Video delivery optimization in cellular heterogeneous wireless networks
Βελτιστοποίηση μετάδοσης βίντεο σε κυψελωτά ετερογενή ασύρματα δίκτυα

Diploma thesis
by
Galanopoulos Apostolos

University of Thessaly
Department of Electrical and Computer Engineering

Supervisors:
Argyriou Antonios
Katsavounidis Ioannis

Volos 2014

Ευχαριστίες

Η παρούσα εργασία υλοποιήθηκε για την απόκτηση διπλώματος στο Τμήμα Ηλεκτρολόγων Μηχανικών και Μηχανικών Υπολογιστών του Πανεπιστημίου Θεσσαλίας. Θέλω να ευχαριστίσω τους καθηγητές, συναδέλφους και γενικά όλα τα μέλη της πανεπιστημιακής κοινότητας του Πανεπιστημίου Θεσσαλίας για τις πολύτιμες γνώσεις κι εμπειρίες που μου έδωσαν κατά τη διάρκεια των σπουδών μου. Ιδιαίτερες ευχαριστίες θα ήθελα να εκφράσω στον καθηγητή μου Αργυρίου Αντώνιο, επιβλέποντα αυτής της εργασίας, για την υποστήριξη που μου έδωσε καθ' όλη τη διάρκεια της προσπάθειάς μου. Επίσης στην οικογένεια μου, για την αμέριστη συμπαράσταση και ενθάρρυνση της σε κάθε μου επιλογή. Τέλος ευχαριστώ τους φίλους μου για όλες τις όμορφες στιγμές που περάσαμε ως φοιτητές.

Abstract

The purpose of this project is to describe and analyze a framework for optimized delivery of video files to users of a heterogeneous (LTE) wireless network. It explains the basic features of such a network and how these can be utilized in order to deliver video data efficiently. This efficiency is meant in terms of users' perceived video quality as well as the minimum possible power consumption, from the network operator's point of view. The method followed was basically the formulation and solving of two optimization problems. The first one decides the optimal association of the users to one of the network's base stations and the throughput each one will achieve consequently. The second problem deals with optimal, in respect to power consumption, resource allocation so that the network's users can achieve the throughput goal set by the solution of the first problem.

Περίληψη

Σκοπός αυτής της εργασίας είναι να περιγράψει και να αναλύσει ένα πλαίσιο για τη βέλτιστη μεταφορά αρχείων βίντεο σε χρήστες ενός ετερογενούς ασύρματου δικτύου. Περιγράφει τα βασικά χαρακτηριστικά ενός τέτοιου δικτύου και πως μπορούν να χρησιμοποιηθούν για να μεταφέρουν δεδομένα βίντεο αποτελεσματικά. Η αποτελεσματικότητα αυτή νοείται ως ποιότητα βίντεο που παρατηρεί ο χρήστης καθώς και ως ελάχιστη δυνατή κατανάλωση ενέργειας από τη μεριά του διαχειρηστή του δικτύου. Η μέθοδος που ακολουθήθηκε ήταν η διατύπωση και επίλυση δυο προβλημάτων βελτιστοποίησης. Το πρώτο αποφασίζει τη βέλτιστη σύνδεση των χρηστών με έναν εκ των σταθμών βάσης του δικτύου καθώς και το πλήθος των δεδομένων που μπορεί ο καθένας να λάβει. Το δεύτερο πρόβλημα αφορά την εκχώρηση πόρων με βέλτιστο τρόπο από άποψη κατανάλωσης ενέργειας, έτσι ώστε οι χρήστες του δικτύου να μπορούν να λάβουν τα δεδομένα που αποφάσισε η λύση του πρώτου προβλήματος.

Contents

List of Figures

List of Tables

1 Introduction

It is known that video data is the type of data that covers most of the capacity of modern wireless networks. Studies show that this situation will continue to grow, thus it is necessary to provide efficient methods to transfer video files in such networks. LTE has greatly improved the data rates, so the users of such wireless networks can watch videos in very good quality. Video quality is a parameter affected by two factors while streaming through a wireless network. The first obvious one is the encoding rate provided by the server streaming the video. A higher encoding rate means a larger file for the server to transmit. The second thing that affects video quality is the number of freezes a user experiences. A freeze occurs when the buffer that stores the video frames to be displayed empties, and thus the video playback stops until the next frames arrive at the buffer. In order to avoid experiencing such freezes, the serving base station has to provide the user with enough data so that the buffer never empties. Rapid changes of the wireless channel can cause fluctuations at the data rate a user experiences so we need to find ways to adapt to the channel's changes. From the above it is clear that in order to provide good video quality to the users of the network the operator has to provide them with as large data rates as possible. Increased data rates though, mean increased power consumption, which is a drawback for the operator. The power consumed now consists of two factors. Firstly, the power required to transmit chunks of the video file from the serving base station to the end user and secondly, the power consumed by the backhaul links in order to deliver the data from the server to the base station associated with the user. Both of these terms of power consumption are being accounted for in this analysis.

1.1 Motivation

From what was mentioned above, it is clear that methods for adaptive streaming are needed to overcome the difficulties that derive from the wireless channel's unstable nature. Dynamic Adaptive Streaming over HTTP (DASH) [1] is such a standardized technique for adaptive streaming that has been used to deliver video content to users of wireless networks requesting such content from web servers. DASH is a standard and thus not implementation specific. Each video a server has stored, is encoded in several bit-rates. A client requesting a video chooses one of the available bit-rates for each chunk of video, depending mostly on average throughput experienced during the previous chunk delivery as well as the buffer's current capacity. As mentioned before, the actual DASH implementation is not standard so the exact rule used to choose the chunk's bit-rate is open for trials. A simple decision policy would be this: Choose the bit-rate that is just less than the throughput experienced in the previous chunk unless the buffer's size drops below a certain threshold. If the latter occurs, then choose the lowest possible bit-rate. Such a method is implemented so that the user can watch the video in a quality as good as the quality of the channel it shares with the base station while trying to avoid buffer under-runs too.

By adopting DASH in a Heterogeneous Network (HetNet) one can see that the throughput each user experiences is affected by a factor that DASH does not consider explicitly, the load of the serving base station. This information is not available to the users and can't help them make a good bit-rate decision. Thus the decision for the bit-rate of each chunk of video, can be made at the base stations, which also receive Channel Quality Indicator (CQI) very often from the users associated to them. This way optimal decisions can be made since all the required information exists at the base stations. The throughput provided to each user can be found, and also the addition of power consumption in the function is imperative, since it is a very important factor for the base stations, as was mentioned before.

