

Broadcast Video Streaming in Cellular Networks: An Adaptation Framework for Channel, Video and AL-FEC Rates Allocation

Daniele Munaretto Dan Jurca Joerg Widmer
 DOCOMO Communications Laboratories GmbH
 Munich, Germany
 E-mail: lastname@docomolab-euro.com

Abstract—Video streaming is one of the most important applications that will make use of the high data rates offered by 4G networks. The current video transport techniques are already very advanced, and the more immediate problems lie in the joint optimization of video coding, AL-FEC, and PHY rate selection with the goal of enhancing the user perceived quality. In this work we provide an analysis of video broadcast streaming services for different combinations of layered coding and AL-FEC, using a realistic LTE PHY layer. Our simulation results show that the scalable content adaptation given by Scalable Video Coding (SVC) and the scheduling flexibility offered by the 3G-LTE MAC-layer provide a good match for enhanced video broadcast services for next generation cellular networks. Our proposed solution is compared to baseline algorithms and broadcast systems based on H.264/AVC streaming solutions. We emphasize the system quality improvement brought by our solution and discuss implications for a wide-scale practical deployment.

I. INTRODUCTION

Wireless broadcast video streaming is one of the most important services offered by the next generation of cellular networks. Adaptability to the time-varying channel conditions and robustness against channel losses have an impact on the reliability of the video transmission, and hence on the video quality perceived by the end users. The more immediate problems related to the current advanced video transport techniques lie in the joint optimization of the coding techniques used to compress the streamed video and to protect it against channel erasures, and the channel rates available at the base stations to transport the video packets.

At the application layer, H.264/SVC can be used to compress the video so as to support heterogeneous devices and heterogeneous sets of users, by enabling a source transmit a video stream composed of multiple quality layers. In addition, at the application layer such a video stream can be protected against channel erasures through the use of Forward Error Correction (FEC) techniques. In contrast to ARQ-based techniques, they avoid the extra delays incurred by packet retransmissions. At the physical layer, adaptive modulation and coding schemes are crucial for broadcast video streaming, since they allow a base station to select different channel rates to transmit the video layers. This provides different video quality levels to a heterogeneous set of users, with respect to their channel conditions, increasing the utility of the overall system. This is appealing given that currently, all multicast packets are

inefficiently transmitted at low order modulation and coding schemes, to serve even the users located far from the base station.

In this paper, we design heuristic cross-layer algorithms which make use of higher order modulation and coding schemes for higher quality video layers, to provide differentiated service quality to a heterogeneous set of users, thus improving the overall video quality. To do so, we solve the problem of jointly optimizing the number of scheduled video layers, AL-FEC, and modulation and coding schemes for a broadcast streaming application, based on a mathematical cross-layer framework.

The goal of our framework is to minimize the overall video distortion perceived by the end clients (e.g. mobile phones, PDAs, laptops), making use of:

- the video quality scalability offered by the H.264/SVC codec at the application layer;
- unequal error protection at the application layer (AL-FEC), which increases the robustness of the transmitted information to channel errors;
- the flexible MAC layer scheduling offered by 3G-LTE (different modulation and coding schemes, MCs), realized both in time and frequency domains using OFDMA, to provide efficient broadcast video streaming solutions in a cellular system.

The contributions of this work can be summarized as follows:

- we define a cross-layer optimization framework for a streaming broadcast system based on the optimization of the overall system perceived distortion;
- We derive a heuristic approach for: 1) the computation of the best modulation and coding schemes for each user; 2) the assignment of the modulation and coding scheme to each single video layer scheduled for transmission; 3) the computation of AL-FEC redundancy required for each video layer, taking into account the channel constraints;
- We perform exhaustive simulations for a wide range of settings in order to validate our framework;
- We compare our solution to other broadcast solutions based on H.264/AVC.
- We discuss a possible system architecture based on our cross-layer framework, in which we can implement our

algorithm as a module at the base station in a cellular system;

- We further discuss issues related to a possible wide-scale deployment of our broadcast system by a mobile operator.

