QoE-Aware Resource Allocation for Small Cells

Anis Elgabli, Ali Elghariani*, Vaneet Aggarwal, Mark Bell
Purdue University, West Lafayette IN 47907
*University of Tripoli

Abstract—In this paper, we study the problem of Quality of Experience (QoE) aware resource allocation in wireless systems. In particular, we consider application-aware joint Bandwidth-Power allocation for a small cell. We optimize a QoE metric for multi-user video streaming in a small cell that maintains a trade-off between maximizing the playback rate of each user and ensuring proportional fairness (PF) among users. We formulate the application-driven joint bandwidth-power allocation as a non-convex optimization problem. However, we develop a polynomial complexity algorithm, and we show that the proposed algorithm achieves the optimal solution of the proposed optimization problem. Simulation results show that the proposed QoE-aware algorithm significantly improves the average QoE. Moreover, it outperforms the weighted sum rate allocation which is the state-of-the-art physical resource allocation scheme.

Index Terms—QoE, Video streaming, Bandwidth-Power allocation, non-convex optimization

I. INTRODUCTION

Video streaming is the dominant contributor to the cellular traffic. Currently, video streaming accounts for 50% of cellular traffic and it is expected to to grow to 75% of the mobile data traffic by the year of 2020 [1]. This increase has forced service providers to enhance their infrastructures to support high-quality video streaming. Despite these efforts, users frequently experience low Quality of Experience (QoE) metrics such as choppy videos and playback stalls [2].

In the modern video coding, the video is divided into chunks with $L$ second duration each. Each chunk is either encoded into $N$ independent versions with different qualities, or encoded into ordered layers, i.e., one base layer (Layer 0) with the lowest playable quality, and multiple enhancement layers (Layer $i > 0$) that further improve the chunk quality. As an example of the former technique is H.264/MPEG-4 AVC (Advanced Video Coding) which was standardized in 2003 [3]. Moreover, an example that represents the layered encoding technique is Scalable Video Coding (SVC) which was standardized in 2007 as an extension to H.264 [4]. In both encoding techniques, the video is available at different rates, and a rate adaptation logic needs to make a decision about the quality of the next few chunks to request based on the current buffer and/or bandwidth prediction.

The recent adoption of the open standard MPEG-DASH [5] has made Adaptive Bit Rate (ABR) streaming the most popular video streaming solution. Commercial systems such as Apple’s HLS [6], Microsoft’s Smooth Streaming [7], and Adobe’s HDS [8] are all ABR streaming algorithms. In recent studies, researchers have investigated various approaches for making streaming decisions, for example, by using control theory [9], [10], Markov Decision Process [11], machine learning [12], client buffer information [13], and data-driven techniques [14], [15], [16]. However, all of these rate adaptation techniques resides at the application layer of the client. Hence, the client decides based on the bandwidth prediction and/or the current buffer size what rate should next chunk of the video be fetched at. Therefore, the wireless base station provides no help in ensuring a certain rate to the client. In this paper, we still assume that each user is running his/her own rate adaptation logic at the application layer to adapt to the network changes. The QoE aware resource allocation algorithm proposed in this paper does not replace the application layer rate adaptation technique that runs at the client side. However, it allows the base station (BS) to consider the application requirements in assigning the resources. Therefore, the physical layer resources (bandwidth and power) can be allocated such that the average QoE among users is maximized.

In this paper, driven by the following aspects; (i) most of the cellular traffic is video, (ii) the video is available at a specific set of rates which is a discrete in nature, and (iii) the new paradigm of a small cell in which a small BS with low power and more customizability resides in a small place can serve few users, we propose a video QoE-Aware bandwidth-power allocation algorithm for a small cell scenario.