1.2 Related work

There is extensive research in the field of HetNets and resource allocation, though not in the context of video delivery and backhauling cost that are important parts of this study. In [2] the authors provide a framework for optimal resource allocation and user quality of experience. The authors in [3] propose an algorithm that stands between function optimization and biasing rule to balance the load of a HetNet. In [4] and [5] user association and resource allocation are studied jointly. In [6] the authors develop a framework to analyze user association in order to maximize the sum rate of the users.

The rest of the report is formed as follows: Section 2 describes the basic features of a HetNet, gives an insight of LTE's downlink resource grid and backhauling techniques to deliver video files. In section 3 the first optimization problem is formulated, in which the users' association as well as the amount of video data each one will receive are defined in the time scope in which we solve the problem. In section 4 the solution of the previous section's problem is used as input, to formulate the resource allocation problem that will decide at what power the base stations will transmit data to the users each time. Section 5 demonstrates and analyzes a numerical analysis and finally, section 6 concludes with the results of this project and propositions for future work.

2 HetNets and video characteristics

In this section the main features of a heterogeneous network that are needed to understand how such a network operates are displayed, in order to be able to formulate the optimization problems that follow. These features include mostly the description of the downlink resource grid of LTE and the backhauling techniques that are used to transfer data to the users of the network. In addition, video files encoding to different bit-rates is analyzed and how this relates to the data rate provided to the users.

Firstly, some basic characteristics and terms that are often used when referring to heterogeneous networks are mentioned. But what makes a cellular wireless network to be heterogeneous? Heterogeneity characterizes a network when different types of serving base stations are deployed within a cell by the operator. These types usually are:

Macrocells, cover a large area with a radius of about 1 kilometer and a high maximum transmission power of about 20 Watts.

Picocells, (also known as smallcells) are low power operator deployed cells that transmit at an order of magnitude smaller power than macro base stations. They are usually deployed in hot-spots to offload the macro base station.

Femtocells, that are more or less like picocells. The difference is that femtocells are deployed by the users in an indoor environment and are accessed only by subscribed users.

Relays, are operator deployed and in contrast with the other base stations, backhaul their traffic through the wireless medium. Typically, they are used to strengthen the signals transmitted by other base stations by retransmitting them after some amplification.

Femtocells and relays are out of the scope of this study for the following reasons. Femtocells are unique since they are not operator deployed. Relays are much more different from every other type as they do not backhaul traffic through a wired medium and this study's backhauling focuses on wired techniques.

Each cell covers the area that the corresponding Base Station (BS) has the ability to cover depending on its maximum transmission power. Each user, or User Equipment (UE) as is often called in LTE, is associated to one of the BSs that can cover its location. This decision is usually made by measuring the Signal to Interference and Noise Ratio (SINR) that each UE gets from each BS. The fact that the macro BS can transmit in a much

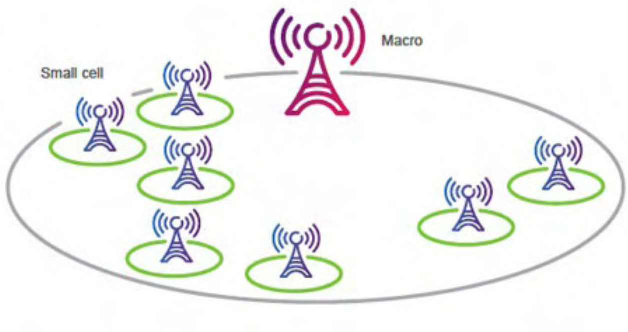

Figure 1: Heterogeneous Network

higher power than pico BSs may lead UEs to associate to the former. This will increase the load on the macro BS and will make the deployment of picocells almost useless. For this reason Cell Range Expansion (CRE) was introduced in LTE. UEs bias the SINR values they get from pico BSs and might choose to associate to one of them, instead of choosing the macro BS even if they get a bigger SINR value from it. Studies show that CRE can improve the efficiency of a HetNet but an optimal association for UEs can be found by embedding the decision in an optimization problem to ensure that UEs enjoy the best possible data rate.

2.1 LTE resource grid

In order to understand how an LTE scheduler works, one should know the structure of the resource grid. This grid is actually a frequency/time domain from which the scheduler chooses the pieces that allocates to the users. This means that the users can only use the downlink channel with the serving base station (macro or pico) only in frequency/time slots that are allocated by the scheduler.

LTE is based on OFDM. The overall spectrum available varies from 1.4 to 20 MHz and is divided to chunks of 12 subcarriers. The number of these chunks is standard for each value of the overall spectrum. In the time domain now, the largest chunk of time is the frame with a duration of 10 msecs. Frames are divided to subframes that last 1 msec, thus each frame consists of 10 subframes. Subframes are further divided to 2 slots, each one lasting 0.5 msecs and can transmit 7 OFDM symbols. A Physical Resource Block (PRB) is a piece of the frequency/time domain that consists of 12 subcarriers in the frequency domain and 7 OFDM symbols (1 slot) in the time domain. The smallest unit of resource allocation in LTE is a pair of PRBs that last 2 slots

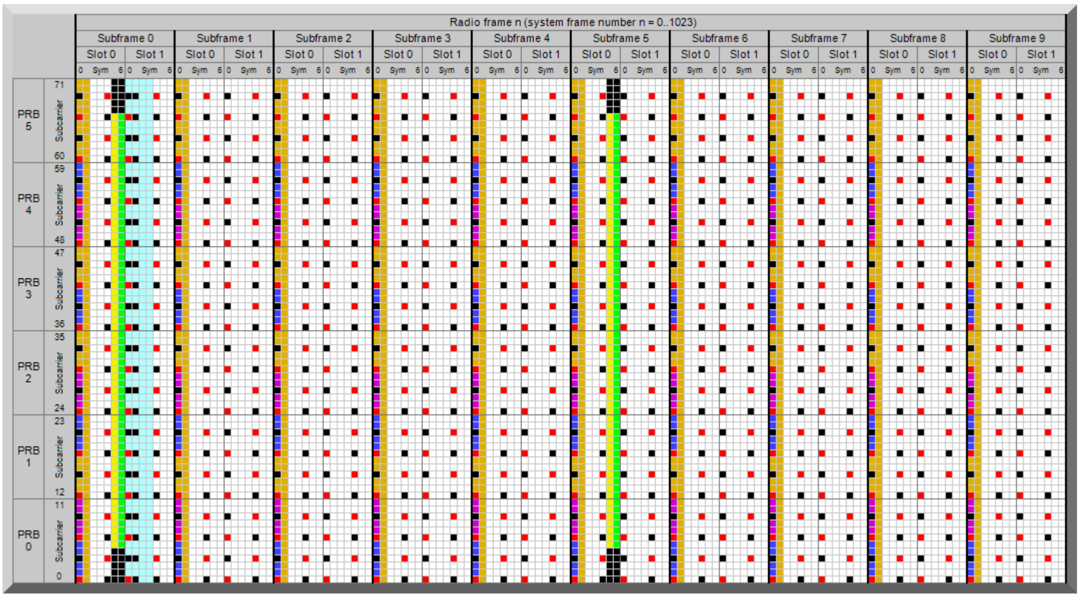

Figure 2: LTE downlink resource grid at 1.4 MHz

or 1 subframe and will be simply referred to as Resource Block (RB) from now on. Depending on the system spectrum now, one can easily calculate the total number of RBs available for the duration of an LTE frame. For example, in a 3 MHz system, there exist 15 PRBs in 1 slot, 15 RBs in 1 subframe and 150 RBs in 1 frame.