The remainder of the paper is organized as follows. In Section II we report on the related work. In Section III we introduce our cross-layer framework and discuss the optimization problem. Section IV presents our heuristic approach to solve the optimization problem. Our delivery architecture is presented in Section V and the simulation setup and results are provided in Section VI. A short discussion on the practical issues to be evaluated in a real deployment is given in Section VII and we conclude the paper in Section VIII.

II. RELATED WORK

In advanced wireless technologies such as 3GPP/LTE [1], [2], broadcast video streaming is considered to be one of the most important applications that need to meet the requirements and capabilities of a wide range of mobile devices, with respect to available computing resources, power consumption, spatial resolution of the screen, rate, etc. Scalable Video Coding (H.264/SVC) [3] supports heterogeneous devices and users with different channel conditions by allowing a source (server) to send a video with scalability provided in either the temporal, spatial, or quality domain [4]. In order to provide protection to the video packets against losses, several FEC techniques have been proposed for broadcast scenarios. Raptor Codes [5], one of the first known classes of fountain codes with linear time encoding and decoding, offer a widely used and highly efficient FEC solution. It has been adopted in 3GPP for mobile cellular wireless broadcast and multicast applications and is also used by DVB-H [6] standards for IP datacast to hand-held devices. The FEC redundancy sent alongside with the original packets can be increased for the most important video layers (Unequal Error Protection, UEP), outperforming regular robustness schemes [7],[8]. In general, the video transport techniques adopted by the standards for broadcast video streaming applications in cellular networks are very advanced. At the same time, several cross-layer studies for robust scalable unicast/broadcast video transmission over WLANs have been proposed, e.g. [9], showing that real-time video quality can be improved by cross-layer signaling [10]. Studies on how to determine which PHY layer modulation and coding scheme the base station should use to broadcast packets were recently presented in [11], but they mainly focus on ARQ-based protection mechanisms rather than FEC techniques.

In this work, we propose a heuristic cross-layer framework which provides differentiated service quality to a heterogeneous set of users in a cellular environment. Our framework is designed to improve overall video quality, optimizing the joint use of scalable video coding, AL-FEC, and modulation and coding schemes. The optimization framework is based on a theoretical distortion model encompassing both the source video distortion and the influence of channel errors. We also present novel algorithms for solving our optimization problem in a tractable way, by isolating the most important factors

influencing the distortion metric. Finally, our results open interesting discussion points related to the practical deployment of our proposed broadcast system by a mobile operator.

III. MODEL

A. System Model

We consider a wireless video broadcast system in which a base station delivers video content to N clients inside its coverage area. The channel used for the transmission has a capacity of C symbols per second, which can be translated into an application rate, based on the modulation and coding scheme (MC_i) used for transmission. We assume that the base station can choose the appropriate MC_i out of a set of available schemes $\mathcal{MC} = \{MC_i | 1 \leq i \leq M\}$. Then, the channel capacity C_i used for transmission is a function of the application rate ρ_i and the modulation and error correction code rate used by the chosen MC_i : $C_i = f(\rho_i, MC_i)$. Each client j in the coverage area of the base station can decode the transmitted information sent with MC_i with a certain error probability $p_{i,j} \leq 1$, based on the observed signal-to-noise ratio of its down-link channel. We assume that the base station obtains feedback about the SNR levels for all its clients and, based on the SNR, can estimate the loss probability observed by each user, for each of the available MC schemes.

The video stream broadcasted by the base station is encoded into L scalability layers using a scalable video encoder, i.e., one base layer, containing the most important video packets for a minimum video quality, and $L - 1$ enhancement layers, each one further improving the quality over the lower layers. We assume that video layer k , of rate ρ_k , can only be decoded at the clients if all lower video layers are successfully decoded, and that each video layer can be truncated before transmission, based on the total available channel rate. The base station has the flexibility of choosing an appropriate $MC(k)$ for each individual transmitted video layer k . Furthermore, it can protect each layer k against channel errors through an application layer forward error correction scheme FEC_k . Finally, we assume that the wired backbone link between the application server, providing the video content, and the base station is over-provisioned and lossless.

