures of the network protocol stack. Here each layer corresponds to a decomposed subproblem of the original NUM, and the interfaces among layers are quantified as functions of the optimization variables coordinating the subproblems. This approach has been successfully used to study various cross-layer optimization problems, such as TCP/PHY interaction [Chiang05b] and TCP/IP interaction [Wang03]. This paper attempts to utilize the same framework to jointly design and optimize the application layer (video coding), transport layer (rate control) and data link layer (scheduling) for a multiple user video session.

References

[Chiang05a] M. Chiang, S. H. Low and R A. Calderbank, "Layering as Optimization Decomposition", to appear in Proceedings of IEEE, 2006

[Chiang05b] M. Chiang, "Balancing Transport and Physical Layers in Wireless Multihop Networks: Jointly Optimal Congestion Control and Power Control", IEEE J. Sel. Areas Comm., vol. 23, no. 1, pp. 104-116, Jan. 2005

[Cover91] T. Cover and J. Thomas, *Elements of Information Theory*, Wiley and Sons, New York, 1991.

[Huang04] J. Huang, R. Berry, M. L. Honig, "Auction-based Spectrum Sharing", Allerton Conference of Communication, Computing and Control, Sept. 2004

[Kumaran03] K. Kumaran and Lijun Qian, "Uplink scheduling in CDMA packet-data systems", IEEE INFOCOM, 2003

[Lee02] J. W. Lee, R. R. Mazumdar, and N. B. Shroff, "Downlink power allocation for multi-class CDMA wireless networks", INFOCOM'2002, N.Y., Jun. 2002.

[Li05a] Z. Li, A. Katsaggelos, G. Schuster and B. Gandhi, "Rate-Distortion Optimal Video Summary Generation", IEEE Trans. on Image Processing, vol. 14, no. 10, October, 2005

[Li05b] Z. Li, F. Zhai, and A. K. Katsaggelos, "Video Summarization for Energy Efficient Wireless Streaming", Proceedings of SPIE Visual Communication and Image Processing (VCIP), Beijing, China, 2005

[Li05c] Z. Li, Q. Cheng, A. K. Katsaggelos, and F. Ishtiaq, "Video Summarization and Transmission Adaptation for Very Low Bit Rate Multiuser Wireless Uplink Video Communication", Proceedings of IEEE International Workshop on Multimedia Signal Processing (MMSP), Shanghai, China, October, 2005

[Li06] Z. Li, J. Huang and A. K. Katsaggelos, "Pricing based collaborative multi-user video streaming over power constrained wireless down link", Proc. IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP), 2006

[Liu05] S. Liu, and C. J. Kuo, "Joint temporal-spatial bit allocation for video coding with dependency", IEEE Trans. on Circuits & System for Video Tech, vol. 15(1), pp. 15-26, Jan. 2005.

[Sampath95] A. Sampath, P. S. Kumar, J. M. Holtzman, "Power control and resource management for a multimedia CDMA wireless system", Proc. IEEE PIMRC, 1995

[Sara02] C. Saraydar, N. Mandayam, and D. Goodman, "Efficient Power Control via Pricing in Wireless Data Networks", IEEE Trans. On Wireless Communications, Vol. 50 (2), Feb., 2002

[Schaar03] M. van der Schaar, S. Krishnamachari, S. Choi and X. Xu, "Adaptive cross-layer protection strategies for robust scalable video transmission over 802.11 WLANs", IEEE Journal on Selected Areas in Communications, 2003

[Srikant04] R. Srikant, The Mathematics of Internet Congestion Control, Birkhauser, Boston, 2004

[Tse05] D. Tse and P. Viswanath, "Fundamentals of Wireless Communication", Cambridge University Press, May 2005

[Wang03] J. Wang, L. Li, S. H. Low, and J. C. Doyle, "Can TCP and shortest path routing maximize utility," IEEE INFOCOM, vol. 3, April 2003.

[Yates95] R. Yates, "A framework for uplink power control in cellular radio systems", IEEE J. Sel. Areas Commun., 13(7):1341-1348, September 1995

[Yoo04] T. Yoo, E. Setton, X. Zhu, A. Goldsmith and B Girod, "Cross-layer design for video streaming over wireless ad hoc networks", Multimedia Signal Processing Workshop (MMSP), 2004

[Zhang01] P. Zhang, S. Jordan, P. Liu, M. L. Honig, "Power control of voice users using pricing in wireless networks", ITCom'01, 2001

[Zhang05] Q. Zhang, W. Zhu, and Y.-Q. Zhang, "End-to-end QoS for video delivery over wireless internet", Proceedings of the IEEE, 2005

[Zhao02] S. Zhao, Z. Xiong, and X. Wang, "Joint error control and power allocation for video transmission over CDMA networks with multiuser detection", IEEE Trans. on Circuits & System for Video Tech, vol.12(6), June, 2002

[Zheng03] H. Zheng, "Optimizing wireless multimedia transmissions through cross layer design", IEEE ICME, 2003