2.2 Traffic backhauling

Each cell of the deployment needs to receive all the data the users associated to it request. The server that contains the video file requested is usually far away from the cell and the base stations rely on the backbone network so that this data can reach them. Usually, only the last transmission from the BS to the end user is wireless. The cells require some wired backhaul deployment to deliver data to all BSs within a macrocell. As shown in [9] there are several backhaul architectures with respective mathematical models for the calculation of power consumption. The star topology is the more efficient regarding power consumption and thus is used in this study. A central switch is used to aggregate data destined for the cell and are forwarded to the appropriate BS. Each BS is connected to exactly one interface of the switch, so the rate that each BS can provide to the associated UEs is limited by the interface's rate. If more BSs than the total number of interfaces are deployed,

Parameter	Value	Summary
P_{dl}	1 W	Power consumed by one downlink interface of the aggregation switch
max_{dl}	24	Number of downlink interfaces of the aggregation switch
$P_{max,s}$	300 W	Maximum power consumption of the switch
C_{max}	24 Gbps	Maximum amount of traffic the switch can handle
$C_{interface}$	1 Gbps	Data rate of a single downlink interface

Table 1: Backhaul switch parameters

a second switch is required to backhaul traffic. In [9] and [10] the authors also provide the parameters of the backhaul switch. Those parameters are displayed in Table 1.

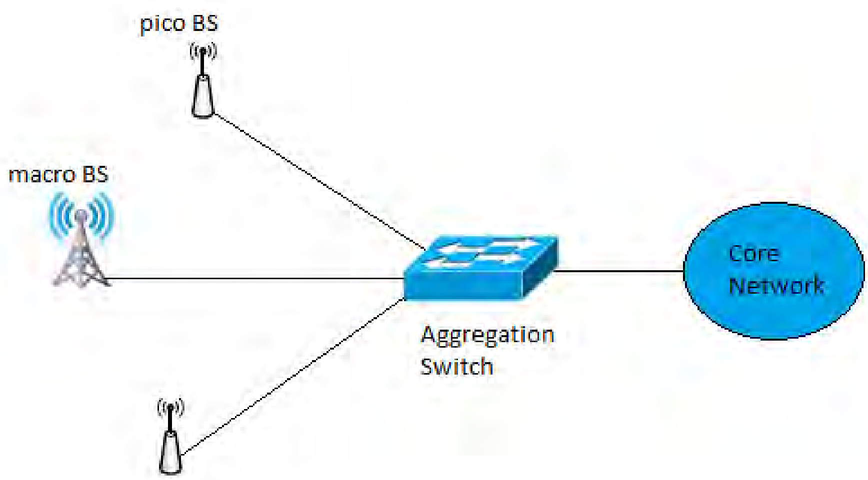

Figure 3: Star backhaul architecture

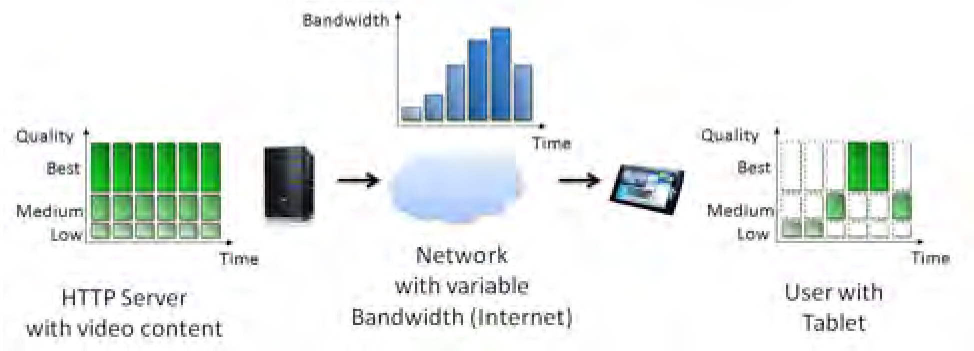

Figure 4: Adaptive bit-rate streaming demonstration

2.3 Adaptive bit-rate video streaming

Adaptive video streaming entails a set of tools that allow the client of an http application that streams video via a web server, to choose between multiple available bit-rates and resolutions. Thus, the client can enjoy as good video quality as the current channel quality and also avoid playback interruption when buffer under-runs occur. A scenario of how a DASH client would work is displayed in Figure 4. The client depending on the network's bandwidth requests video chunks through http. If network conditions allow it, better quality chunks with respective larger size in bits are requested. The process begins when the client obtains the Media Presentation Description (MPD) file. This file contains the different resolutions and bit-rates the video segments are available. More specifically, video resolution is defined through a width/height pair. An example of a MPD file is shown in Figure 5. One can notice the Representation tags that contain information about the bandwidth the width and height of each video segment that follows. There may be multiple triplets of width/height/bit-rate available and there is a method to measure the quality of the displayed video using this triplet. To measure it though, the acceptable frame rate needs to be defined, so that the video has a physical flow for the human eye. A typical value for the frame rate is around 30 frames per second (fps). The value that measures the quality is bits per pixel and is calculated as:

pixels per second = Width x Height x Frame Rate
bits per pixel = bit-rate / pixels per second

