Iterative Pricing-Based Rate Allocation for Video Streams With Fluctuating Bandwidth Availability

Meng Yang, *Student Member, IEEE*, Theodore Groves, Nanning Zheng, *Fellow, IEEE*, and Pamela Cosman, *Fellow, IEEE*

***Abstract*—We consider rate allocation for video users in the case where the available bandwidth fluctuates. Simply minimizing the objective distortion or optimizing the stability of video qualities does not optimize subjective quality. We formulate a utility-based solution, considering that a user's preference of video quality often varies over a range with upper and lower thresholds of quality. Our iterative pricing-based resource allocation procedure reallocates the bandwidth not only between different users within a time slot but also between different time slots, such that no user suffers quality degradation on average by participating in the multiplexing process. Experimental results show that, compared with equal resource allocation and existing rate allocation solutions, the subjective result becomes increasingly better with the increase of bandwidth fluctuation rate or bandwidth fluctuation range. Moreover, as the number of users increases, the results improve.**

***Index Terms*—Cognitive radios, H.264/AVC, pareto optimality, rate allocation, video compression.**

I. Introduction

RESOURCE allocation for multiple video streams, often referred to as video multiplexing, concerns situations in which a limited resource (bandwidth in Hz or channel capacity in bits/sec) must be allocated to a set of video users. Prior work has shown that proper bandwidth scheduling benefits the average objective video quality of all users in the constant bandwidth scenario, compared with equal bandwidth allocation among users [1]–[6]. The goal is often to minimize the average distortion across all users (MINAVE) [5] or the weighted distortion [7], which in general means that users with high complexity scenes will be allocated more bits than those with low complexity scenes. Pricing-based mechanisms for resource allocation were extensively studied in recent years [8]–[12]. In some of these solutions, the quality of some users is often improved by sacrificing the quality of other users. Unlike these solutions, the equilibrium pricing method [6], [13] optimizes the quality of each video user, so any potential user will not suffer quality degradation (compared to equal resource allocation) by voluntarily participating in the multiplexing process. Some researchers have tried to minimize the overall variance of the videos (MINVAR), aiming at steady quality [14], [15]. However, this solution is achieved by increasing the overall distortion [3], which goes against the usual goal of many other video multiplexing studies. Tagliasacchi *et al.* exploited the correlation between the MINAVE and MINVAR problems in [3], achieving a compromise between them. Some researchers considered other goals, such as minimizing the distortion of the worst-case user [16], constant distortion ratio [7], or allocating bit rate to each user according to their video complexity [17]–[19]. With all of these different goals, the distortion has typically been measured by mean squared error (MSE).

In this paper, we consider scenarios where the bandwidth available to the group of users can vary substantially over time. One situation where this can occur is cognitive radio (CR). In a CR system, primary users can begin transmission at any time and have guaranteed use of the system resources. In contrast, secondary users can opportunistically make use of whatever resources are not being used by the primaries. The bandwidth available to the secondaries can fluctuate dramatically. Even though the overall distortion or stability of videos is optimized, it is possible that, for all secondary users, the quality of video is very low for some period of time when the available CR bandwidth is scarce. In this case, a better solution might be to give acceptable quality to each secondary user for as much time as possible, rather than sharing very low quality for some period of time and very high quality during some other period of time. So we consider this problem from a simple subjective point of view. Subjectively a user's preference of video quality varies over a range defined by upper and lower thresholds of quality.

The upper threshold of video quality (lower threshold of video distortion), occurs because additional quality beyond a certain point is not necessary for the viewer to be satisfied with watching the video. Indeed, it is clear that a quality increase above some threshold is quite imperceptible, as having a single pixel differ by one amplitude unit for 24-bit color cannot be seen, and so correcting that pixel to allow lossless quality provides no benefit. This paper does not attempt to determine the distortion threshold at which further distortion decrease is of no additional value, but just assumes its existence, and takes

Manuscript received April 26, 2013; revised September 25, 2013 and March 06, 2014; accepted June 04, 2014. Date of publication July 29, 2014; date of current version October 13, 2014. This research was supported in part by the NSF under Grant CCF-1160832, in part by Program 973 No. 2012CB316400, and in part by the NSFC Key Project No. 61231018. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Klara Nahrstedt. *(Corresponding author: Nanning Zheng.)*

M. Yang and N. Zheng are with the Department of Automation, Xi'an Jiaotong University, Xi'an 710049, China (e-mail: myang59@gmail.com; nnzheng@mail.xjtu.edu.cn).

T. Groves is with the Department of Economics, University of California at San Diego, La Jolla, CA 92093-0508 USA (e-mail: tgroves@ucsd.edu).

P. Cosman is with the Department of Electrical and Computer Engineering, University of California at San Diego, La Jolla, CA 92093-0407 USA (e-mail: pcosman@ece.ucsd.edu).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TMM.2014.2343943

1520-9210 © 2014 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission.
See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.

the video utility to be equal to one (maximum) if the distortion is below some threshold. By not seeking to reduce the distortion below the threshold, resources can be freed up for other users. This resource reallocating problem over users and over time slots can be solved with the pricing-based method in exchange economics, similar to [6] and [13].

In a similar manner, the video utility is set to zero when the video distortion is higher than some threshold. When the distortion is very high, additional distortion does not incur additional penalty, as the user considers the video useless already. At such a point, these users can temporarily not transmit (frame freeze) in a time slot to free up resources for others.

Both the upper threshold and lower threshold of video quality would have to be determined empirically by users' preferences on video quality, and would likely depend on spatial resolution and viewing conditions, among other factors. Given the overall bandwidth availability, a lower value for the upper quality threshold means more bandwidth resources will be freed up for other users and time slots. The lower threshold of video quality is related to the rate of frame freezes. Raising the lower threshold of video quality will free up more resources for others, however, it will also increase the frequency of frame freezes. The selection of the lower threshold would need to reflect the users' preference of video quality as a tradeoff between frequency of frame freezes and average quality when not frozen.

We are concerned here with applications where a user might accept frame freezes, if it allows higher quality elsewhere on the average. We consider, for example, using Skype for a video chat. Such sessions often have periods of freezes or of very low quality, but as long as the audio component continues throughout at high quality, the user is generally content, because the alternative is usually a conventional phone call with no video component at all. So if there are some periods of high quality and most periods of acceptable quality, the session is clearly better than the alternative, even with freezes. Other video applications which fit this paradigm include lectures and various types of instructional materials, news broadcasts, music videos, etc. Users might prefer having acceptable video quality most of the time with the occasional freeze, rather than having very low quality for some period of time and very high quality during some other period of time. That is, when the number of users is high relative to the total resource available, it might be preferable for all users to reallocate the resource in a highly unequal manner within each time slot, so that most videos are defended with acceptable quality, rather than having the allocation be roughly equal.

In this paper, we formulate an iterative pricing-based resource allocation procedure to optimize the utility of each potential secondary user, in which the price-guided procedure is similar to those discussed in the economics literature [20]. Unlike existing rate allocation solutions [1]–[5] and our previous work [6], [13] which consider a constant bandwidth scenario, we consider secondary users in a CR system, so available bandwidth fluctuates according to the utilization of the primary users. Unlike traditional rate allocation solutions [1], [3], [5] and our previous work [6], [13], which try to minimize the objective distortion or the variance of the video qualities of all users, we use a piecewise function with two thresholds to control the resource reallocation not only between different users of each time slot but also between different time slots over time. Unlike most traditional solutions [1]–[5], which often try to minimize the average distortion, we aim to find a Pareto optimal (PO) [21] bandwidth allocation for the secondary users such that each user is at least as well off as he would be with his initial endowment - an equal allocation of bandwidth to all secondary users. It may seem surprising that no user suffers quality degradation by participating in the multiplexing process, because it seems intuitive that when one user is better off, someone else must be worse off. It is true that in any one time period, one user getting higher allocation means another user must get lower allocation. However, when all time periods are considered together, each user gets higher allocation at times when it helps them most, and gets lower allocation when it hurts them least. So overall, all users improve their quality on the average by participating in the multiplexing process. This counter-intuitive result was established in [6] and [13].

Our main contributions are the following: 1) Rather than minimizing objective MSE or the variance in MSE, we use a piecewise linear function of MSE with an upper threshold and lower threshold of video quality, which captures some simple but important perceptual threshold effects in users' preference, such that acceptable quality of video is given to as many time slots as possible for all users, rather than sharing very low quality of video some time and very high quality of video some other time. 2) Rather than considering a constant bandwidth supply, we consider the resource allocation problem for CR secondary users, for whom the available bandwidth fluctuates over a wide range. 3) We further consider the case of too many potential secondary users with limited bandwidth availability. In this scenario, it may be possible that not all the users can have acceptable quality of video. Then only some users can be active in order to guarantee acceptable quality. The rest of this paper is organized as follows. In Section II, we review the exchange economics theory for video applications. In Section III, we provide details on our solution of the resource allocation problem. Simulation results are shown in Section IV, and conclusions are in Section V.

II. Utility-Based Economics Theory for Video Applications

A. Utility-Based Economics Theory

In the theory of exchange economics, users trade their endowments among themselves in a market for mutual advantage. Our goal, as in [6], [13], is that users in the market will exploit all potential gains from trade by participating in voluntary exchanges with others. Interested readers can read [13], [21] for more details. We briefly describe the method here. We consider $N, (i = 1, 2, \ldots, N)$ users and $T, (t = 1, 2, \ldots, T)$ goods. User i's consumption vector is $x_i = (x_i^1, x_i^2, \ldots, x_i^T)$. User i is initially endowed with an amount c_i^t of each good t. The total endowment of good t is denoted by $c^t = \sum_{i=1}^{N} c_i^t$. An allocation x is an assignment of a non-negative consumption vector to each user: $x = (x_1^1, x_1^2, \ldots, x_1^T; x_2^1, \ldots, x_2^T; \ldots; x_N^1, \ldots, x_N^T)$. A feasible allocation is one such that $c^t = \sum_{i=1}^N x_i^t$ for all goods $t = 1, 2, \ldots, T$. A PO allocation x^* is a feasible allocation such that there is no other feasible allocation that would be preferred by all users. Thus, at a PO allocation it is not possible to make any user better off without making some other user worse off. Given the default allocation of providing every user an equal amount of bandwidth, e.g. $1/N$ of the total available, in each period, we seek to find a procedure that will yield a PO allocation of bandwidth that is at least as good as the default allocation for every user.

User i's wealth is defined by the market value of his goods endowed initially. Supposing users can buy or sell good t in the market for price p^t, then the wealth of user i is $W_i = \sum_{t=1}^T p^t c_i^t$. For any consumption allocation x, the expenditure for each user should not exceed his wealth, that is $\sum_{t=1}^T p^t x_i^t \le W_i$.