To motivate our problem, consider a scenario in which $U$ users are connected to a small cell and watching videos. The videos are encoded into $N+1$ rates: $r_0, r_1, ..., r_N$. If a user cannot achieve a rate of $r_0$, then the user will run into rebuffering which significantly degrades the user’s QoE. The user’s QoE is generally evaluated by the mean opinion score (MOS) [17], [18], which practically ranges, e.g., from 1 to 4.5. Moreover, video MOS is shown to be bound logarithmic with respect to the Quality of Service (QoS) parameter [19], [20]. The QoS parameter in video streaming is the video playback rate. Therefore, the increase in the QoE is more significant when switching from rate $r_{n-1}$ to $r_n$ compared to switching from $r_n$ to $r_{n+1}$. Hence, in the case of multiple users connected to a small cell scenario, pushing a large number of users to the rate $r_n$ is more preferable than just increasing the rates of only a few users to rates higher than $r_n$ at the cost of dropping the rates of the majority of users to rates below $r_n$. This strategy improves the playback rate of each user and ensures proportional fairness among users.
In the physical layer, utility-based resource allocation in wireless networks that ensures fairness among users with different application requirements was studied in [21]. However, the authors in [21] assumed that the traffic of the users with hard QoS are available at a single rate which is not true for the modern videos. Moreover, all users are sharing a single total rate which does not hold when the users are experiencing different channels. The authors of [22] proposed a QoE-based proportional fairness utility function in which a continuously differentiable MOS function is maximized based on relaxing the discrete constraints to simplify the problem. In contrast to the previous work, we exploit the discrete nature of the available video rates and show that the non-convex optimization problem is solvable in polynomial time.

In this paper, we first formulate the joint bandwidth-power allocation for a video traffic in a multi-user small cell scenario as a non-convex optimization problem. The objective is to optimize a QoE metric that maintains a trade-off between maximizing the playback rate of each user and ensuring proportional fairness among all users. Secondly, we develop a novel algorithm that solves this problem optimally in a polynomial time complexity. Finally, we compare our proposed algorithm with the weighted sum rate, which is the state-of-the-art physical resource allocation technique. Moreover, we show that our algorithm outperforms the weighted sum rate in terms of avoiding rebuffering times and maintaining high QoE for more users.

II. System Model

Consider a single cell wireless network as shown in Fig. 1 with U users, \( u \in U = \{ 1, 2, 3, \ldots, U \} \). A spectrum of a total bandwidth, \( B \), is available for the downlink transmissions from the Base Station (BS) or Access Point (AP) to all users. Each user is allocated a downlink portion of the spectrum for online video streaming. This spectrum is assumed to be flat fading and can be divided into distinct and non-overlapping channels of unequal bandwidths, so that the users share the available spectrum through a frequency division manner. Let \( p_u \) and \( b_u \) represent the allocated transmit power and channel bandwidth of the BS to serve the \( u \)-th user. The received SNR at the \( u \)-th user is \( \text{SNR}_u = \frac{p_u H_u}{N_o b_u} \), and according to the Shannon theorem, the channel capacity, \( C_u \), of the \( u \)-th user is:

\[
C_u = b_u \log(1 + \frac{p_u H_u}{N_o b_u})
\]

where \( H_u \) denotes the channel power of the link between the BS and the \( u \)-th user, which can be modeled based on both large-scale and small-scale fading effects as follows:

\[
H_u = |\alpha_u h_u|^2
\]

where \( h_u \) represents the small scale fading which is modeled as \( \mathcal{CN}(0, 1) \). \( \alpha_u \) represents the large scale propagation effect for the \( u \)-th user with \( \alpha_u = \sqrt{c/d_u^m} \). \( d_u \) is the relative distance between the \( u \)-th user and the base station, \( m \) is the path loss exponent which is typically between 1.6 and 4 depending on the environment, and \( c \) is a constant related to the propagation loss and antenna gains. Large scale fading that results from shadowing can be modeled as a log normal distribution for the \( u \)-th user channel, but since in this paper we focus on a small cell such as the one located inside a building, the shadowing effect can be ignored. In this model, we assume that each user watches a video which is divided into \( C \) chunks with \( L \)-second duration each. This video is encoded at one of the \( N+1 \) playback rates, \( r_0, \ldots, r_N \). Higher playback rate leads to a better QoE. However, higher playback rate requires more physical resources (Bandwidth and Power).