The higher the bits per pixel value, the better the video quality. Usually for certain values of resolution and frame rate, the bit-rate is decided so that the bits per pixel value is above a threshold to ensure good video quality. Usually, a bits per pixel value around 0.1 gives a good video quality [11]. For example, for a 1280 x 720 resolution video and a 30 fps value, 2765 kbps will give a good encoding bit-rate. Since the bit-rate decides the quality, if the available bandwidth suffices and the BS can provide this bit-rate to the user, the highest available encoding rate is chosen. To sum up, there are two boundaries we set to the chosen video encoding rate. The lowest one, ensures a smooth video playback without freezes and the highest one, is limited by the BSs ability to provide certain bit-rates to its associated users. The next section proposes a solution to calculate the latter for each user. If there is no encoding rate available between the two boundaries, the lowest one is relaxed enabling the BS to chose an even lower encoding rate providing lower quality while ensuring no freezes. In any case the bits per pixel value should not drop below 0.03 that will make the video unwatchable.

```
<?xml version="1.0" encoding="UTF-8" ?>
- <MPD type="Live" availabilityStartTime="2010-07-01T05:00:00Z"
    availabilityEndTime="2010-07-08T05:00:00Z" mediaPresentationDuration="PT2H"
    minimumUpdatePeriodMPD="PT10S" minBufferTime="PT10S"
    timeShiftBufferDepth="PT30M" baseUrl="http://www.example.com/"
    xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
    xsi:schemaLocation="urn:3GPP:ns:PSS:AdaptiveHTTPStreamingMPD:2009
    3GPP-MPD-r1.xsd"
    xmlns="urn:3GPP:ns:PSS:AdaptiveHTTPStreamingMPD:2009">
- <ProgramInformation moreInformationURL="http://www.example.com">
  <Title>Example</Title>
  <Source>Example</Source>
  <Copyright>Example</Copyright>
  </ProgramInformation>
- <Period start="PT0S" segmentAlignmentFlag="true" bitstreamSwitchingFlag="true">
- <Representation bandwidth="239000" width="320" height="240" lang="en"
    mimeType="video/3gpp; codecs=avc1.42E00b, mp4a.40.2">
- <SegmentInfo duration="PT10S">
  <InitialisationSegmentURL sourceURL="p1rep1.3gp" range="0-985" />
  <Url sourceURL="p1rep1.3gp" range="985-293761" />
  <Url sourceURL="p1rep1.3gp" range="293761-592501" />

          . . .

<Url sourceURL="p1rep1.3gp" range="17304004-17600064" />
  <Url sourceURL="p1rep1.3gp" range="17600064-17894640" />
  </SegmentInfo>
  </Representation>
- <Representation bandwidth="892000" width="480" height="240" lang="en"
    mimeType="video/3gpp; codecs=avc1.42E015, mp4a.40.2">
- <SegmentInfo duration="PT10S">
  <InitialisationSegmentURL sourceURL="p3rep3.3gp" range="0-985" />
  <Url sourceURL="p3rep3.3gp" range="985-1126190" />
  <Url sourceURL="p3rep3.3gp" range="1126190-2228575" />

          . . .

l sourceURL="p3rep3.3gp" range="64712374-65844317" />
  <Url sourceURL="p3rep3.3gp" range="65844317-66966044" />
  </SegmentInfo>
  </Representation>
  </Period>
  </MPD>
```


Figure 5: MPD file example

3 Association and data delivery maximization problem

In this section, an optimization problem that will decide the best possible association for the UEs of the network to one of its BSs is formulated. This decision will have effect on the problem for a duration of 10 seconds. This means that the solution of the problem will determine both the association and the amount of data, that each UE expects to receive in the next 10 seconds.

3.1 Optimization theory prerequisites

In order to understand the optimization problems that follow it is important to provide some insight on optimization theory. An optimization problem consists of a multi-variable function which value should be either minimized (minimization problem) or maximized (maximization problem). This function is called the objective function. The rules that must apply for a solution of the problem to be feasible are called constraints. Constraints are simply relations (equalities and/or inequalities) between the problem's variables. One is the optimal solution of the problem when:

1. it is a feasible solution, meaning that all constraints are satisfied
2. the objective function takes its optimal value, minimum or maximum, depending on the problem

There are several types of optimization problems. The most common one is when the objective function and the constraints are linear functions of the problem's variables. This is called a Linear Program (LP). The problems we are going to face though are more complicated than a LP because they involve a concave function either in the objective, or in the constraints. This is called a Convex Program. In addition, some of the variables take only integer values (binary to be more specific) in both problems. In the first one, these are the variables that decide if UE n will be associated to BS i. The rest of the variables indicate the average transmission power. In the second problem, integer variables are used to decide if RB m will be allocated to UE n. The other variables indicate the transmission power of each resource block. When a problem contains both continuous and integer variables is called a Mixed Integer Program. If a convex program contains integer variables however, it becomes a Non Linear Program. So, both problems that will be studied are characterized as Mixed Integer Non Linear Programs (MINLP), which is a relatively difficult type of problem.

3.2 Problem formulation

In this model a network that consists of N UEs and I BSs (macro and picos) is considered. Each BS has a number of M RBs that can be allocated to the UEs associated to it. The assumption that these RBs are distributed to and not shared by the BSs, so we can avoid interference is made in the study. The solution of the problem that follows, aims to maximize the utility of each user, while respecting the total power consumed by the operator (radio and backhaul). The variables and constants used in the problem formulation are summarized in Table 2.

$$\max_{\boldsymbol{y},\boldsymbol{P}} a \sum_n a_n \log D_n - b(\sum_i P_i^{tx} + P_i^{bh}) \tag{1}$$

subject to:

$$D_n = \lfloor \frac{M}{K_i} \rfloor W_b \log(1 + \frac{\sum_i \overline{h_{ni}} P_{ni}}{\sigma^2}),\ \forall n \in \mathcal{N} \tag{2}$$

$$y_{ni} \in \{0, 1\} \tag{3}$$

$$\sum_i y_{ni} = 1,\ \forall n \in \mathcal{N} \tag{4}$$

$$P_i^{tx} = \sum_{n \in N_i} P_{ni},\ \forall i \in \mathcal{I} \tag{5}$$

$$P^{bh} = \lceil \frac{I}{max_{dl}} \rceil c P_{max,s} + \sum_i P_i^{bh} \tag{6}$$

$$P_i^{bh} = (1-c) \frac{\sum_{n \in \mathcal{N}_i} D_n}{C_{max}} P_{max,s} + P_{dl},\ \forall i \in \mathcal{I} \tag{7}$$

$$\sum_{n \in \mathcal{N}_i} D_n \leq C_{interface},\ \forall i \in \mathcal{I} \tag{8}$$

$$K_i = \sum_{n \in N_i} y_{ni},\ \forall i \in \mathcal{I} \tag{9}$$

$$0 \leq P_{ni} \leq y_{ni} P_{max,i} \tag{10}$$

The objective function is a weighted sum of the UEs utility that models video quality and the network's power consumption. The use of the logarithmic function on D_n is made in order to encourage an increase in the value of D_n when it is small and discourage it as it increases. The parameter a_n for