Given an allocation x, the utility for user i can be denoted as $U_i(x_i^1, x_i^2, \ldots, x_i^T)$, which is an ordinal measure of the user's preferences or "level of satisfaction" with the consumption bundle x_i. When each good's contribution to utility is independent of every other good's contribution, we have

$$U_i\left(x_i^1, x_i^2, \ldots, x_i^M\right) = \sum_{t=1}^T U_i^t\left(x_i^t\right) \tag{1}$$

where $U_i^t(x_i^t)$ denotes the utility of good t for user i. We use this special form of the utility function in this paper to reflect our assumption that a user's preference of the quality of a video in any period does not depend on the video's quality in any other period.

In a competitive exchange economy, each user maximizes his utility subject to the budget constraint. A competitive equilibrium [21] is defined by a feasible allocation x_i^{t*} and a price system p^t such that the utility for each user is maximized, that is

$$\max_{x_i^t} \sum_{t=1}^T U_i^t\left(x_i^t\right), \quad (i = 1, 2, \ldots N) \tag{2}$$

subject to his budget constraint

$$\sum_{t=1}^T p^t x_i^t \le W_i = \sum_{t=1}^T p^t c_i^t, \quad (i = 1, 2, \ldots N) \tag{3}$$

for every user i, and the total demand equals the supply for each good t

$$\sum_{i=1}^N x_i^t = c^t, \quad (t = 1, 2 \ldots, T). \tag{4}$$

As stated above, the objective is to find an efficient allocation of bandwidth that is at least as good as the conventional allocation that provides each user $1/N$ of the available bandwidth in every period. We attempt to implement such a solution by finding a competitive equilibrium for which the initial endowment of the users is the equal allocation of bandwidth in every period.

B. Approximation for Video Applications

The model of N users and T goods in Section II.A can be applied to video bit rate allocation. We suppose there are N secondary video users in a CR system, and each video can be divided into T time slots (TS). As the H.264/AVC video standard is used in our simulation, we consider the slot to be one Group-of-Pictures (GOP). The bandwidth available in time slot t is the tth good. An allocation $x = (x_1^1, x_1^2, \ldots, x_1^T; x_2^1, \ldots, x_2^T; \ldots; x_N^1, \ldots, x_N^T)$ denotes the bandwidth resource allocation across all users and time slots. The resource (bandwidth) allocation optimization problem is the same as (2)–(4). Solving this problem, we would find the equilibrium prices p_t of bandwidth for each slot $t = 1, 2, \ldots, T$. However, for video bit rate allocation in real-time systems, one does not know the utilities of bandwidth and prices of all future slots in advance. In order to solve this problem, we considered the sequential process solution developed in [6], [13]. For each time slot t, each user tries to optimize his decision for the current slot and all future slots using expected values for future prices and future bandwidths. If the future slots are identical in expectation (for example, the future environment is perceived as stationary), then the optimization problem for each slot is a problem with two "slots" only, that is, to optimize the utility of the current slot t and the average expected utility of all future slots $t+1, \ldots, T$. We use x_i^t and $\bar{x}_i^t$ to denote the allocated resource for current slot t and the average future slot of user i, and we use p^t and $\bar{p}^t$ to denote the price of current slot t and the average future slot. Then the optimization problem is similar to (2)–(4), in which (2) becomes

$$\max_{x_i^t, \bar{x}_i^t} \left[U_i^t\left(x_i^t\right) + (T-t)\,\bar{U}_i^t\left(\bar{x}_i^t\right)\right], \quad (i = 1, 2, \ldots N) \tag{5}$$

and the constraint (3) becomes

$$p_t x_i^t + (T-t)\,\bar{p}_t \bar{x}_i^t \le W_i^t \tag{6}$$

where W_i^t denotes the wealth of user i before resource allocation is made for slot t. It is recursively calculated as

$$W_i^t = W_i^{t-1} - p_{t-1} x_i^{t-1}, \quad (i = 1, 2, \ldots N). \tag{7}$$

This solution can be considered as an approximation to the competitive equilibrium allocation for video application. With this solution, we aim to find an efficient allocation for all secondary users, such that each user is at least as well off as with their initial allocation. It is different from conventional rate allocation solutions [1]–[3], since they try to minimize the average distortion of all videos, even if that would make some users worse off compared to the case where they received $1/N$ of the available bandwidth in each period.

III. Pricing-Based Bit-Rate Allocation for Secondary Users

In this section, we provide details on solving the rate allocation problem of secondary users with the optimization solution from Section II.

A. Bandwidth Model for Secondary Users

We suppose there are P primary and S secondary users in a CR system, and the bandwidth resource provided for each primary user is R. For each primary user, the duration of each state (busy/idle) follows an exponential distribution [22]. When the

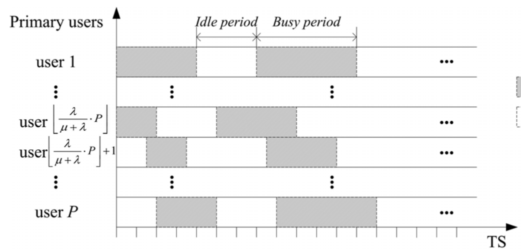

Fig. 1. Bandwidth model generation for our cognitive radio scenario.

state ends, the user switches to the opposite state. For simplicity, we suppose the time model for all the primary users is the same, with mean parameter λ for the busy time period and mean parameter μ for the idle time period. The available bandwidth for secondary users is the remaining resource that is not used by the primary users. This is shown in Fig. 1. We initialize the first $\lfloor \frac{\lambda}{\mu+\lambda} P \rfloor$ primary users as busy while the rest start idle. Then the average bandwidth resource for the secondary users over all time slots (denoted R_{ave} (kbits/TS)) can be estimated as follows:

$$R_{ave} = RP\left(1 - \frac{\lambda}{\mu+\lambda}\right) = \frac{\mu}{\mu+\lambda} RP. \tag{8}$$

In order to avoid the case that all bandwidth resource is utilized by primary users, we reserve a minimum bandwidth $R_{min} = 0.1RP$ for secondary users. So the available bandwidth fluctuates in the range $[R_{min}, R_{max}]$, where $R_{max} = R_{min} + RP$. As the busy/idle time periods can be longer than 1 time slot, the bandwidth model for secondary users has memory. That is, the amount of bandwidth available for secondary users in period t is not independent of the bandwidth available to secondary users in period $t - 1$. The parameters λ and μ can be used to change the memory of the model; larger values of λ and μ correspond to more memory and to slower bandwidth fluctuations. Two realizations are shown in Fig. 2(a)–(b), where the parameters are set to $\lambda = \mu = 1$ and $\lambda = \mu = 5$.

As baselines for comparison, we also consider the case where the bandwidth available to secondary users is constant at value R_{ave}, and the memoryless case where the bandwidth available in each time slot is generated uniformly and randomly in the range $[R_{min}, R_{max}]$. In this case, the average bandwidth for the secondary users is $R_{ave} = (R_{max} + R_{min})/2$. The fluctuation range of the bandwidth model is defined as $\delta_R = (R_{max} - R_{ave})/R_{ave}$. Fig. 2(c) shows a generated memoryless bandwidth model with fluctuation range $\delta_R = 0.3$.

B. Definition of Utility for Video Quality

Utility in economic consumer theory is an ordinal numerical measure reflecting a consumer's relative rankings of different allocations or consumption bundles. For video coding applications, a user's satisfaction is often directly measured by the negative of MSE. Then the rate allocation optimization problem for an individual user can be simply specified as minimizing the average MSE over time (MINAVE) [1], [4] subject to the budget constraint. Correspondingly, our optimization problem in (5)–(7) is equivalent to minimizing the average MSE of each secondary user, if the utility is defined as negative of MSE. This tends to work reasonably well for the constant bandwidth model, because the video quality is often steady enough over time, and the goals of minimizing MSE and minimizing MSE variance are not sharply in conflict. However, this does not hold for fluctuating networks.

In this work, we consider that a user's preference of video quality varies over a range defined by upper and lower thresholds of quality. The utility function is taken to be a piecewise function of distortion (measured by MSE), as shown in Fig. 3. That is

$$U(D) = \begin{cases} 1, & (D < D_1) \\ \frac{D_2 - D}{D_2 - D_1}, & (D_1 \le D \le D_2) \\ 0, & (D > D_2) \end{cases} \tag{9}$$

where D_1 and D_2 are predefined thresholds. For values of $D < D_1$, the distortion is so low that the user does not value any further reduction in distortion. So the utility is 1 for all distortions $D < D_1$, which allows the additional resource to be freed up for other users to lower their distortion. At the opposite extreme, if there is a time slot for a user for which the distortion D is too high, that is $D > D_2$, then the video is temporarily so bad that there is no additional penalty for additional distortion, in which case one might as well freeze the frame and thereby free up resources for other users. So all values of $D > D_2$ correspond to zero utility. When $D_1 \le D \le D_2$, that is the normal scenario, the utility maps linearly to the distortion. Here we use a linear mapping with MSE, although a different mapping or another quality metric could be used.

This utility function means that if any user were to be assigned too much or so few resources in a time slot that their distortion is beyond the range $[D_1, D_2]$, the additional or wasted resources for the current time slot is freed up completely and used for other users. For users not allocated zero, the quality is guaranteed not worse than threshold D_2, as well as not better than D_1. The time slots allocated zero experience a freeze, that is, their content is copied directly from the last frame of the previous slot. Then the optimization problem in (5)–(7) maximizes the utility of every secondary user subject to their budget constraint using the utility function defined in (9). However, with these utility functions, it is no longer necessarily true that a competitive equilibrium allocation is Pareto Optimal [21]. But we can conclude that this procedure for allocating bandwidth by defining prices for users to make mutually acceptable trades of bandwidth within a time slot and across time, will also result in an improvement in utility for all users. Any reallocation that would make a user worse off will not be made.