We denote \( B \triangleq \{ b_0, b_1, \ldots, b_U \} \) as the bandwidth allocation, \( \mathcal{P} \triangleq \{ p_0, p_1, \ldots, p_U \} \) as the Power allocation.

III. Problem Formulation

In this section, we describe our problem formulation. Let \( I_n^u \) be the decision variable of the \( n \)-th rate and \( u \)-th user, i.e,

\[
\begin{align*}
I_n^u &= 1, \text{ if the } u \text{-th user is a candidate to the } n \text{-th rate} \\
I_n^u &= 0, \text{ otherwise}
\end{align*}
\]

\( n \in \{0, \ldots, N\} \)

Our objective function is a weighted sum of the decision variables. Let \( \lambda_n \) be the weight of the \( n \)-th rate. In order to prefer pushing all the users to the \( n \)-th rate over any other choice that is pushing some users to the higher rates with the cost of dropping the rate of others to below the \( n \)-th rate, the weights should satisfy the following constraint:

\[
\lambda_n > U \cdot \sum_{k=n+1}^{N} \frac{\lambda_k}{N}
\]

Equation (3) states that the utility that is achieved by pushing a user to the \( n \)-th rate if possible is higher than the utility that is achieved by pushing all other users to rates > \( n \)-th rate. Therefore, maximizing this objective will maximize the playback rate of each user subjected to ensuring proportional fairness among users. The overall optimization problem is formulated as follows:

\[
\text{Maximize } \sum_{n=0}^{N} \sum_{u=1}^{U} I_n^u
\]
subject to

\[ I_0^u = 1 \left( b_u \log(1 + \frac{\gamma p_u |a_u h_u|^2}{N_o b_u}) \geq r_0 \right), \forall u \]  

\[ I_n^u = I_{n-1}^{u', 1} \left( b_u \log(1 + \frac{\gamma p_u |a_u h_u|^2}{N_o b_u}) \geq r_n \right), \forall u, n = 1...N \]  

\[ \sum_u p_u \leq P_{\text{max}} \]  

\[ \sum b_u \leq B \]  

\[ I_n^u \in \{0, 1\}, n = 1...N, u \]  

Where 1\((x > y) = \{1, \text{if } x > y, \text{and } 0 \text{ otherwise}\), \(\gamma \in (0, 1)\) is the rate to the capacity gap, and \(\lambda_0, ..., N\) must satisfy \((3)\). Constraint \((5)\) states that \(I_0^u\) is equal to 1 only if there is a bandwidth-power allocation such that the user \(u\) achieves a rate \(\geq r_0\). Constraint \((6)\) states that \(I_n^u, \forall u, n > 0\) is equal to 1 only if the user \(u\) is a candidate to the immediate lower rate \((I_{n-1}^{u'})\) and there is a bandwidth-power allocation such that the user \(u\) achieves a rate \(\geq r_n\). \((7)\) and \((8)\) are the total power and bandwidth constraints. Finally, \((9)\) is the non-convexity constraint on the feasible set of \(I_n^u, \forall u, n\). Therefore, the optimization problem \((4)-(9)\) is a non-convex. In the next section, we will describe a novel algorithm that can solve this problem optimally in polynomial complexity.

### IV. Proposed Algorithm

In this section, we explain the proposed algorithm which is listed in Table 1. We process rate levels in order such that when making the \(n\)-th rate level decisions, we only consider users who were candidates to the \((n - 1)\)-th rate level. i.e., we consider a user \(u\) if \(I_{n-1}^u = 1\), otherwise, the constraint \((6)\) will be violated. Let’s index the users according to their channel gains where the user with the highest channel gain is given the index 1, and the user with the lowest channel gain is given the index \(U\).