B. Video Distortion Model

We represent the end-to-end video distortion, as perceived by one media client j , as an additive metric D_j depending on both the source distortion and the channel distortion (in terms of MSE). Thus, the received video quality depends on the lossy encoding of the media information (D_S) and on the amount of packet loss experienced in the network during the video transmission (D_L).

The source distortion D_S depends on the total video rate $R_v(l)$ of the $l \leq L$ video layers scheduled for transmission by the server. If $f_k \in [0, 1]$ represents the truncating factor of video layer k , then $R_v(l) = \sum_{k=1}^l f_k \rho_k$. In contrast, D_L is roughly proportional in average to the number of lost pixels/video elements. As network packets contain in general data referring to the same amount of video information (e.g., one frame, one slice, or one encoded video layer of a frame),

the channel distortion is proportional to the number of lost packets, and is differentiated by the importance of the video layer containing the lost packets [12]. If $R_v(l)$ is the video rate scheduled for transmission, we can explicitly formulate the video distortion metric as:

$$D_j(l) = D_S(l) + \lambda \pi_1^j + \sum_{k=2}^l (\pi_k^j (D_S(k-1) - D_S(l)) \prod_{s=1}^{k-1} (1 - \pi_s^j))$$

where $D_S(l)$ can be computed based on the encoding parameters and our transmission assumptions, using a linear approximation for truncated layers. π_k^j represents the loss process affecting the transmission of video packets belonging to layer k after FEC decoding, at user j , and λ is a video sequence dependent constant. In our transmission context, packets from different video layers k are transmitted using potentially different modulation and coding schemes $MC(k)$ and are protected by different amounts of FEC (application layer unequal error protection – FEC_k), hence π_k differs from layer to layer.

Finally the total distortion of our broadcast scenario can be computed as the sum of the individual distortions perceived by all users in the system:

$$D(l) = \sum_{j=1}^N D_j(l)$$

Note that the total distortion of our broadcast scenario depends on the total number of clients in the system N , and their respective channel conditions SNR_j , $1 \leq j \leq N$, the number of video layers transmitted l , the chosen modulation and coding scheme for each video layer $MC(k)$, $1 \leq k \leq l$, and the amount of application layer FEC added to each layer FEC_k , $1 \leq k \leq l$.

We validate the distortion model through video experiments, using more encoded video sequences with 300 frames, using the JSVM encoding software [13]. Due to space constraints, we omit the validation results in this paper. However, we refer the interested reader to [14] for a complete discussion of these results.

C. Unequal Error Protection

We use a systematic error correction code which protects each block of s video packets by adding $n - s$ redundant packets, hence creating a FEC block $FEC(n_k, s_k)$ for each video layer k . We assume that the s video packets can be reconstructed as long as no more than $n - s$ packets are lost from the transmitted FEC block. The average loss probability as seen by video layer k packets after FEC decoding can be computed as:

$$\pi_k = \frac{1}{s} \cdot \sum_{i=1}^s i \cdot e_i(n, s),$$

where $e_i(n, s)$ is the probability of losing at least $n - s + 1$ packets from the FEC block, out of which exactly i packets are video packets. Assuming an independent model for the channel packet losses, $e_i(n, s)$ can be easily computed:

$$e_i(n, s) = \binom{s}{i} p^i (1-p)^{s-i} \sum_{l=\max(b+1-i, 0)}^s \binom{b}{l} p^l (1-p)^{b-l},$$

where $b = n - s$. The rate of video layer k protected by a FEC scheme $FEC(n_k, s_k)$ increases to $\rho'_k = \rho_k \frac{n_k}{s_k}$.

In streaming applications, where transmission delay should be kept small, it is common to form short-to-medium length FEC blocks, based on independent blocks from the video bitstream, e.g., s could represent the number of packets in one GOP of one scalability layer in the encoded bitstream.