C. Estimation of $R \sim D$ Model for Current Slot and Average Future Slot

The utility function $U \sim D$ defined in Section III.B requires the rate-distortion ($R \sim D$) model of the current slot and the average $R \sim D$ model of all future slots for the optimization problem (5). For an encoded video stream, the MSE distortion typically decreases nonlinearly with an increase in bit rate. Different slots have different $R \sim D$ models. We update the $R \sim D$ model once per slot. For user i, $(1 \le i \le N)$ at time

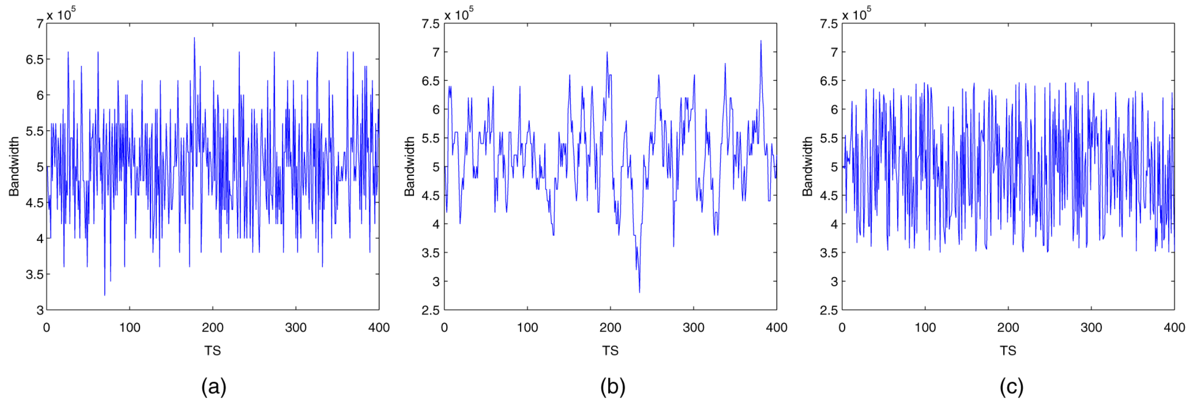

Fig. 2. Bandwidth models for secondary users. (a) CR model ($\lambda = \mu = 1$). (b) CR model ($\lambda = \mu = 5$). (c) Random model ($\delta_R = 0.3$).

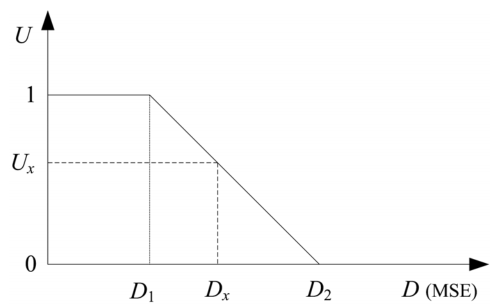

Fig. 3. Utility function.

slot t, $(1 \le t \le T)$, let D_i^t be the MSE distortion at bandwidth x_i^t. We use the common $R \sim D$ model [23], [24]

$$D_i^t\left(x_i^t\right) = a_i^t + \frac{b_i^t}{x_i^t + d_i^t} \tag{10}$$

where the parameters a_i^t, b_i^t, and d_i^t relate to the video complexity of the slot. The traditional least squares method is used for curve fitting to obtain the parameters. We first collect K RD pairs for slot t of user i: $(R_{i,1}^t, D_{i,1}^t)(R_{i,2}^t, D_{i,2}^t), \ldots, (R_{i,K}^t, D_{i,K}^t)$. From (10), we obtain

$$\left(D_i^t - a_i^t\right)\left(x_i^t + d_i^t\right) = b_i^t. \tag{11}$$

Using contiguous RD pairs $(R_{i,k}^j, D_{i,k}^j)$ and $(R_{i,k+1}^j, D_{i,k+1}^j)$, we obtain $K-1$ equations as follows:

$$\left(D_{i,k}^t - a_i^t\right)\left(R_{i,k}^t + d_i^t\right) = \left(D_{i,k+1}^t - a_i^t\right)\left(R_{i,k+1}^t + d_i^t\right). \tag{12}$$

Then by solving the set of equations with a least squares numerical method, we obtain the values of a_i^t and d_i^t for user i at time slot t. Then b_i^t can be estimated as follows:

$$b_i^t = \frac{1}{K}\sum_{k=1}^{K}\left(D_{i,k}^t - a_i^t\right)\left(R_{i,k}^t + d_i^t\right). \tag{13}$$

With the calculated coefficients a_i^t, b_i^t and d_i^t, the utility of the current slot $U_i^t(x_i^t)$ can be calculated by (9) and (10). We also need to predict the average utility of all future slots $\bar{U}_i^t(\bar{x}_i^t)$ to solve problem (5). To do this, we need to predict the average $R \sim D$ of future slots. Our simple prediction is to average over the $R \sim D$ information of all previous time slots, which was justified in [6]. The average future $R \sim D$ is estimated by averaging the individual coefficients (a, b, and d in (10)) separately. For current slot t of user i, the $R \sim D$ coefficients in (10) of averaged future slots are predicted as

$$\bar{a}_i^t = \frac{1}{t}\sum_{t'=1}^{t} a_i^{t'}, \quad \bar{b}_i^t = \frac{1}{t}\sum_{t'=1}^{t} b_i^{t'}, \quad \bar{d}_i^t = \frac{1}{t}\sum_{t'=1}^{t} d_i^{t'}. \tag{14}$$

The average utility $\bar{U}_i^t(\bar{x}_i^t)$ of all future slots can be calculated by (9), (10) and (14) with the predicted coefficients $\bar{a}_i^t$, $\bar{b}_i^t$, and $\bar{d}_i^t$, where $\bar{x}_i^t$ is the predicted average bandwidth for all future slots.

D. Utility-Based Resource Allocation for Multiple Video Streams

In order to solve the optimization problem in (5) with an inequality constraint (6) and an equality constraint (7), numerical methods can be used. However, the computational complexity of this natural problem is high. We try to find a closed-form solution of this problem. We first simplify the optimization problem (5)–(7). As the constraint (6) is an inequality, we solve it with dynamic programming [25]. We denote $W_i^t = V_i^t W_{step}$. Larger W_{step} means lower computational complexity with a lower precision result. We set $W_{step} = 100$ kbits in this paper. Then we obtain a set of optimization problems as follows:

$$\begin{aligned} &\max_{x_i^t, \bar{x}_i^t} \left[U_i^t\left(x_i^t\right) + (T-t)\,\bar{U}_i^t\left(\bar{x}_i^t\right)\right], \quad (i = 1, 2, \ldots N), \\ &s.t. \qquad p_t x_i^t + (T-t)\,\bar{p}_t \bar{x}_i^t \le w_i^t W_{step}, \end{aligned} \tag{15}$$

where $(w_i^t = 0, 1, \ldots, V_i^t)$. We denote the solution of (15) as $A(w_i^t)$. It has the following properties:

$$\begin{cases} A(0) = 0, \\ A\left(w_i^t\right) = \max\left\{A\left(w_i^t - 1\right), B\left(w_i^t\right)\right\}, \quad \left(w_i^t = 0, 1, \ldots, V_i^t\right) \end{cases} \tag{16}$$

where $B(w_i^t)$ is the solution of the following optimization problem with equality constraints:

$$\begin{aligned} &\max_{x_i^t, \bar{x}_i^t} \left[U_i^t\left(x_i^t\right) + (T-t)\,\bar{U}_i^t\left(\bar{x}_i^t\right)\right], \quad (i = 1, 2, \ldots N), \\ &s.t. \qquad p_t x_i^t + (T-t)\,\bar{p}_t \bar{x}_i^t = w_i^t W_{step}. \end{aligned} \tag{17}$$

Tabulating the results from $A(0)$ up through $A(V_i^t)$ gives the solution of the problem (15) for user i at time slot t. The next step is to solve the optimization problem (17), which has an equality constraint. As the target function (9) is not convex, some classical optimization methods, such as Lagrange multipliers, can not be used. It can be solved numerically with high complexity. From the Appendix, we obtain the closed-form solution of (17) as follows:

$$x_i^t\left(w_i^t\right) = \arg\max_{x_i^t \in X} U_i^t\left(x_i^t\right) + (T-t)\,\bar{U}_i^t\left(\frac{w_i^t W_{step} - p_t x_i^t}{(T-t)\,\bar{p}_t}\right) \tag{18}$$

where $U_i^t()$ and $\bar{U}_i^t()$ are defined in (29) and (30). The set X is defined as follows. If $\max\{X_2, X_4\} \leq X_5 \leq \min\{X_1, X_3\}$, $X = \{0, X_1, X_2, X_3, X_4, X_5\}$. Otherwise, $X = \{0, X_1, X_2, X_3, X_4\}$. Here $X_1, \ldots,$ and X_5 are defined in (19)–(23).

$$X_1 = \frac{b_i^t}{D_1 - a_i^t} - d_i^t, \tag{19}$$

$$X_2 = \frac{b_i^t}{D_2 - a_i^t} - d_i^t, \tag{20}$$

$$X_3 = \frac{w_i^t W_{step}}{p_t} - (T-t)\,\frac{\bar{p}_t}{p_t}\left(\frac{\bar{b}_i^t}{D_2 - \bar{a}_i^t} - \bar{d}_i^j\right) \tag{21}$$

$$X_4 = \frac{w_i^t W_{step}}{p_t} - (T-t)\,\frac{\bar{p}_t}{p_t}\left(\frac{\bar{b}_i^t}{D_1 - \bar{a}_i^t} - \bar{d}_i^j\right) \tag{22}$$

$$X_5 = \sqrt{\frac{b_i^t}{p_t}}\,\frac{w_i^t W_{step}\bar{p}_t + (T-t)\,\bar{d}_i^t\bar{p}_t + d_i^t p_t}{\sqrt{b_i^t p_t} + (T-t)\sqrt{\bar{b}_i^t \bar{p}_t}} - d_i^t. \tag{23}$$

Then by combining (16) and (18), we obtain the final result of the optimization problem (15) as follows:

$$x_i^{t*} = x_i^t\left(w_i^{t*}\right) \tag{24}$$

where

$$w_i^{t*} = \arg\max_{w_i^t}\left\{A\left(w_i^t\right), w_i^t = 0, 1, \ldots, V_i^t\right\}. \tag{25}$$

E. Iterative Pricing for Competitive Equilibrium

The solution in Section III.D is related to the price of the current slot p_t and that of the average future slot $\bar{p}_t$. We set $\bar{p}_t = 1$ because the results in (19)–(23) actually only depend on the price ratio $p_t/\bar{p}_t$. That is, the solution is homogeneous of degree zero in the prices $(p_t, \bar{p}_t)$ and this can be normalized by setting $\bar{p}_t = 1$. The price is optimal only when the total bandwidth of all secondary users in slot t equals the bandwidth resource of the current slot t (denoted $R_{t,total}$), and also the total average expected demands for future slots equals the average expected supply of future bandwidth, that is the constraint (4). So we can solve it in a centralized scenario with the constraint (4) to obtain the optimal price p_t, which we called the equilibrium price in [13]. In this solution, the computational burden of all users is shifted to the central server. In [6], we adjusted the price slot-by-slot in a decentralized scenario to lower the burden of the central server, an approach which we called 1BID. This solution suited the constant bandwidth model well, because the price of the current slot is related to that of the neighbor slots. However, for fluctuating bandwidth, prices are expected to fluctuate randomly and thus using the previous slot's price would not necessarily be expected to be an informative signal for the current price.