We start with rate \(r_0\) by considering only user 1. Then we check if the user 1 is allocated the whole bandwidth and power \((b_1 = B, \text{ and } p_1 = P_{\text{max}})\), would he/she be able to achieve the 0-th rate. i.e., If \(b_1 \log(1 + \frac{\gamma p_1 |a_1 h_1|^2}{N_o b_1}) \geq r_0\), then the user is a candidate to the 0-th rate, and thus \(I_0^{(1)} = 1\). Otherwise, no user will be a candidate to this rate and the algorithm should terminate. In general if the \(u\)-th user is not a candidate to the \(n\)-th rate, every user \(\in \{u + 1, \ldots, U\}\) will not be a candidate to rate levels \(\geq n\)-th rate. Hence the algorithm proceeds by checking the possibility of finding candidate users to the next rate among users 1 to \(u - 1\). If user 1 is a candidate to the 0-th rate (i.e \(I_0^{(1)} = 1\)), we solve the following optimization problem to find if user 2 can also be a candidate to the rate \(r_0\) given that user 1 is already a candidate to this rate level. Note, the candidacy of user 1 to rate \(r_0\) is formally stated by constraint \((11)\):

\[ \text{argmax}_{b_1, b_2, p_1, p_2} b_2 \log(1 + \frac{\gamma p_2 |a_2 h_2|^2}{N_o b_2}) \]  

subject to

\[ b_1 \log(1 + \frac{\gamma p_1 |a_1 h_1|^2}{N_o b_1}) \geq r_0 \]  

\[ p_1 + p_2 \leq P_{\text{max}} \]  

\[ b_1 + b_2 \leq B \]  

Since we already know that user 1 is a candidate to the 0-th rate, there must be a bandwidth-power allocation such that the constraint \((11)\) is satisfied. By solving the optimization problem \((10)-(13)\), we find whether user 2 is a candidate to this rate or not.

**Remark** The bandwidth-power allocation for both user 1 and user 2 obtained by solving \((10)-(13)\) overweights the previous decision when considering only user 1. In general, solving the allocation problem at the \(u\)-th user and \(n\)-th rate overweights the previous allocation of all users.

Now, for the \(u\)-th user and \(0\)-th rate, we need to solve the following optimization problem:

\[ \text{argmax}_{b_u, p} b_u \log(1 + \frac{\gamma p_u |a_u h_u|^2}{N_o b_u}) \]  

subject to

\[ b_k \log(1 + \frac{\gamma p_k |a_k h_k|^2}{N_o b_k}) \geq r_0, \forall k < u \]  

\[ \sum_{k=1}^{u-1} p_k + p_u \leq P_{\text{max}} \]  

\[ \sum_{k=1}^{u-1} b_k + b_u \leq B \]  

If \(b_u \log(1 + \frac{\gamma p_u |a_u h_u|^2}{N_o b_u}) \geq r_0\), then the \(u\)-th user is a candidate to this rate, otherwise, all users \(\{u, \ldots, U\}\) will not be considered for all rates.

In order to generalize the case to any rate \(> r_0\), let’s first define \(r_m\) to be the maximum rate \(< r_n\) for which the user \(k\) is a candidate. Then to find out if the \(u\)-th user is a candidate to the \(n\)-th rate, we solve the following optimization problem:

\[ \text{argmax}_{b_u, p} b_u \log(1 + \frac{\gamma p_u |a_u h_u|^2}{N_o b_u}) \]  

subject to

\[ b_k \log(1 + \frac{\gamma p_k |a_k h_k|^2}{N_o b_k}) \geq r_{m}, \forall k = 1...U, \neq u \]  

\[ \sum_{k=1}^{u} p_k + p_u \leq P_{\text{max}} \]
\[
    \sum_{k=1, k \neq u}^{U} b_k + b_u \leq B
\]  
(21)