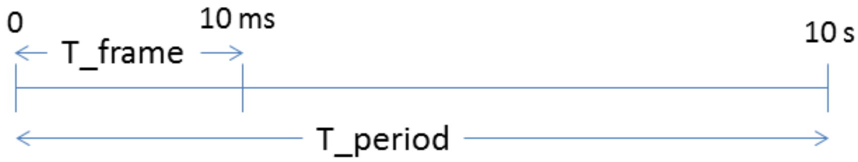

Figure 6: Problem time scale

user n can be set according to the quality of the requested video content. As stated in [12], it usually depends on the motion of the video content, meaning that the more motion the video contains the more quality it demands so that it will be played in a pleasing for the human eye manner. The result is that sequences that have higher quality will have a higher value for a_n and they will eventually receive higher fraction of the resources. In constraint (2) the term in front of the logarithm ensures equal allocation of the M RBs between the users associated to each BS. The number of resource blocks M, is the same for all BSs since we assume orthogonal distribution among them. The expression that produces D_n comes from Shannon's theorem that specifies the maximum rate of a communications channel ($C = B\log(1 + SNR)$). (3) and (4) ensure that each UE will be associated to exactly one BS. Constraint (5) gives the average radio power consumed by BS i. Constraints (6) and (7) are about the total backhaul power consumed. It consists of 2 factors weighted by variable c. The first one is a standard operation cost of the switch. The second is proportional to the rate provided to each BS's users [10]. Constraint (8) ensures that the total rate provided to the UEs associated to each BS i is less than the interface rate of the switch used for backhauling. Constraint (9) calculates the number of users associated to BS i. Finally, in (10) the power in which BS i will transmit to UE n is constrained between zero and the maximum power BS i can transmit.

The solution of the above problem gives us the association decisions for the users, the average power for each user during the 10 second period (the vector $\boldsymbol{P}$), as well as the users expected data rate. This solution applies for a duration of 10 seconds or as shown in Figure 6, T_period. Having solved this problem, the resource allocation problem can be studied.

Parameter	Summary
D_n	The data rate enjoyed by user n
a_n	Video stream related value of user n
$\overline{h_{ni}}$	Average value of the channel between user n and base station i
P_{ni}	The average power allocated to each resource block of user n by base station i
P_i^{tx}	Average Radio power consumed by BS i
P_i^{bh}	The power consumed at the backhaul switch due to the rate provided to base station i to serve the users associated to it.
$\mathcal{I}$	Set of base stations, macro and picos
I	Total number of base stations, i.e. $\lvert\mathcal{I}\rvert$
$P_{max,i}$	The maximum transmission power of BS i
W_b	Bandwidth of a Resource Block
$\mathcal{N}$	Set of users in the cell
$\mathcal{N}_i$	Set of users associated to BS i
M	The number of resource blocks available to the BS for allocation
K_i	The number of UEs associated to BS i
y_{ni}	Takes the value of 1 if user n is associated to BS i and 0 otherwise

Table 2: Parameters description

4 Resource allocation problem

The time scope of this problem is the duration of an LTE frame (10 ms). The amount of data each user will receive has already been decided, so now it has to be made certain that the users will be receiving enough data at each frame so that the "promise" made to the users in the solution of the first problem is respected. The problem is solved every 10 msecs during the 10 seconds period in which the association and data demands apply. The reason that this problem has to be solved repeatedly and not just use the power solution of the previous problem is that a CQI feedback from UEs is available in every frame (7-8 msecs) and thus the exact power needed to deliver the required amount of data can be allocated. Remember that in the previous section's problem equal allocation of RBs among the associated UEs of a BS was assumed. This is generally not true and the fluctuations of the channel require that RBs should be allocated accordingly. In Table 3 there is the description of some additional parameters used in the resource allocation problem formulation that follows.

$$\min_{\boldsymbol{x},\boldsymbol{P}} \sum_i P_i^{tx} + P_i^{bh} \tag{11}$$

subject to:

$$r_n = \sum_m W_b \log(1 + \frac{h_n P_{mn}}{\sigma^2}),\ \forall n \in \mathcal{N} \tag{12}$$

$$r_n 10^{-2} \geq D_n^{frame},\ \forall n \in \mathcal{N} \tag{13}$$

$$R_i = \sum_{n \in N_i} r_n,\ \forall i \in \mathcal{I} \tag{14}$$

$$P_i^{tx} = \sum_m \sum_{n \in N_i} P_{mn},\ \forall i \in \mathcal{I} \tag{15}$$

$$P_i^{bh} = aR_i + b,\ \forall i \in \mathcal{I} \tag{16}$$

$$\sum_{n \in N_i} x_{mn} \leq 1,\ \forall\ m \in [1, M] \tag{17}$$

$$x_{mn}\ \in\ \{0, 1\} \tag{18}$$

$$0 \leq P_{mn} \leq x_{mn} P_{max,i} \tag{19}$$

Parameter	Summary
r_n	Total rate enjoyed by user n
h_n	Channel between BS i and user n
D_n^{frame}	Amount of data to be delivered to user n during a frame
R_i	Total rate BS i provides to the associated users
P_{mn}	Radio power allocated to user n for Resource Block m
x_{mn}	Takes the value of 1 if RB m is allocated to user n and 0 otherwise

Table 3: Resource allocation additional parameters

Due to the orthogonal allocation of RBs among the BSs, the problem can be solved separately and independently for each BS. The association of UEs has already been decided, as well as D_n^{frame} for each one. The actual rate for user n will be according to (12). It is the sum rate over all RBs that can be allocated to n. Constraint (13) satisfies that r_n will be sufficient enough to cover the user's demand in data, while constraint (14) calculates the total rate BS i provides to its associated users. In (15) the total power consumed by BS i is the sum of powers it allocates to its RBs. Equation (16) states that the backhaul power consumed by BS i is a linear function of the total rate it supplies. This is similar to constraint (7) of the previous section's problem, only now $\sum_{n \in \mathcal{N}_i} D_n$ is replaced by R_i which is a more accurate value of the rate BS i provides its users with. Constraints (17) and (18) mean that each RB can be allocated to only one UE. Finally, (19) restricts P_{mn} to take value from 0 (if x_{mn} is 0) to $P_{max,i}$.

5 Numerical analysis

In this section results from simulations made to evaluate the behavior of the proposed framework are shown. The implementation was made using the "Opti Toolbox" matlab optimization toolbox [13]. The solver used, more specifically was BONMIN, which solves smooth twice differentiable Mixed Integer Non Linear Programs. In these simulations the evaluation of the performance of a heterogeneous cellular network is made in respect to users achieved data rate and network's average power consumption.The simulations concern a macrocell and several picocells and UEs randomly spread in a 3x3 kilometer area. The system's bandwidth chosen is 3 MHz which means 150 RBs are to be distributed among the Base Stations. The bandwidth W_b of each RB is 200 KHz. The biasing factors of the first problem's objective function a and b are chosen empirically as 2 and 1 respectively.