We consider an iterative pricing method, denoted as ITER. Considering the memoryless bandwidth model, the price of the current slot p_t has little relationship with that of the neighbor slots. So we try to find the optimal price for each slot separately. We set the initial price of each slot t to be $p_t^{(1)} = 1$. Then the price is iteratively adjusted by

$$p_t^{(k+1)} = p_t^{(k)}\left(1 + \delta_p \frac{\sum\limits_{i=1}^{N} x_i^{t*} - R_{t,total}}{R_{t,total}}\right) \tag{26}$$

until $\sum_{i=1}^{N} x_i^{t*} \approx R_{t,total}$ is met within 5% error for the current slot t, where k is the iteration index. That is the equilibrium price. In (26), x_i^{t*} is the bandwidth demand of user i at slot t calculated by (24), and $\delta_p, (0 \leq \delta_p \leq 1)$ is the iterative price adjustment parameter. It affects the speed of convergence for the price. An iterative pricing method and its convergence issues are discussed in [8]. Finally we normalize the resource demand of each user as

$$\hat{x}_i^{t*} = x_i^{t*}\frac{R_{t,total}}{\sum\limits_{i=1}^{N} x_i^{t*}}, \tag{27}$$

so that the total allocation to the users equals the total bandwidth resource $R_{t,total}$.

F. Slot Freeze Rate Control

From Section III.B, we know that the utility U of a user for some slots may equal 1 or 0 depending on the amount of bandwidth allocated to the user. The fraction of slots where $U = 1$ is called the slot saturation rate, and the fraction for which $U = 0$ is called the slot freeze rate. If the slot freeze rate is too high, it seems reasonable for some users to cease attempting to use the system and thus to drop out, or become inactive. When a user becomes inactive, whatever bandwidth is being allocated to him becomes available to the remaining *active* users, thus enabling their video quality to be improved. In this section we discuss how our procedure of the previous section is modified to take account of some users becoming inactive because their slot freeze rate is too large.

We suppose there are P primary and S secondary users, and the bandwidth resource for each primary user is R, with a total bandwidth resource of 1.1 RP. All S secondary users are initially active. After some period of time (50 TS in our simulation), we examine the slot freeze rates of all active secondary users in that period. If any of them exceed a predefined threshold T_{drop}, then the worst user is dropped from service. This is repeated until all the slot freeze rates of the active secondary users are below T_{drop}. The number of remaining active secondary users is denoted S_A. A simple example of the procedure is

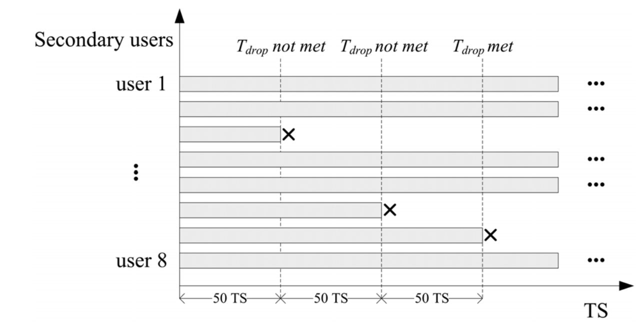

Fig. 4. Slot freeze rate control. At the end of each period, the worst user is dropped out until all active users meet T_{drop}.

shown in Fig. 4, in which $S = 8$, and $S_A = 5$. S_A is related to T_{drop} and to the total bandwidth resource.

IV. Experiments

The simulation was implemented using the baseline profile of H.264/AVC reference software JM16 [1] and MATLAB. The GOP size is 15 frames (I-P-P-P). The rate control module of H.264/AVC is activated. The test sequences for secondary users were taken from travel documentaries with substantial variation in scene complexity from one scene to another. The channel is assumed to be lossless. We chose 12 test sequences for simulation and each one is 6000 frames (400 slots) long at a resolution of 720×480. For every simulation, N secondary users are randomly chosen from them. N was set to 4 and 8. To estimate the video $R \sim D$ model for the current slot and the average future slot in Section III.C, $K = 20$ RD pairs were generated for rates ranging from 100 kbps to 2000 kbps.

The satisfaction for each secondary user is measured by the utility defined in (9). To provide results, we first obtain the utility value of each user at each slot, and transform it to MSE by the inverse function of (9) in the range $[D_1, D_2]$. Then the average MSE of each user for all time slots is taken and the PSNR is calculated (denoted UPSNR). Note that UPSNR here stands for the subjective measurement in our paper. An improvement in subjective measured utility means an improvement in UPSNR; we report results as UPSNR as the log units of PSNR are more familiar. The two thresholds D_1 and D_2 in (9) are chosen as 38 dB and 30 dB in this paper. The high threshold of 38 dB is taken to be the value above which further gains in quality provide no additional utility to the user. This choice of 38 dB might be valid for small screens and certain types of content, whereas a higher value of 43 dB might be valid for other viewing conditions or other content. The low threshold of 30 dB is taken to be the value below which further loss in quality incurs no further penalty, and the user can equivalently be shown a freeze. For some applications, this low threshold might be as low as 25 dB for example. In practical applications, the users' preference of video quality should be further investigated to determine the thresholds for specific scenarios. The average simulation bandwidth R_{ave} of the network ranged from 400 to 600 kbps per secondary user.

[1]http://iphome.hhi.de/suehring/tml

As discussed in Section III.A, λ and μ can change the time correlation or fluctuating rate of the bandwidth availability. We consider four bandwidth time correlation cases: $\lambda = \mu = 1$ and $\lambda = \mu = 5$ shown in Fig. 2(a)–(b), in which the bandwidth fluctuating rate of the former is lower than that of the latter, constant bandwidth model, and random bandwidth model shown in Fig. 2(c). Bandwidth fluctuation ranges $\delta = 0.2$, 0.4, and 0.6 are considered. Six combinations of these parameters are simulated. Case 1: Constant bandwidth; Case 2: CR $\lambda = \mu = 5$ and $\delta = 0.6$; Case 3: CR $\lambda = \mu = 1$ and $\delta = 0.6$; Case 4: Random bandwidth and $\delta = 0.6$; Case 5: Random bandwidth and $\delta = 0.4$; Case 6: Random bandwidth and $\delta = 0.2$. In this paper, the equal allocation method, denoted EQL, is the baseline. For each time slot, it equally allocates the available resource among all secondary users. We denote the MSE-based optimization solutions as MSE_ITER (for the iterative solution) and MSE_1BID (for the non-iterative solution) where the utility is simply the objective distortion MSE [6]. The proposed utility-based optimization solution is denoted UTILITY_ITER.

A. Comparison of Pricing Methods

We first examine MSE_1BID and MSE_ITER for different bandwidth models. The price adjustment parameter δ_p in (26) is set to 0.2 empirically. If δ_p is large, the price will converge fast but inaccurately. If δ_p is too small, the speed of convergence will be very slow. For our simulations, δ_p was not optimized for any set of video streams. It is known from Section III.E that the converged price of ITER is the equilibrium price. It is expected that the result of ITER is equal to or better than others.

For constant bandwidth, the available resource for each slot is the same. So the equilibrium price of the current slot is often highly related to that of the previous slot. In this case, it is expected that the MSE_1BID price, MSE_ITER price and UTILITY_ITER price would be similar, which the simulation results in Fig. 5(a)–(c) suggest is true. For fluctuating bandwidth, the equilibrium price for each slot is less correlated. Simulation results are shown in Fig. 6(a)–(c). The price fluctuates sharply together with the bandwidth.

We tested the objective PSNR results of MSE_1BID, MSE_ITER and EQL for different bandwidth models. That is to average MSE of each user over all time slots and then calculate the PSNR (denoted PSNR). The subjective UPSNR results are similar with the objective results, and will be shown later in next section. Fig. 7(a)–(d) show the results for the predefined bandwidth models Case1-Case4, which have different fluctuating rates. As expected, the improvement of MSE_ITER over MSE_1BID becomes larger for the network model with rapid fluctuations. The improvement of MSE_ITER is up to 0.9 dB comparing with EQL and up to 0.4 dB comparing with MSE_1BID for the high fluctuating rate case. As the resource availability becomes less correlated from slot-to-slot, the equilibrium price is also less correlated from one slot to another. Similarly, for the network models Case4-Case6 with different fluctuating ranges, a larger fluctuating range means lower correlation, so the result of MSE_ITER is much better than that of MSE_1BID and EQL. The results are shown in Fig. 7(d)–(f), whose fluctuating ranges are 0.6, 0.4 and 0.2 respectively.

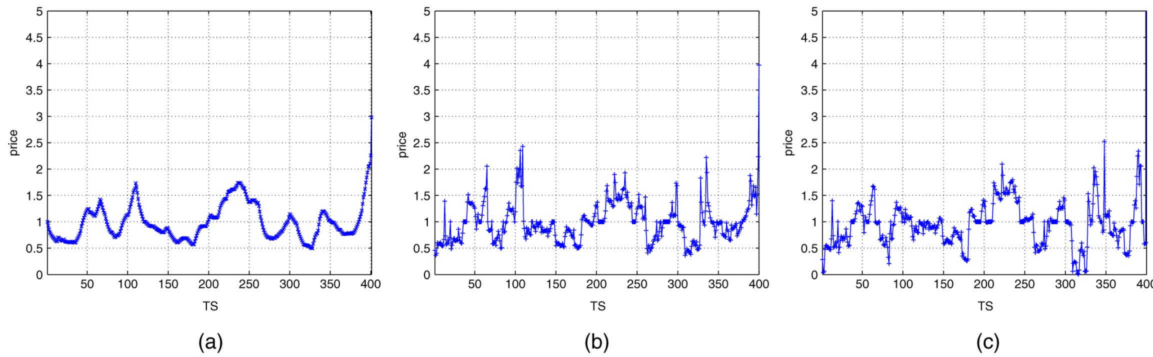

Fig. 5. Price for constant bandwidth model. (a) MSE_1BID pricing. (b) MSE_ITER pricing. (c) UTILITY_ITER pricing.

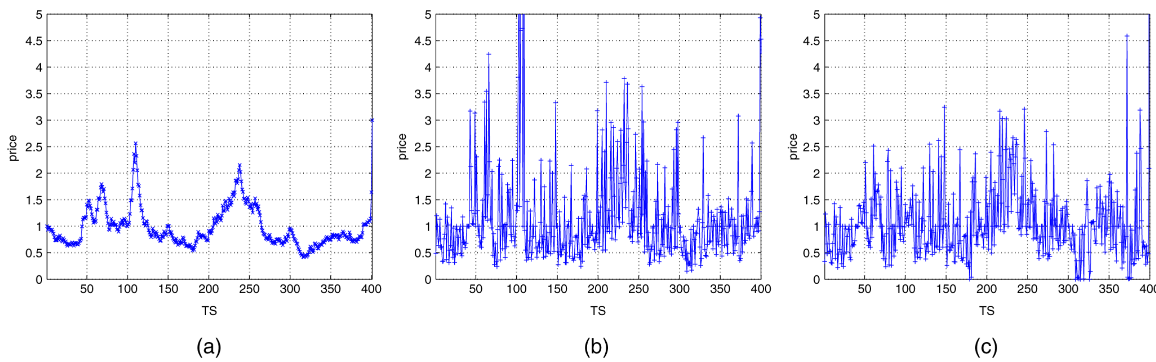

Fig. 6. Price for fluctuating bandwidth model in Fig. 2. (a) MSE_1BID pricing. (b) MSE_ITER pricing. (c) UTILITY_ITER pricing.