If \( b_u \log(1 + \gamma p_u |h_u|^2) \geq r_n \), then the \( u \)-th user is a candidate to the \( n \)-th rate, otherwise, it is not a candidate to any rate \( \geq r_n \). This means that we exclude all users \((u, \ldots, U)\) for all rates \( \geq r_n \). Finally, solving the optimization problem (18)-(21) for the highest rate \( N \) \((r_N)\) and for the last user that is a candidate to the rate \( N - 1 \) yield the optimal solution to the original problem described in (19). Next, we proof the optimality of the proposed algorithm in solving the optimization problem described in (4)-(9).

Lemma 1: Given \((I_0^u, \ldots, I_{n-1}^u), \forall u \in \{1, \ldots, U\}\), the execution of the algorithm for the \( n \)-th rate decisions yields the maximum value of \( \sum_{u=1}^{U} I_n^u \). In other words, the proposed algorithm obtains the maximum number of users at rate \( n \) as compared to any feasible algorithm which has the same rate decisions of every user up to the rate level \( n - 1 \).

Proof: We first note that using the proposed algorithm, the user \( u_s \) is skipped at the \( n \)-th rate level. i.e. the decision variable of the \( n \)-th rate and user \( u_s \) is equal to 0 \((I^u_s = 0)\) in two scenarios, which are described as follows.

Case 1: If the user \( u_s \) is not a candidate to the \( (n-1) \)-th rate i.e., \( I_{n-1}^u = 0 \), no other feasible algorithm can increase the rate of this user, \( u_s \), to the \( n \)-th rate level due to the violation of the constraint \( (3) \).

Case 2: There is no feasible bandwidth-power allocation such that constraints (19)(21) are satisfied and \( b_u \log(1 + \frac{\gamma p_u |h_u|^2}{N_0 b_n}) \geq r_n \)

Skips which are due to Case 1 are the same for any feasible algorithm. Thereby, we do not consider skips that are of Case 1. On the other hand, for the skips of Case 2, we note that the proposed algorithm starts with the user that has the highest channel gain since this user can achieve the \( n \)-th rate with the lowest resources if that is feasible. Therefore, if there is no bandwidth-power allocation such that this user can achieve the \( n \)-th rate, there will be no other user can achieve this rate. The algorithm proceeds in descending order with respect to the channel gains. Hence, if user \( u_s \) is skipped at the \( n \)-th rate level, any algorithm would have skipped a user or more with an index \( u \leq u_s \). However, skipping user \( u_s \) as compared to user \( u \leq u_s \) allows for more bandwidth-power to increase the quality of users with higher channel gains. i.e. since user \( u_s \) has lower channel gain, he/she needs more resources than user \( u \leq u_s \) to reach rate \( r_n \). Therefore, skipping the user \( u_s \) offers more resources to the remaining users. Thus, we see that the number of \( n \)-th rate skips in any feasible algorithm can be no less than the proposed algorithm.

Theorem 1: Up to a given rate \( N \geq 0 \), if \( I_n^u \) is the decision variable of the \( n \)-th rate and \( u \)-th user \((n \leq N)\) that is found by the proposed algorithm, and \( I_n^{u'} \) is the decision variable of the \( n \)-th rate and \( u \)-th user that is found by any other feasible algorithm such that all constraints (4)-(9) are satisfied, then the following holds when \( \lambda \)’s satisfy (9).

\[
    \sum_{n=0}^{N} \sum_{u=1}^{U} I_n^u \leq \sum_{n=0}^{N} \lambda_n \sum_{u=1}^{U} I_n^{u'}
\]
(22)

In other words, the proposed algorithm achieves the optimal solution of the optimization problem (4)-(9) when \( \lambda \)’s satisfy (3).