5.1 Picocell deployment

Firstly, the system's performance for different operator deployments concerning the number of picocells is displayed. In Figure 7 one can see how the users' average data rate relates to the number of users served by the cell. The proposed framework reduces the data rate provided to the users as they increase, in order to save power. The total number of RBs is the same in case of 4 picocells as is in the case of 5, but we can see a slight increase in the average rate because 1 more picocell can give more flexibility on which BS will serve the users.

In Figure 8 the impact of the number of picocells in total power consumption is shown. This power consumption concerns radio transmissions by the BSs as well as backhaul power consumption. It increases as the number of users increases since we are trying to provide them with enough data rate using the same number of RBs, but after a while the increase rate drops as the power consumption increase is a drawback in our problem's objective function. At this point it is made clear that this power consumption that concerns a LTE frame is not based on the power allocation problem, but is calculated from the first optimization problem that decides the users rate and transmission power based on the average channel quality value observed in the past. This means that it is not a power consumption during some frame but an estimation of the average frame power consumption. A much better performance when adding a fifth picocell in our deployment can be observed as well, since it can offload other BSs that don't share good quality channel conditions with some users and consume less power.

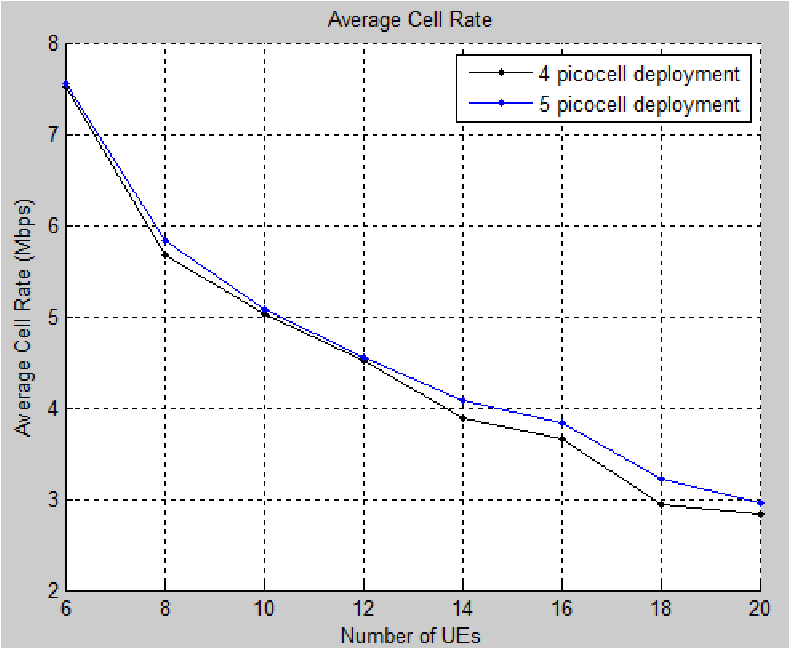

Figure 7: Average Cell Rate I

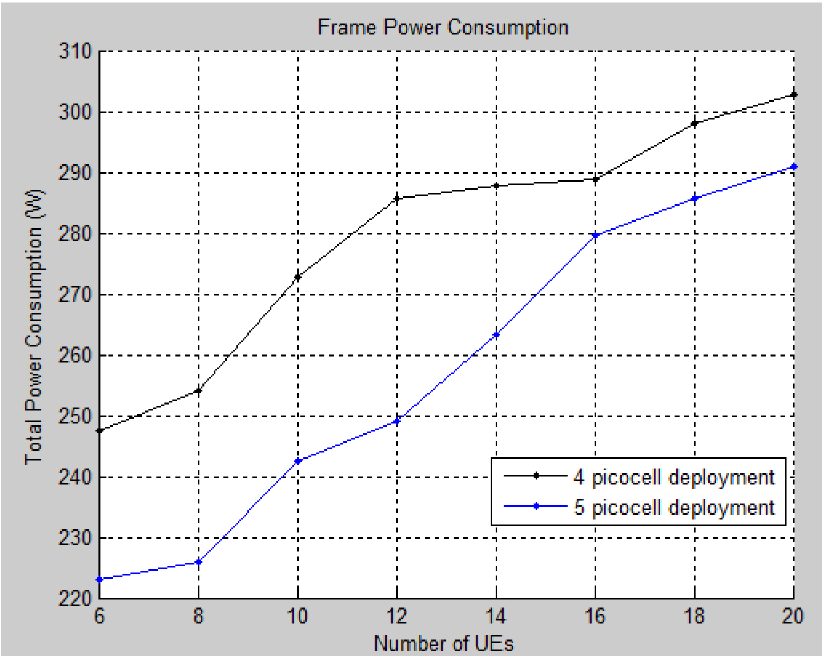

Figure 8: Frame Power Consumption I

5.2 Biasing data rate and power consumption

The more UEs within a cell, the less data rate is provided in order to save power. But what if the operator wants to save even more power or on the contrary, offer a better streaming service to the UEs? It is all decided by the objective function's biasing factors a and b, or otherwise the fraction a/b. In the previous subsection the value used for a/b was 2. In this subsection the effect on data rate and power consumption when using different values for a/b is studied. The deployment for the following simulations is 4 picocells plus 1 macrocell.

In Figure 9 the average cell data rate relation with the number of UEs is displayed for different values of a/b. As a/b increases, more weight goes to the users data rate rather than the operator's power consumption.