B. Result of Utility-Based Iterative Pricing

In Section III.D, we simplified the original utility-based optimization problem (5)–(7) (denoted ORIGINAL) with an inequality constraint (6) to (15) (denoted UNEQL_CONSTRAINT) and solved it with dynamic programming. However, the computational complexity for solving this problem is high. A further approximate solution of (15) is the problem with an equality constraint (17) by setting $w_i^t = V_i^t$, whose closed-form solution is shown in (18) (denoted EQL_CONSTRAINT). We averaged multiple simulation results of the closed-form solution EQL_CONSTRAINT, comparing with the simplified solution UNEQL_CONSTRAINT and the original problem ORIGINAL solved with numerical methods. One of the UPSNR-versus-bandwidth result is shown in Fig. 8. EQL denotes the baseline, which equally allocates the resource among all users for each time slot. The approximate solution is slightly worse than the original solution by 0-0.3 dB and slightly worse than the simplified solution by 0-0.1 dB. However, the average simulating times (by MATLAB) for EQL_CONSTRAINT, UNEQL_CONSTRAINT, and ORIGINAL are 0.03, 3.4, and 1352 (seconds) respectively. So in the following simulations, we only consider the approximate solution EQL_CONSTRAINT to accelerate the algorithms.

The utility-based solution in Section III.D is based on seeking a competitive allocation that will maximize the utility of each secondary user subject to his budget constraint. The following four solutions are compared for different bandwidth models. 1) UTILITY_ITER: The utility-based optimization solution in Section III.D is combined with iterative pricing (ITER) in Section III.E. 2) MSE_ITER: The MSE-based optimization solution in [6], which tries to optimize the global MSE of each user, is combined with iterative pricing. 3) MSE_1BID: The MSE-based optimization solution in [6] is combined with 1-bid pricing (1BID). 4) EQL: It is the baseline solution, which equally allocates the resource among all users for each time slot.

Fig. 9 shows the UPSNR results of these methods for different bandwidth models with different fluctuating rates and fluctuating ranges for four secondary users. The results for MSE_1BID, MSE_ITER and EQL are similar to those of Section IV.A. For UTILITY_ITER, it is expected that the result is much better than the other three methods for bandwidth models with higher fluctuating rate or fluctuating range. Fig. 9(a)–(d) shows the results for bandwidth models with different fluctuating rates. The result of UTILITY_ITER becomes increasingly better with the increase in fluctuating rate, compared with the other three solutions. Fig. 9(d)–(f) shows the results for bandwidth models with different fluctuating ranges. The result of UTILITY_ITER becomes increasingly better with the increase of the bandwidth fluctuation range, compared with other solutions. Table I shows the average UPSNR improvement of each method compared with EQL for different bandwidth models. The improvement of the proposed method is up to 1 dB when the bandwidth fluctuates sharply and frequently.

We additionally increase the number of secondary users N from 4 to 8. The results shown in Fig. 10 and Table II are similar to, but slightly better than, the case of $N = 4$ by 0.05-0.2 dB, because having more secondary users allows more efficient rate

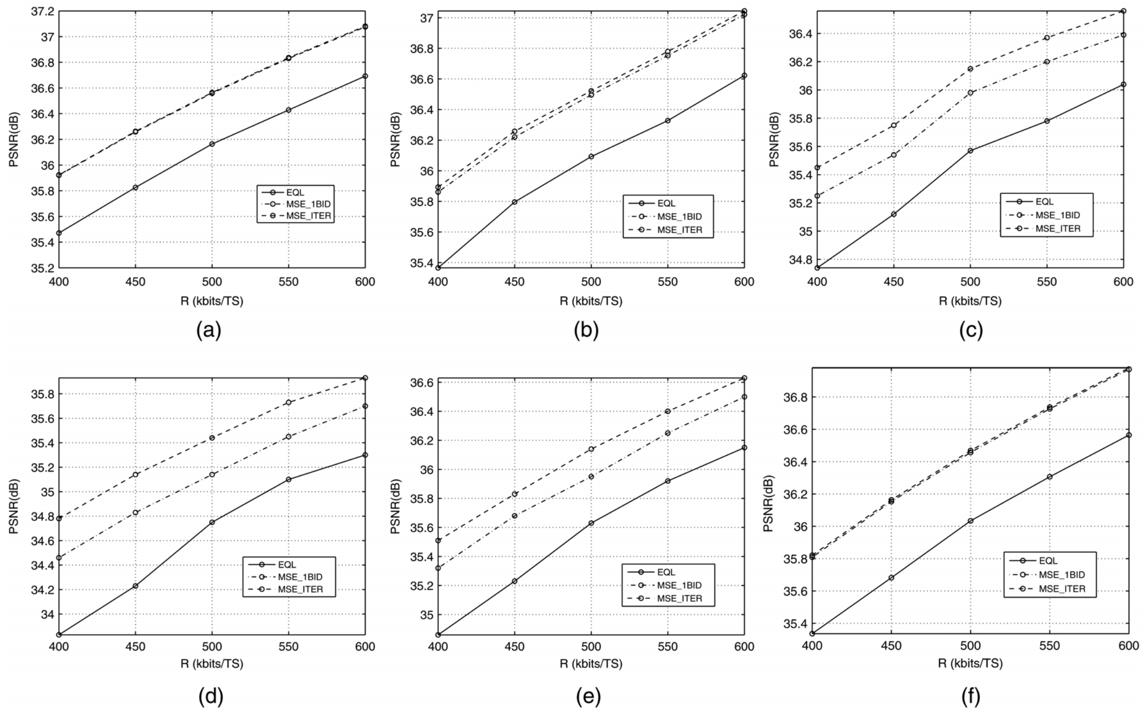

Fig. 7. Comparison of pricing methods for the bandwidth models Case 1 through Case 6. (a) Case 1. (b) Case 2. (c) Case 3. (d) Case 4. (e) Case 5. (f) Case 6.

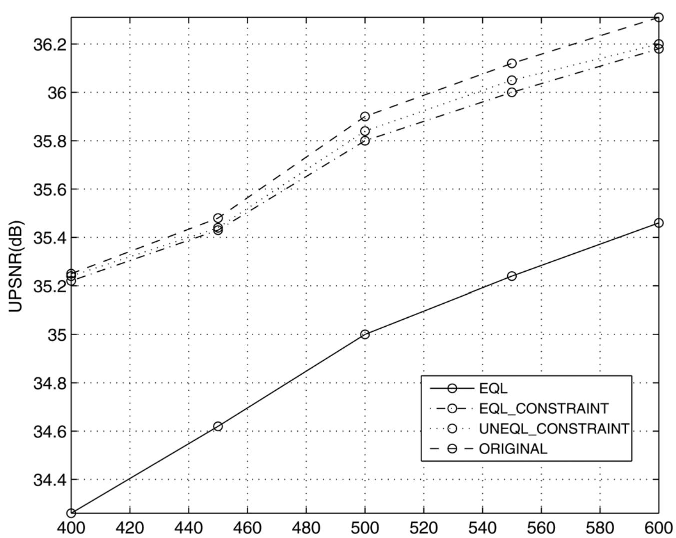

Fig. 8. Comparison of original solution, simplified solution, and closed-form solution.

allocation among users as there are more gains from trading bandwidth among users in any slot.

Besides the subjective test, we also test the usual objective result PSNR of our method. Here, we average MSE of each user over all time slots and then calculate the PSNR. One simple simulation result is shown in Fig. 11. The proposed method is still better than EQL by 0.5-0.7 dB but slightly worse than the method we presented in [6] (which minimized the objective quality MSE of videos) by 0.1-0.2 dB. So the proposed method which aims to maximize utility has only a very slight loss in conventional MSE compared to the method [6] which aims to maximize MSE; this slight loss is because utility maximization ignores changes in MSE at levels above and below the utility thresholds.

C. Simulation of Slot Freeze Rate Control

The effectiveness of the proposed bandwidth allocation approach mainly comes from the freed-up resource from the saturated slots. We tested the slot saturation rate for the simulation in Section IV.B with 4 secondary users, and some results are shown in Fig. 12. The slot saturation rate is up to 20%-50% for most cases. With the increase of the total available bandwidth, the slot saturation rate increases, which means that more resources are freed up. When bandwidth is scarce and the resources are not sufficient to give acceptable video quality to all users, some slots experience a freeze, which is the other mechanism for freeing up resources for others. While high slot saturation rate is not a concern (it means users are at their maximal utility), in practical applications, the slot freeze rate (slots where $U = 0$) should not be too high. Without using a mechanism to control the slot freeze rate, we simply examined the slot freeze rate for the simulation in Section IV.B with 4 secondary users. Some results are shown in Fig. 13; the slot freeze rate is less than 4% for most cases. By comparison, the rate of slots with quality below the lower threshold is often up to 30% for the baseline solution EQL. Note that frame freezes are not required by our scheme; the frame freeze rate can be controlled to as small as zero by decreasing the lower threshold, according to the users' preference of video quality in different conditions.