Proof: First, the constraint (3) on \( \lambda_0, \ldots, N \) imposes a strict priority on the rate levels. Therefore, pushing 1 user to the rate \( r_n \) yields higher utility than pushing all the other users to rates \( > r_n \). Hence, processing rates in their order and sequentially finding the maximum users served at rates 0 to \( N \) in their order will yield the optimal solution to the proposed problem. Lemma 1 proves that the proposed algorithm achieves the maximum number of users at the rate \( n \leq N \) given the decisions of all rates up to the rate \( n - 1 \). Projecting that on rate \( r_0 \), we conclude that the algorithm finds the maximum number of users that can be served at rate \( r_0 \). Given the \( r_0 \) decision which is optimal, running the algorithm for \( r_1 \) will produce the optimal rate decisions for both \( r_0 \) and \( r_1 \). Hence, continuing this way up to the rate \( r_N \) will yield the optimal solution of all rates and users up to the rate \( N \) and that concludes the proof.

V. NUMERICAL RESULTS

In this section, we provide simulation results for the studied scenario of a single small cell multiuser with video streaming in the downlink. We consider the case of 5 users downloading a video of length 1 minute. The video is divided into 60 chunks, each chunk is of 1 second duration. The video is encoded into 5 rate levels \([r_1, r_2, r_3, r_4, r_5]\) which are \([0.5, 1, 1.5, 2, 2.5]\) Mbps, respectively. A user streams with rate 1 \((r_1)\) means that it can span a range from 0.5 Mbps to less than 1 Mbps, and it streams with rate 2 \((r_2)\), means that it can span a range between 1 Mbps to less than 1.5

| TABLE I: Proposed Algorithm |
|-------------------------------|
| **Input:** \( r_0, \ldots, N \): set of the rates, \( \{1, \ldots, U\} \): set of the users, \( B, P_{max}, H_k, \forall k, N_0, \gamma \) |
| **Output:** \( p_u \text{ and } b_u \forall u = 1, \ldots, U \) |
| 1 Initialize: \( r_0^u = 0, I_0^u = 0 \) |
| 2 for \( n = 0 : N \) |
| 3 \( U = \{1, \ldots, U\} \) |
| 4 While \((U \text{ is not empty})\) |
| 5 \( u \leftarrow \text{index of the user with maximum channel gain } \in U \) |
| 6 Solve the optimization problem defined by Eqs (18)-(21) |
| 7 if \( (b_u \log(1 + \frac{\gamma p_u |h_u|^2}{N_0 b_n}) \geq r_n) \) then |
| 8 \( I_n^u = 1, r_n^u = r_n \) |
| 9 \( U = U \setminus \{u\} \) |
| 10 else |
| 11 \( I_n^k = 0, \forall k \in U \) |
| 12 \( U = \phi \) |
| 13 endif |
| 14 endWhile |
| 15 end |


Mbps, and so on for the other rate levels. We add rate level 0 ($r_0$) to represent the case when the user cannot achieve the basic rate ($r_1$). The path loss and the Rayleigh fading effects are both considered in the downlink. The path loss gain is computed based on the path loss exponent of 2. The maximum available bandwidth at the BS is 20 MHz and the maximum power is 10 watt. The noise PSD, $N_o$, is taken as $10^{-9}$ for SNR calculation. For the sake of performance comparison with other algorithms, we compare our algorithm with the weighted sum rate maximization when window size, $T$, is 1 and 5.

The environment of a small cell is considered such that a low mobility scenario can be studied. The channel is considered slowly fading with coherence time that can span a one complete video chunk. Since the proposed algorithm explicitly consider the fairness in the formulation (no weight based on the history is needed), we solve the optimization once per chunk. However, for weighted sum algorithms, in order to ensure fairness among users per chunk we solve the optimization problem 40 time per chunk. In this experiment, the users’ distance from the BS is randomly generated between 5m and 80m and the whole simulation experiment is repeated 100 times.