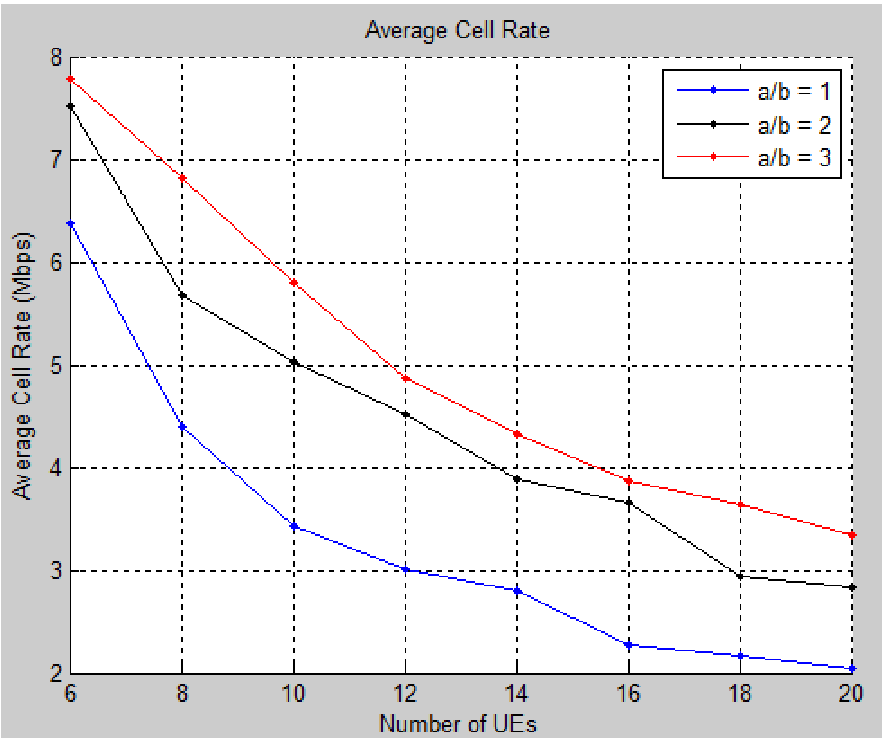

Figure 9: Average Cell Rate II

The impact of the increased data rate as a result of a bigger a/b value on power consumption is displayed on Figure 10. As UEs within a cell increase, the total power consumption increases in order to provide a certain level of data rate to them. But as this level of service increases (a/b increases), the

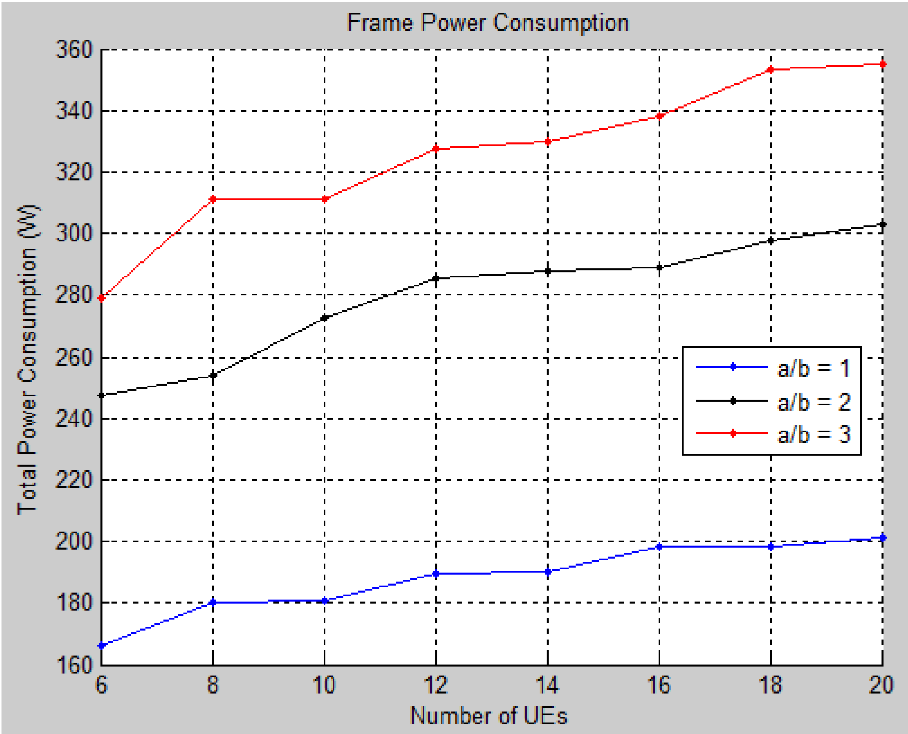

Figure 10: Frame Power Consumption II

total power consumption increases too.

5.3 Resource allocation

After video encoding rate for the deployments users is decided, each base station performs resource allocation to satisfy the data rate required in order to ensure in time delivery of the video segments. This subsection presents the results of such resource allocation by solving the problem formulated in section 4.The following measurements were taken by simulating a network with 4 picocells and 10 users, and presents the data rate provided by each BS as well as their average power consumption during a frame. Comparison is made with the results of the *expected* data rate and power consumption of the solution of the problem in section 3.1.

Figure 11 displays the data rate provided by each base station to its associated users. It is expected that each base station will provide a certain and stable rate for all subframes in a 10 second duration that decides a certain encoding rate for video segment delivery. The resource allocation problem's solution validates this assumption, since every base station provides actual data rate equal to the expected data rate.

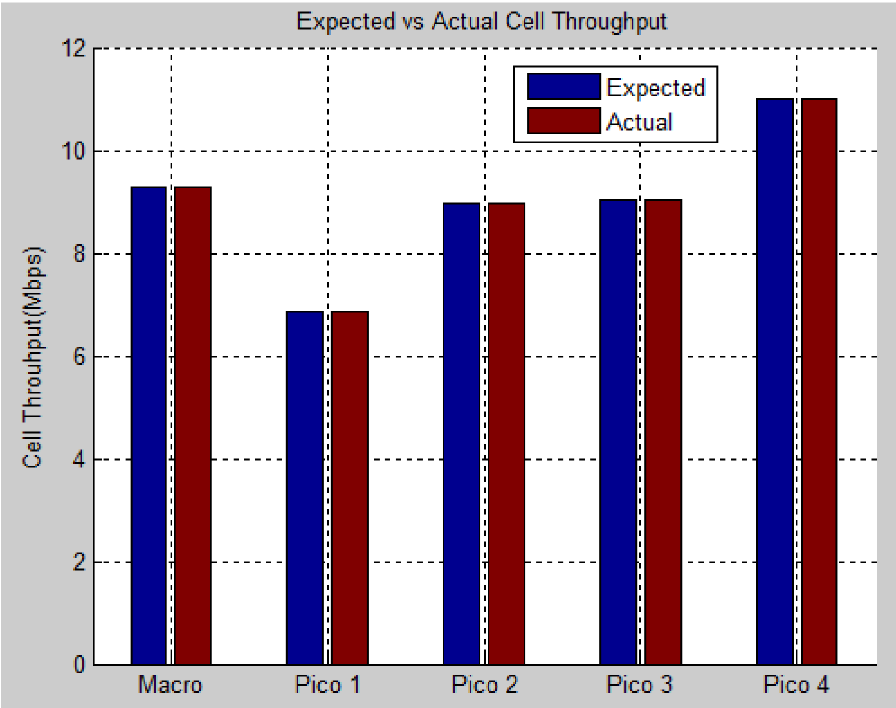

Figure 11: Base station provided rate

In Figure 12 the average radio power consumption during a frame is displayed along with the expected one. Backhaul power consumption is not accounted in this particular measurement since it is only affected by the data rate provided by the base station. Since Figure 11 shows no difference between expected and actual data rate, adding backhaul power to radio power will not make any more difference between expected and actual power consumption. Radio power consumption is affected by the condition of the channel each base station has, with the associated users. We observe similar results between expected and actual power consumption due to the random changes of the channel quality. If CQI values reported by the UEs mostly indicate channel quality improvement, actual power consumption is found less than the expected one and vice versa.