We further conducted a simple observer test to verify the visual effect of the frame freezes. Six videos at low bit rate (about 30 dB) were provided to ten users to watch, including four practical network videos (these were 6000 frames long at 30 fps with

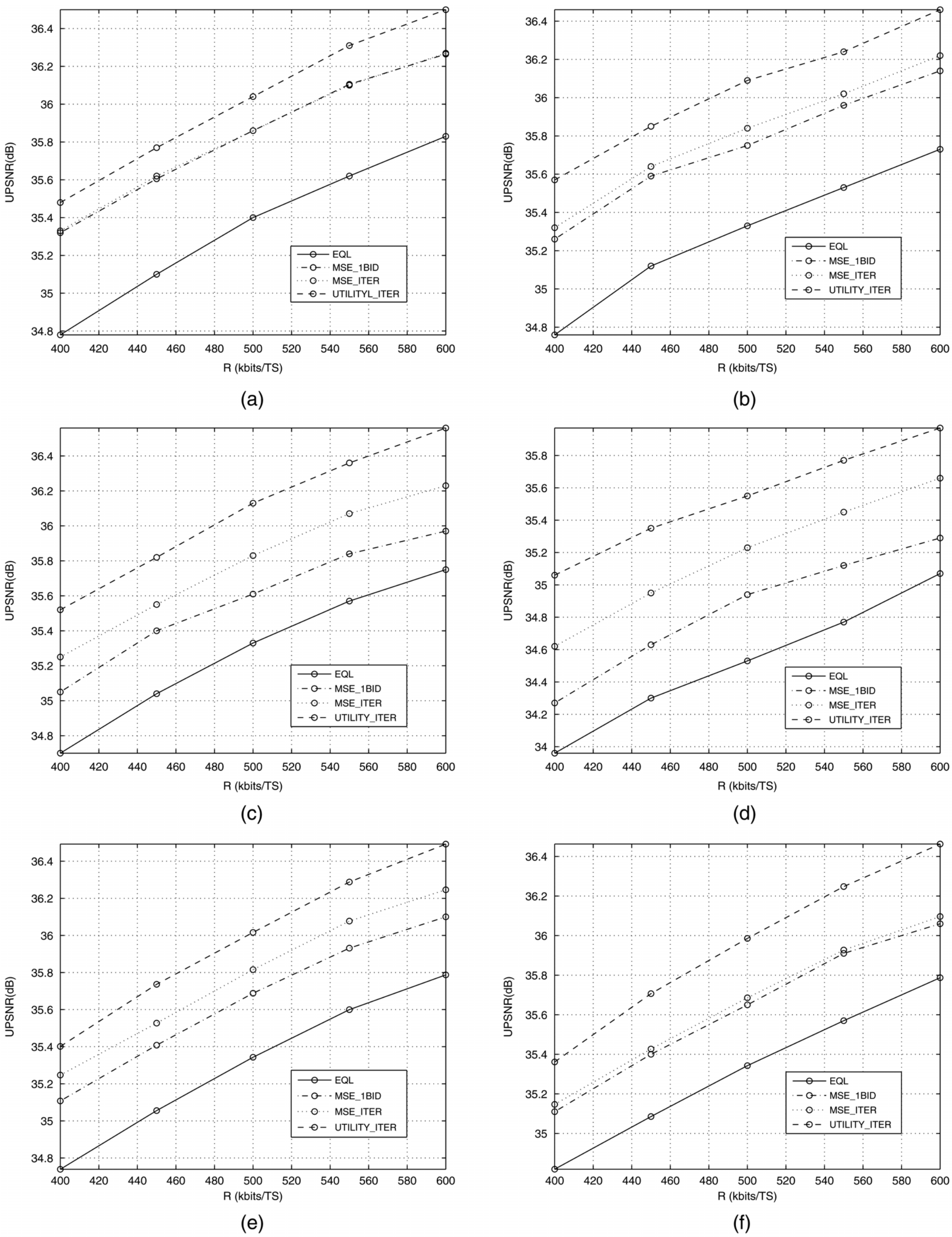

Fig. 9. Subjective results of utility-based pricing method for four secondary users. (a) Case 1. (b) Case 2. (c) Case 3. (d) Case 4. (e) Case 5. (f) Case 6.

TABLE I
AVERAGE UPSNR IMPROVEMENT (dB) OF EACH METHOD FOR FOUR SECONDARY USERS

	Total	Invisible	Low	Medium	High
Network videos	5.0%	1.4%	2.8%	0.7%	0.1%
Standard videos	5.0%	0.8%	3.1%	0.9%	0.2%

resolution 720×480), and two standard YUV videos ("Paris" and "News" which were 300 frames long at 30 fps with resolution 352×288). The GOP size was 15. Five percent of the slots were randomly selected to be frozen, which is the highest freeze rate in our simulation result. The videos were presented to users on a personal computer in a quiet room. The four network videos were presented with audio, and the audio continued throughout even when the video experienced a freeze. The other two YUV videos do not have audio information. Users were asked to think of the videos as on-line material, and rate the annoyance of each freeze they saw as 3 levels: low, medium, or high. Undetected freezes were considered invisible. Results averaged over videos and users are in Table III. Most freezes were either invisible or of low annoyance for this video content. Moreover, the annoyance of the videos with audio is lower than those without audio. We assume that users will accept a low rate of annoying freezes,

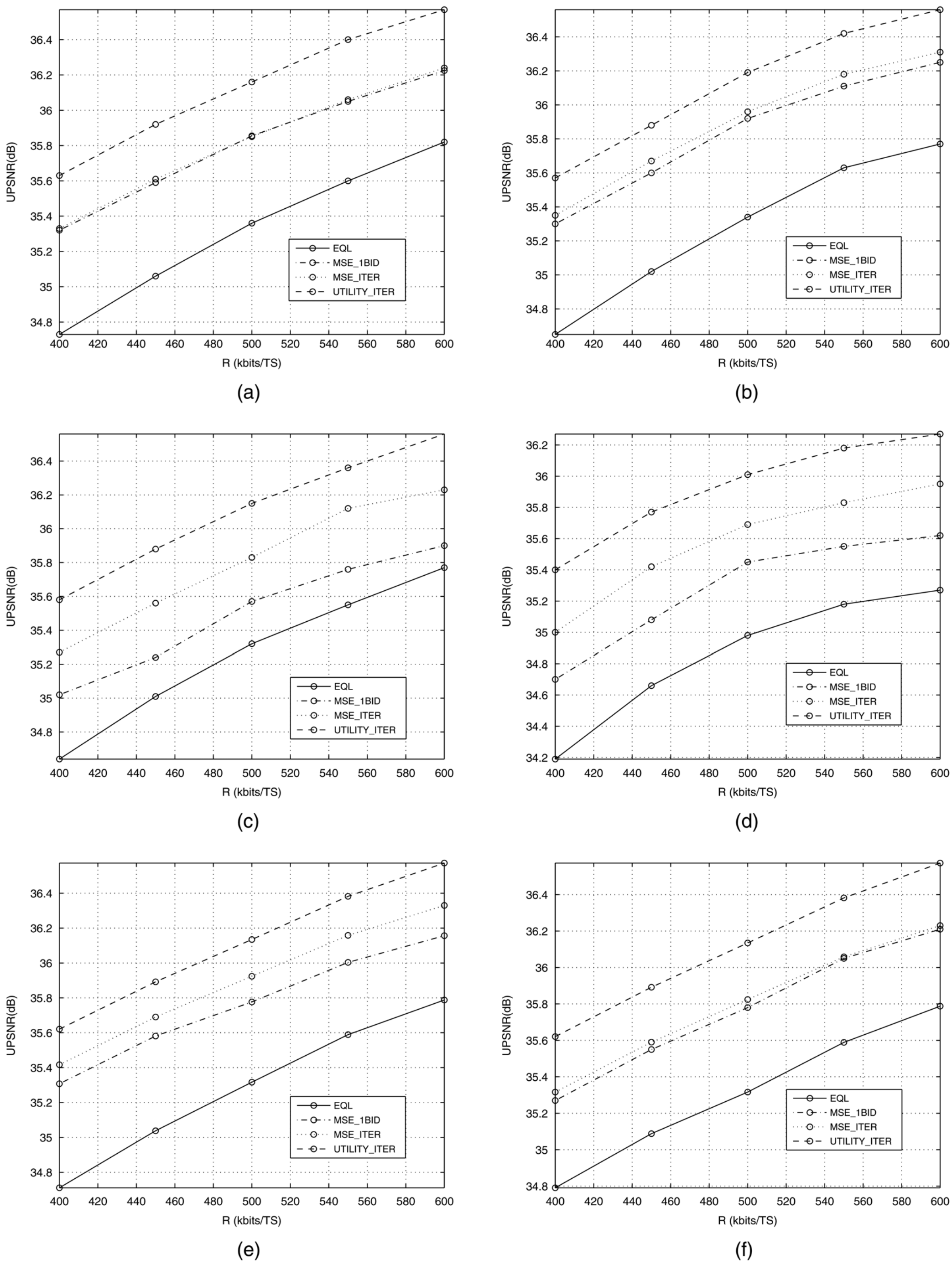

Fig. 10. Subjective results of utility-based pricing method for eight secondary users. (a) Case 1. (b) Case 2. (c) Case 3. (d) Case 4. (e) Case 5. (f) Case 6.

but above a certain point the video will be considered not worth watching.

So far, no user was dropped from service. From Fig. 13 we see that the slot freeze rate for secondary users increases with the decrease of the average bandwidth R_{ave}. If there are too many secondary users relative to the limited bandwidth resource in the CR networks, we can not guarantee the slot freeze rate to be acceptable. In this case, the mechanism described in Section III.F is implemented to drop some of the secondary users from service near the beginning until the predefined slot freeze rate is met.

V. Conclusion

In this paper, we considered the rate allocation problem of multiple secondary users in cognitive radios, for which the available bandwidth often fluctuates over a wide range. A utility-based iterative pricing procedure, which is derived from economics theory, is formulated to optimize the subjective video quality of each user, such that each user is at least as well off as his initial endowment. Experimental results show that the subjective result becomes increasingly better with the increase of the bandwidth fluctuating rate or the bandwidth fluctuating

TABLE II
AVERAGE UPSNR IMPROVEMENT (dB) OF EACH METHOD FOR EIGHT SECONDARY USERS

	Bandwidth Model					
	Case 1	Case 2	Case 3	Case 4	Case 5	Case 6
EQL	35.346	35.294	35.278	34.526	35.302	35.321
	Improvement over EQL					
MSE_1BID	0.485	0.446	0.296	0.324	0.346	0.305
MSE_ITER	0.490	0.514	0.508	0.656	0.482	0.336
UTILITY_ITER	**0.674**	**0.748**	**0.800**	**1.014**	**0.686**	**0.632**

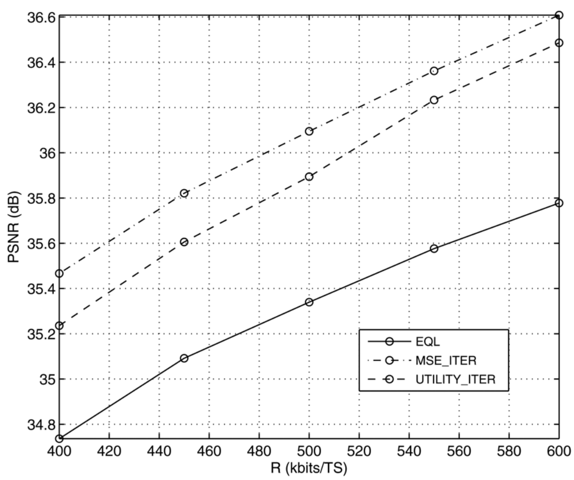

Fig. 11. Objective results comparison.

range, compared with that of equal resource allocation and existing rate allocation solutions. The results also improve with more secondary users. When too many secondary users share limited resources, dropping some users from service at the beginning can guarantee the quality of all active ones to be acceptable.