Fig. 2 shows the rate percentage across all users and all video rates (0.5, 1, 1.5, 2, 2.5 Mbps) averaged over all repeated simulations. From this figure, it can be said that users with the proposed algorithm spend less percentage of time experiencing rate 0 level (i.e. $< 0.5$ Mbps), which is represented by the dark blue color in Fig. 2, as compared to the weighted sum rate algorithms. This is an advantage of the proposed algorithm from the point of view of the users’ QoE. Moreover, the advantage of the proposed algorithm can also be seen in the rate levels 1, 2, 3, and 4, in which users spend more percentage of time experiencing these rates compared to the other algorithms. The only situation when other algorithms outperform the proposed algorithm is in achieving the rate level 5. However, this advantage comes at the cost of making users experience the rate $r_0$ for more than 30% of the time. The importance of our algorithm is that it reduces the number of time users run below the basic rate, i.e. it minimizes the probability of rebuffing which otherwise significantly degrades the QoE.

To further analyze the above experiment, we consider a one particular case when the users 1 to 5 are located at distances 30, 35, 40, 55, 60 meters from the BS, respectively. Fig. 3 shows the CDF of the rate achieved by all users. This figure demonstrates the superiority of the proposed algorithm in guaranteeing a satisfactory rate level for video streaming most of the times, even for the users with the furthest distance from the BS (e.g. users 4 and 5). On the average, our algorithm can guarantee the basic rate or higher for 90% of the time. While on the other hand, users based on the weighted sum rate algorithm suffers to reach the basic rate for about 30 – 40% of the time, on the average. In particular, if we consider users 1 and 5 for all algorithms, we can see from Fig. 3a that user 1, which is the closest to the BS, can stream with the basic rate or higher for 97%, 89%, and 53% of the time based on the proposed algorithm, weighted sum rate ($T=5$), and weighted sum rate ($T=1$), respectively. Similarly, from Fig 3-e, it can be depicted that user 5, which is the furthest user from the BS, can stream with the basic rate or higher for 78%, 40%, and 40% of the time based on the proposed algorithm, weighted sum rate ($T=5$), and weighted sum rate ($T=1$), respectively. Similar to the observation we made above in Fig (2), the advantage of the weighted sum rate techniques in Fig’s. 3c-d-e in achieving better rate levels (>1.5 Mbps) for users 3, 4, and 5 does not prevent these users from suffering for a noticeable percentage of time.

The objective function in (4) represents the QoE metric in this work. The higher the value of this objective function, the higher the playback rate of each user with guaranteeing proportional fairness among all users. Note that the values of $\lambda_n$ in (4) are chosen to satisfy (3), and thus, the values of $\lambda_0, \ldots, \lambda_4$ are chosen, for this simulation experiment, as $[13310, 1210, 110, 10, 1]$, respectively. This choice of $\lambda_0, \ldots, \lambda_4$ tells that pushing more users to the basic rate level achieves significantly higher objective than pushing one user to the highest rate. Fig. 4 plots the CDF of the objective function for all the considered algorithms. It can be clearly observed that the proposed algorithm significantly outperforms the other algorithms.

VI. CONCLUSION

In this paper, we formulate the joint bandwidth-power allocation for a small cell and video traffic as a non-convex optimization problem whose objective is to optimize the QoE metric that maintains a trade-off between maximizing the playback rate of each user and ensuring users’ proportional fairness (PF). We developed a novel algorithm that solves this
problem optimally in a polynomial time complexity. Simulation results show that the proposed algorithm outperforms the weighted sum rate in terms of maintaining a high QoE for all users.