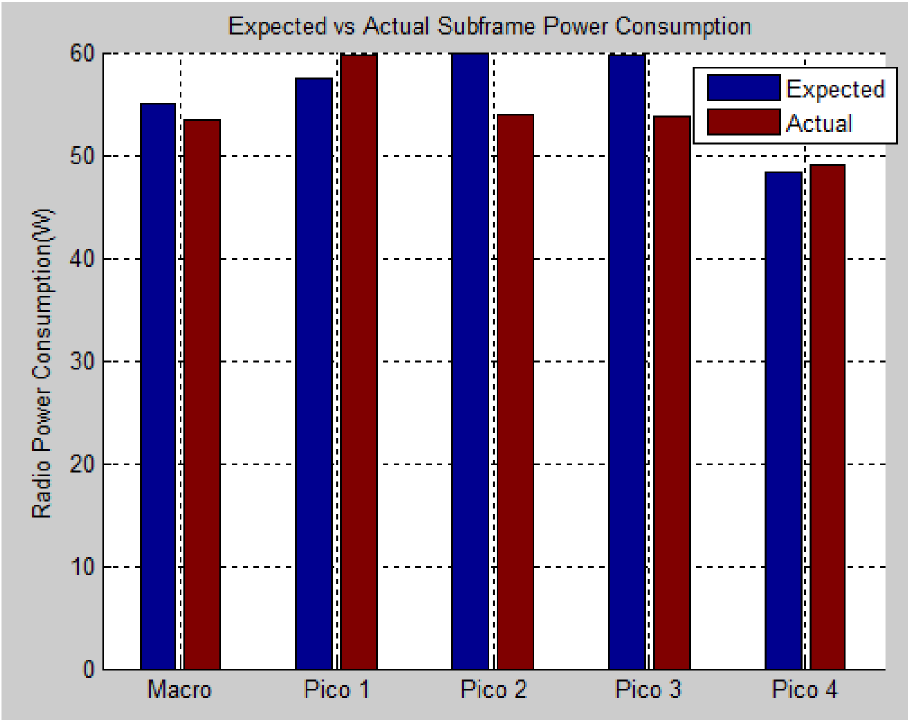

Figure 12: Base station average power consumption

6 Conclusion

This project presented a framework for the delivery of video files to users of a heterogeneous network. Video files are stored in several encoding rates and the operator of the network can decide the encoding rate each user will enjoy, so that the users can watch the requested video without interruptions in playback. These decisions adapt to the channels between the network's base stations and the respective associated users, so that each segment of a video may be transmitted in different encoding rate. As a result, the video quality enjoyed by each user varies as the channel's condition changes. The higher the video quality though, the more power is consumed by the network, both radio and backhaul, so the operator must consider it and balance between the two by providing as good quality as possible while consuming as less energy as possible. This is an obvious trade-off that is closely studied in the proposed framework. Finally, the scheduler in each base station allocates resources according to the encoding rates decided for each video stream, so as to minimize power consumption.

The mathematical modeling in the resource allocation problem and especially equation (13) requires stable data rate for each subframe in a 10 second period. This constraint restricts the scheduler to allocate sufficient power for it to be satisfied. When the channel's condition deteriorates more power is required and vice versa as simulations suggest. With this in mind, it is possible to alter the formulation in a way that when experiencing good channel condition the scheduler can decide to allocate a bit more power, providing a larger instant data rate. On the contrary when experiencing bad channel quality the subframe data rate constraint can be relaxed and thus not waste a lot of power on a bad channel. This may cause buffer under-runs but if the total amount of bits of a 10 second video segment are received, this probability is small and depends on the random channel fluctuations.

References

[1] T. Stockhammer, "Dynamic adaptive streaming over http –: Standards and design principles," in *Proceedings of the Second Annual ACM Conference on Multimedia Systems*, MMSys '11, (New York, NY, USA), pp. 133–144, ACM, 2011.

[2] J. Chen, R. Mahindra, M. A. Khojastepour, S. Rangarajan, and M. Chiang, "A scheduling framework for adaptive video delivery over cellular networks," in *Proceedings of the 19th Annual International Conference on Mobile Computing & Networking*, MobiCom '13, (New York, NY, USA), pp. 389–400, ACM, 2013.

[3] Q. Ye, B. Rong, Y. Chen, M. Al-Shalash, C. Caramanis, and J. G. Andrews, "User association for load balancing in heterogeneous cellular networks," *CoRR*, vol. abs/1205.2833, 2012.

[4] D. Fooladivanda and C. Rosenberg, "Joint resource allocation and user association for heterogeneous wireless cellular networks," *Wireless Communications, IEEE Transactions on*, vol. 12, pp. 248–257, January 2013.

[5] S. Deb, P. Monogioudis, J. Miernik, and J. Seymour, "Algorithms for enhanced inter-cell interference coordination (eicic) in lte hetnets," *Networking, IEEE/ACM Transactions on*, vol. 22, pp. 137–150, Feb 2014.

[6] S. Corroy, L. Falconetti, and R. Mathar, "Dynamic cell association for downlink sum rate maximization in multi-cell heterogeneous networks," in *IEEE International Conference on Communications (ICC 2012)*, (Ottawa, Canada), pp. 2485–2489, June 2012.

[7] `http://wcnc2013.ieee-wcnc.org/WCNC.T7.Slides.pdf`.

[8] `http://www.pewscorner.host-ed.me/LTE/lte_resource_grid.html`.

[9] P. Monti, S. Tombaz, L. Wosinska, and J. Zander, "Mobile backhaul in heterogeneous network deployments: Technology options and power consumption," in *Transparent Optical Networks (ICTON), 2012 14th International Conference on*, pp. 1–7, July 2012.

[10] S. Tombaz, P. Monti, K. Wang, A. Vastberg, M. Forzati, and J. Zander, "Impact of backhauling power consumption on the deployment of heterogeneous mobile networks," in *Global Telecommunications Conference (GLOBECOM 2011), 2011 IEEE*, pp. 1–5, Dec 2011.

[11] http://community.ooyala.com/t5/Developers-Knowledge-Base/Understanding-Bitrate-Resolution-and-Quality/ta-p/1740.

[12] C. An and T. Nguyen, "Analysis of utility functions for video," in *Image Processing, 2007. ICIP 2007. IEEE International Conference on*, vol. 5, pp. V – 89–V – 92, Sept 2007.

[13] http://www.i2c2.aut.ac.nz/Wiki/OPTI/index.php.