APPENDIX I

In this appendix, we solve the optimization problem (17) in Section III.D, and give its closed-form solution.

First, we aim to simplify the optimization problem. By substituting the utility function $U \sim D$ in (9) with the $R \sim D$ models in (10) and (14), we obtain the utility functions of the current slot t and the average future slot for user i as follows:

$$U_i^t\left(x_i^t\right)=\begin{cases}1, & \left(x_i^t>X_1\right)\\ \frac{\left(x_i^t+d_i^t\right)\left(D_2-a_i^t\right)-b_i^t}{\left(D_2-D_1\right)\left(x_i^t+d_i^t\right)}, & \left(X_2\le x_i^t\le X_1\right)\\ 0, & \left(x_i^t<X_2\right)\end{cases} \tag{28}$$

$$\bar{U}_i^t\left(\bar{x}_i^t\right)=\begin{cases}1, & \left(\bar{x}_i^t>\bar{X}_1\right)\\ \frac{\left(\bar{x}_i^t+\bar{d}_i^t\right)\left(D_2-\bar{a}_i^t\right)-\bar{b}_i^t}{\left(D_2-D_1\right)\left(\bar{x}_i^t+\bar{d}_i^t\right)}, & \left(\bar{X}_2\le \bar{x}_i^t\le \bar{X}_1\right)\\ 0, & \left(\bar{x}_i^t<\bar{X}_2\right)\end{cases} \tag{29}$$

The boundaries X_1, X_2, $\bar{X}_1$, and $\bar{X}_2$ in (28) and (29) can be calculated from (10), that is

$$X_1=\frac{b_i^t}{D_1-a_i^t}-d_i^t,\quad X_2=\frac{b_i^t}{D_2-a_i^t}-d_i^t \tag{30}$$

$$\bar{X}_1=\frac{\bar{b}_i^t}{D_1-\bar{a}_i^t}-\bar{d}_i^t,\quad \bar{X}_2=\frac{\bar{b}_i^t}{D_2-\bar{a}_i^t}-\bar{d}_i^t. \tag{31}$$

The average bandwidth for the future slots $\bar{x}_i^t$ in (29) can not be known in advance. However, we could obtain the relationship between $\bar{x}_i^t$ and x_i^t from the equality constraint in (17), that is

$$\bar{x}_i^t=\frac{w_i^t W_{step}-p_t x_i^t}{(T-t)\,\bar{p}_t}. \tag{32}$$

Then by substituting (32) into (29), we obtain

$$\bar{U}_i^t\left(x_i^t\right)=\begin{cases}1, & \left(x_i^t<X_4\right)\\ \frac{\left(D_2-\bar{a}_i^t\right)\left(\frac{w_i^t W_{step}-p_t x_i^t}{(T-t)\bar{p}_t}+\bar{d}_i^t\right)-\bar{b}_i^t}{\left(D_2-D_1\right)\left(\frac{w_i^t W_{step}-p_t x_i^t}{(T-t)\bar{p}_t}+\bar{d}_i^t\right)}, & \left(X_4\le x_i^t\le X_3\right)\\ 0, & \left(x_i^t>X_3\right)\end{cases} \tag{33}$$

where X_3 and X_4 are calculated as follows:

$$X_3=\frac{w_i^t W_{step}}{p_t}-(T-t)\frac{\bar{p}_t}{p_t}\left(\frac{\bar{b}_i^t}{D_2-\bar{a}_i^t}-\bar{d}_i^t\right) \tag{34}$$

$$X_4=\frac{w_i^t W_{step}}{p_t}-(T-t)\frac{\bar{p}_t}{p_t}\left(\frac{\bar{b}_i^t}{D_1-\bar{a}_i^t}-\bar{d}_i^t\right). \tag{35}$$

Next, with the preparations above, we aim to find the closed-form solution of (17). For simplicity, we denote the objective function in (17) as follows:

$$U_{total}\left(x_i^t\right)=U_i^t\left(x_i^t\right)+(T-t)\,\bar{U}_i^t\left(x_i^t\right). \tag{36}$$

Note that both $U_i^t(x_i^t)$ and $\bar{U}_i^t(x_i^t)$ are now functions of x_i^t, as shown in (28) and (33).

It is easy to prove that $U_i^t(x_i^t)$ is monotonic increasing within the region $X_2 \le x_i^t \le X_1$, and $\bar{U}_i^t(x_i^t)$ is monotonic decreasing within the region $X_4 \le x_i^t \le X_3$. Then the optimization of $U_{total}(x_i^t)$ can be considered in 2 cases: (a) $\max\{X_2, X_4\} < \min\{X_1, X_3\}$, (b) $\max\{X_2, X_4\} \ge \min\{X_1, X_3\}$.

First we consider case (a). The x-axis is divided into 5 regions: $[0, \min\{X_2, X_4\})$, $[\min\{X_2, X_4\}, \max\{X_2, X_4\})$, $[\max\{X_2, X_4\}, \min\{X_1, X_3\})$, $[\min\{X_1, X_3\}, \max\{X_1, X_3\})$, and $[\max\{X_1, X_3\}, +\infty)$, denoted *region* $1, 2, \ldots$, and 5 respectively. For *region*1, the values of $U_i^t(x_i^t)$ and $\bar{U}_i^t(x_i^t)$ are constant. So the optimal point in this region lies at $x_i^t = 0$. *Region*5 is similar to *region*1. So the optimal point lies at $x_i^t = \max\{X_1, X_3\}$. For *region*2, there are two possible scenarios. If $X_2 < X_4$, $U_{total}(x_i^t)$ is monotonic increasing, and $\bar{U}_{total}(x_i^t)$ is constant. If $X_2 \ge X_4$, $U_{total}(x_i^t)$ is constant, and $\bar{U}_{total}(x_i^t)$ is monotonic decreasing. So for *region*2, the maximal utility point lies at $x_i^t = \min\{X_2, X_4\}$ or $x_i^t = \max\{X_2, X_4\}$. That is equivalent to $x_i^t = X_2$ or $x_i^t = X_4$. *Region*4 is similar to *region*2. So the optimal point is $x_i^t = X_1$ or $x_i^t = X_3$. For *region*3, $U_{total}(x_i^t)$

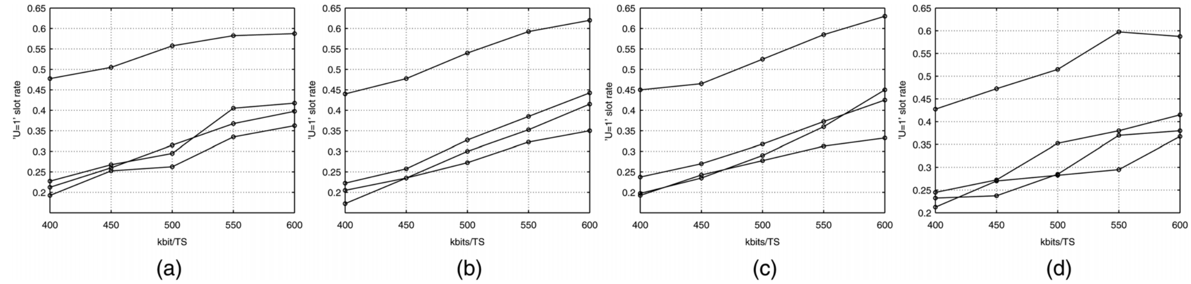

Fig. 12. Slot saturation rate versus average bandwidth for four secondary users. Each curve in the figures corresponds to one user. (a) Case 1. (b) Case 2. (c) Case 3. (d) Case 4.

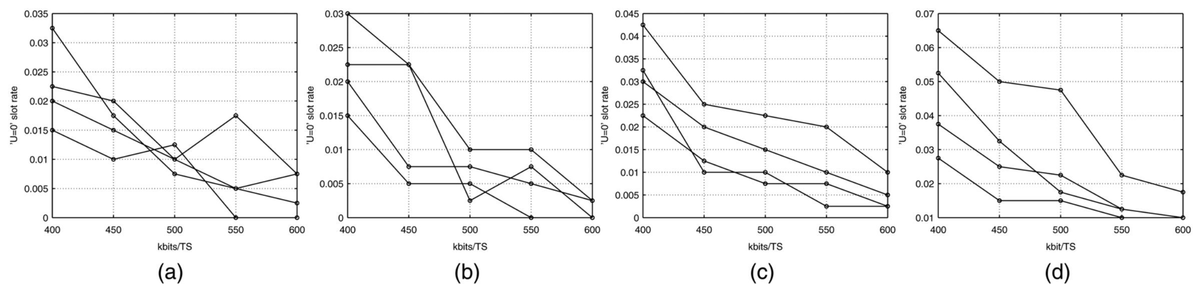

Fig. 13. Slot freeze rate versus average bandwidth for four secondary users. Each curve in the figures corresponds to one user. (a) Case 1. (b) Case 2. (c) Case 3. (d) Case 4.

TABLE III
AVERAGE FRAME FREEZE RATE FOR EACH EFFECT LEVEL

	Bandwidth Model					
	Case 1	Case 2	Case 3	Case 4	Case 5	Case 6
EQL	35.314	35.282	35.258	34.856	35.290	35.315
	Improvement over EQL					
MSE_1BID	0.494	0.554	0.240	0.424	0.474	0.458
MSE_ITER	0.504	0.612	0.544	0.722	0.616	0.489
UTILITY_ITER	**0.822**	**0.842**	**0.848**	**1.070**	**0.836**	**0.806**

is monotonic increasing and $\bar{U}_{total}(x_i^t)$ is monotonic decreasing. In this case, it is known that $U_{total}(x_i^t)$ is always differentiable. By solving $\frac{\partial U_{total}(x_i^t)}{\partial x_i^t} = 0$, we obtain the sole maximal/minimal result as

$$X_5 = \sqrt{\frac{b_i^t}{p_t}} \frac{w_i^t W_{step} \bar{p}_t + (T-t)\,\bar{d}_i^t \bar{p}_t + d_i^t p_t}{\sqrt{b_i^t p_t} + (T-t)\sqrt{\bar{b}_i^t \bar{p}_t}} - d_i^t. \quad (37)$$

If X_5 is within *region*3, it is valid. Otherwise, it is discarded. Considering the boundary cases, the optimal point for *region*3 is $x_i^t = \max\{X_2, X_4\}$, $x_i^t = \min\{X_1, X_3\}$, or $X_i^t = X_5$.

Summing up, the solution for problem (17) in case (a) is as follows:

$$x_i^t\left(w_i^t\right) = \arg\max_{x_i^t \in X} U_{total}\left(x_i^t\right). \quad (38)$$

Here X is defined as $X = \{0, X_1, X_2, X_3, X_4, X_5\}$, where $X_1, \ldots$, and X_5 are shown in (30), (34), (35), and (37).