REFERENCES
[1] “Ericsson Mobility Report on the pulse of the networked society,” Jun. 2016.
[2] X. K. Zou, J. Erman, V. Gopalakrishnan, E. Halepovic, R. Jana, X. Jin, J. Rexford, and R. K. Sinha, “Can accurate predictions improve video streaming in cellular networks?” in Proceedings of the 16th International Workshop on Mobile Computing Systems and Applications, ACM, 2015, pp. 57–62.
[3] “H.264/MPEG-4 AVC,” http://handle.itu.int/11.1002/1000/6312
[4] H. Schwarz, D. Marpe, and T. Wiegand, “Overview of the scalable video coding extension of the h.264/avc standard,” IEEE Transactions on Circuits and Systems for Video Technology, vol. 17, no. 9, pp. 1103–1120, Sept 2007.
[5] “MPEG-DASH,” http://goo.gl/Qxtp9Y
[6] “Apple HTTP Live Streaming,” https://goo.gl/6yYWg
[7] “Microsoft Smooth Streaming,” http://goo.gl/4QHLW
[8] “Adobe HTTP Dynamic Streaming,” http://goo.gl/I2WE8d
[9] X. Yin, A. Jindal, V. Sekar, and B. Sinopoli, “A Control-Theoretic Approach for Dynamic Adaptive Video Streaming over HTTP,” in Proc. ACM SIGCOMM, 2013.
[10] K. Miller, D. Bhanahabotla, G. Caire, and A. Wolisz, “A Control-Theoretic Approach to Adaptive Video Streaming in Dense Wireless Networks,” IEEE Trans. Multimedia, vol. 17, no. 8, 2015.
[11] D. Jarnikov and T. Özçelebi, “Client intelligence for adaptive streaming solutions,” Sig. Proc.: Image Comm., vol. 26, no. 7, 2011.
[12] M. Claeyts, S. Latré, J. Famaey, and F. D. Turck, “Overview of the scalable video coding extension of the h.264/avc standard,” IEEE Transactions on Circuits and Systems for Video Technology, vol. 18, no. 4, 2014.
[13] T.-Y. Huang, R. Jothi, N. McKeown, M. Trunnell, and M. Watson, “A buffer-based approach to rate adaptation: Evidence from a large video streaming service,” in Proc. ACM SIGCOMM, 2014.
[14] X. Liu, F. Dobrian, H. Milner, J. Jiang, V. Sekar, I. Stoica, and H. Zhang, “C3: Internet-Scale Control Plane for Video Quality Optimization,” in Proc. NSDI, 2015.
[15] Y. Sun, X. Yin, J. Jiang, V. Sekar, F. Lin, N. Wang, T. Liu, and B. Sinopoli, “CS2P: Improving Video Bitrate Selection and Adaptation with Data-Driven Throughput Prediction,” in Proc. ACM SIGCOMM, 2016.
[16] S. Singh, J. G. Andrews, and G. de Veciana, “Interference shaping for improved quality of experience for real-time video streaming,” IEEE Journal on Selected Areas in Communications, vol. 30, no. 7, pp. 1259–1269, 2012.
[17] I. Rec, “P. 800: Methods for subjective determination of transmission quality,” International Telecommunication Union, Geneva, 1996.
[18] A. Saul and G. Auer, “Multiserver resource allocation maximizing the perceived quality,” EURASIP Journal on Wireless Communications and Networking, vol. 2009, no. 1, p. 341689, 2009.
[19] B. Li, S. Li, C. Xing, Z. Fei, and J. Kuang, “A qoe-based ofdm resource allocation scheme for energy efficiency and quality guarantee in multiserver-multiservice system,” in Globecom Workshops (GC Wkshps), 2012 IEEE. IEEE, 2012, pp. 1293–1297.
[20] W.-H. Kuo and W. Liao, “Utility-based resource allocation in wireless networks,” IEEE Transactions on Wireless Communications, vol. 6, no. 10, 2007.
[21] Y. H. Cho, H. Kim, S.-H. Lee, and H. S. Lee, “A qoe-aware proportional fair resource allocation for multi-cell ofdma networks,” IEEE Communications Letters, vol. 19, no. 1, pp. 82–85, 2015.