Then we consider case (b) for the optimization problem. In this case, *region*3 is empty. So X_5 is directly discarded. For the other regions, they are the same as that of case (a). So the solution for problem (17) in case (b) is (38), where X is defined as $X = \{0, X_1, X_2, X_4, X_5\}$.

REFERENCES

[1] G. Su and M. Wu, "Efficient bandwidth resource allocation for low delay multiuser video streaming," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 15, no. 9, pp. 1124–1137, Sep. 2005.

[2] A. Fattahi, F. Fu, M. Schaar, and F. Paganini, "Mechanism-based resource allocation for multimedia transmission over spectrum agile wireless networks," *IEEE J. Sel. Areas Commun.*, vol. 25, no. 4, pp. 601–612, Apr. 2007.

[3] M. Tagliasacchi, G. Valenzise, and S. Tubaro, "Minimum variance optimal rate allocation for multiplexed H.264/AVC bitstreams," *IEEE Trans. Image Process.*, vol. 17, no. 7, pp. 1129–1143, Jul. 2008.

[4] L. Wang and A. Vincent, "Joint rate control for multi-program video coding," *IEEE Trans. Consum. Electron.*, vol. 42, no. 3, pp. 300–305, Mar. 1996.

[5] X. Zhu and B. Girod, "Distributed rate allocation for multi-stream video transmission over ad hoc networks," in *Proc. IEEE Int. Conf. Image Process.*, Genoa, Italy, Sep. 2005, pp. 157–160.

[6] M. Tiwari, T. Groves, and P. Cosman, "Bit-rate allocation for multiple video streams using a pricing-based Mechanism," *IEEE Trans. Image Process.*, vol. 20, no. 11, pp. 3219–3230, Nov. 2011.

[7] J. Ronda, M. Eckert, F. Jaureguizar, and N. Garcia, "Rate control and bit allocation for MPEG-4," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 9, no. 12, pp. 1243–1258, Dec. 1999.

[8] F. Fu, T. Stoenescuand, and M. Schaar, "A pricing mechanism for resource allocation in wireless multimedia application," *IEEE J. Sel. Topics Signal Process.*, vol. 1, no. 2, pp. 264–279, Aug. 2007.

[9] A. Mohsenian-Rad, V. Wong, and V. Leung, "Two-fold pricing to guarantee individual profits and maximum social welfare in multi-hop wireless access works," *IEEE Trans. Wireless Commun.*, vol. 8, no. 8, pp. 4110–4121, Aug. 2009.

[10] D. Niyato and E. Hossain, “Market-equilibrium, competitive, and cooperative pricing for spectrum sharing in cognitive radio networks: Analysis and comparison,” *IEEE Trans. Wireless Commun.*, vol. 7, no. 11, pp. 4273–4283, Nov. 2008.
[11] D. Niyato and E. Hossain, “A game theoretic analysis of service competition and pricing in heterogeneous wireless access networks,” *IEEE Trans. Wireless Commun.*, vol. 7, no. 12, pp. 5150–5155, Dec. 2008.
[12] D. Niyato and E. Hossain, “Competitive pricing for spectrum sharing in cognitive radio networks: Dynamic game, inefficiency of Nash equilibrium, and collusion,” *IEEE J. Sel. Areas Commun.*, vol. 26, no. 1, pp. 192–202, Jan. 2008.
[13] M. Tiwari, T. Groves, and P. Cosman, “Competitive equilibrium bitrate allocation for multiple video streams,” *IEEE Trans. Image Process.*, vol. 19, no. 4, pp. 1009–1021, Apr. 2010.
[14] Z. Chen and K. Ngan, “Distortion variation minimization in real-time video coding,” *Signal Process.: Image Commun.*, vol. 21, pp. 273–279, Jan. 2006.
[15] Y. Wang, L. Chau, and K. Yap, “Joint rate allocation for multiprogram video coding using FGS,” *IEEE Trans. Circuits Syst. Video Technol.*, vol. 20, no. 6, pp. 829–837, Jun. 2010.
[16] W. Rhee and J. Cioffi, “Increase in capacity of multiuser OFDM system using dynamic subchannel allocation,” in *Proc. IEEE Veh. Spring Conf.*, Tokyo, Japan, May 2000, vol. 2, pp. 1085–1089.
[17] A. Vetro, H. Sun, and Y. Wang, “MPEG-4 rate control for multiple video objects,” *IEEE Trans. Circuits Syst. Video Technol.*, vol. 9, no. 2, pp. 186–199, Feb. 1999.
[18] L. Wang and A. Vincent, “Bit allocation and constraints for joint coding of multiple video programs,” *IEEE Trans. Circuits Syst. Video Technol.*, vol. 9, no. 6, pp. 949–960, Sep. 1999.
[19] M. Balakrishan and R. Cohen, “Global optimization of multiplexed video encoders,” in *Proc. Int. Conf. Image Process.*, Santa Barbara, CA, USA, Oct. 1997, pp. 377–380.
[20] E. Malinvaud and M. Bacharach, “Decentralized procedures for planning,” in *Activity Analysis in the Theory of Growth and Planning*. New York, NY, USA: St. Martin’s, 1967, pp. 170–208.
[21] A. Mas-Colell, M. Whinston, and J. Green*, Microeconomics Theory*. Oxford, U.K.: Oxford Univ. Press, 1995.
[22] L. Tang, Y. Chen, E. Hines, and M. Alouini, “Effect of primary user traffic on sensing-throughput tradeoff for cognitive radios,” *IEEE Trans. Wireless Commun.*, vol. 10, no. 4, pp. 1063–1068, Apr. 2011.
[23] K. Stuhlmuller, N. Farber, M. Link, and B. Girod, “Analysis of video transmission over lossy channels,” *IEEE J. Sel. Areas Commun.*, vol. 18, no. 6, pp. 1012–032, Jun. 2000.
[24] H. Park and M. van der Schaar, “Bargaining strategies for networked multimedia resource management,” *IEEE Trans. Signal Process.*, vol. 55, no. 7, pp. 3496–3511, Jul. 2007.
[25] S. Martello and P. Toth*, Knapsack Problems: Algorithms and Computer Implementations*. New York, NY, USA: Wiley, 1990.

Meng Yang (S’12) received the B.S. degree in electrical engineering and the Ph.D. degree in automation from Xi’an Jiaotong University, Xi’an, China, in 2008 and 2014, respectively.

He was a Visiting Scholar with the Department of Electrical and Computer Engineering, University of California at San Diego, La Jolla, CA, USA, from October 2011 to November 2012. He is currently an Assistant Professor with the School of Electronic and Information Engineering, Xi’an Jiaotong University, Xi’an, China. His research interests focus on 2-D/3-D video processing and optimization.

Theodore Groves received the B.A. degree from Harvard University, Cambridge, MA, USA, and the Ph.D. degree in economics from the University of California, Berkeley, CA, USA, in 1970.

He was previously a faculty member with the University of Wisconsin, Madison, WI, USA, Northwestern University’s Kellogg School of Management, Evanston, IL, USA, and Stanford University, Stanford, CA, USA. In 1979, he joined the University of California at San Diego (UCSD), La Jolla, CA, USA, as a Professor of Economics. He is the Director of the Center for Environmental Economics in the Department of Economics, UCSD, La Jolla, CA, USA, and is involved in ongoing research on international and national fisheries, water pricing, and numerous projects for managing marine resources and the protection of endangered species. He was a founder of mechanism design theory, and the discoverer of the “Groves Mechanism” for eliciting truthful information in an incentive-compatible manner. He and coauthor John Ledyard also developed the first general equilibrium solution to the “free rider problem” of public goods. He has also studied the Chinese economy’s transition to a market economy, optimal policies for minimizing the occurrence of oil spills, the incentive compatibility of stated-preference surveys, and improved methods for video multiplexing.

Prof. Groves is a Fellow of the Econometric Society and the American Academy of Arts and Sciences.

Nanning Zheng (SM’94–F’06) received the B.S. degree and the M.S. degree in information and control engineering from Xi’an Jiaotong University (XJTU), Xi’an, China, in 1975 and 1981, respectively, and the Ph.D. degree in electrical engineering from Keio University, Yokohama, Japan, in 1985.

In 1975, he joined XJTU, Xi’an, China, where he is currently a Professor and the Director of the Institute of Artificial Intelligence and Robotics. His research interests include computer vision, pattern recognition and image processing, and hardware implementation of intelligent systems.

Dr. Zheng became a member of the Chinese Academy of Engineering in 1999 and is the Chinese Representative on the Governing Board of the International Association for Pattern Recognition. He also serves as Executive Deputy Editor of the Chinese Science Bulletin.

Pamela Cosman (S’90–M’93–SM’00–F’08) received the B.S. degree with honors in electrical engineering from the California Institute of Technology, Pasadena, CA, USA, in 1987, and the M.S. and Ph.D. degrees in electrical engineering from Stanford University, Stanford, CA, USA, in 1989 and 1993, respectively.

She was an NSF postdoctoral fellow at Stanford University, Stanford, CA, USA, and a Visiting Professor at the University of Minnesota from 1993 through 1995. In 1995, she joined the Department of Electrical and Computer Engineering, University of California, San Diego, La Jolla, CA, USA. From 2006 to 2008, she was the Director of the Center for Wireless Communications at the University of California, San Diego. She is currently a Professor and Vice Chair at the University of California, San Diego, La Jolla, CA, USA. Her research interests are in the areas of image and video compression and processing and wireless communications.

Dr. Cosman is a member of Tau Beta Pi and Sigma Xi. She has been a member of the Technical Program Committee or the Organizing Committee for numerous conferences, including ICIP 2008–2011, QOMEX 2010–2012, ICME 2011–2013, VCIP 2010, PacketVideo 2007–2013, WPMC 2006, ICISP 2003, ACIVS 2002–2012, ICC 2012, Asilomar Conference on Signals, Systems, and Computers 2003, and EUSIPCO 1998. She was the Technical Program Chair of the 1998 Information Theory Workshop in San Diego, CA, USA. Her awards include the ECE Departmental Graduate Teaching Award, a Career Award from the National Science Foundation, a Powell Faculty Fellowship, Globecom 2008 Best Paper Award, and HISB 2012 Best Poster Award. She was a Guest Editor of the June 2000 special issue on error-resilient image and video coding of the IEEE JOURNAL ON SELECTED AREAS IN COMMUNICATIONS She was an Associate Editor of the IEEE COMMUNICATIONS LETTERS (1998–2001), an associate editor of the IEEE SIGNAL PROCESSING LETTERS (2001–2005), and the Editor-in-Chief (2006–2009) as well as a Senior Editor (2003–2005, 2010–present) of the IEEE JOURNAL ON SELECTED AREAS IN COMMUNICATIONS.