D. Optimization Problem

Finally, we formulate the optimization problem. We are interested in finding the optimal broadcast strategy at the base station, which minimizes the total video distortion perceived at all the clients, under a given channel capacity constraint. With this respect we must find the optimal number of video layers to be transmitted, the appropriate modulation and coding scheme for each video layer, and the right amount of AL-FEC protection for each transmitted layer. We aim at optimizing our broadcast strategy periodically, in order to reflect the changes in the overall system, be it number of users and their respective channel conditions, availability of different modulation and coding schemes, or changes in the video stream parameters.

Formally, our optimization problem can be stated as follows. Given the total number of users N and their respective channel conditions $SNR_j, \forall j \leq N$, the available modulation and coding scheme set $MC = \{MC_i | i \leq M\}$, the statistics of the video bitstream $\rho_k, \forall k \leq L$ and s , and the channel capacity constraint C , find the optimal resource allocation tuple $(l, MC(k), FEC_k)^*$ with $1 \leq k \leq l$ such that:

$$(l, MC(k), FEC_k)^* = \arg \min_{\forall (l, MC(k), FEC_k)} D(l) \quad (1)$$

under the total channel rate constraint:

$$\sum_{k=1}^l f_k f(\rho_k, MC(k)) \frac{n_k}{s} \leq C$$

Given the dependency of the distortion metric on so many parameters, the optimization problem can become intractable even for small broadcasting scenarios. Hence, in the following section, we detail our fast heuristic approach solving this problem.

IV. ALGORITHMS

In this section we present our heuristic approach and algorithms to solve the optimization problem. Our solution is based on decoupling the factors that influence the behavior of the distortion metric D , and analyzing them separately. Namely, we take three steps in our proposed approach:

- 1) First, we compute the best operational modulation and coding scheme MC_j and its associated packet loss probability p_j , for each individual user j , based on the channel conditions SNR_j . Given MC_j , we can compute the subset of modulation and coding schemes that

can be used by the broadcast system for transmission, that are decodable by user j .

- 2) Next, we compute the appropriate assignment of modulation and coding schemes $MC(k)$ for each transmitted video layer k , based on the first step. We assume that a video layer k transmitted with $MC(k)$ can be decoded (possibly with some errors) by all users j that support a better or equal MC_j , and hence reduces the video distortion for these users. All other transmitted video layers are considered to be undecodable by users j .
- 3) Finally, once the modulation and coding schemes are assigned to each scheduled video layer, we assign the appropriate amount of application layer FEC to each of the layers, taking into account the channel capacity constraint and the trade-off between source and error-correction rate.

A. Selection of the operational MC scheme MC_j

We derive a simple selection algorithm which, based on the SNR level of each user j , assigns an operational MC scheme MC_j , for each active user j . Considering the same desired throughput as perceived by the user j , we compute MC_j as the MC scheme that utilizes the smallest channel symbol rate. Let r be the desired unit throughput at the application layer, as perceived by user j . In case the base station uses a given modulation and coding scheme MC_i for data transmission, user j experiences a packet loss probability $p_{i,j} = g(MC_i, SNR_j)$. Hence, the base station will require an average application layer bandwidth $b_j = \frac{r}{1-p_{i,j}}$ in order to relay a useful rate r to user j . In turn, the total channel capacity needed to transmit b_j is $C_j = f(b_j, MC_i)$. The operational MC scheme is chosen to be the one that minimizes the channel use C_j . Algorithm 1 represents the pseudo-code of the algorithm.

Algorithm 1 Selection of best MC_j for user j .

Input: channel condition at user j , SNR_j , set of available MC schemes, \mathcal{MC} , packet loss probability for each MC scheme, $p_{i,j} = g(MC_i, SNR_j), \forall MC_i \in \mathcal{MC}$;

Procedure: MC selection for user j

Compute the application layer bandwidth $b_j = \frac{r}{1-p_{i,j}}$, $\forall MC_i \in \mathcal{MC}$;

Compute the required channel rate $C_{i,j} = f(b_j, MC_i)$;

Compute the operational MC scheme $MC_j = \arg \min_{MC_i \in \mathcal{MC}} C_{i,j}$;

Output: best operational MC scheme for user j , MC_j .

The algorithm requires a single pass through all available MC schemes, in order to compute MC_j . If the set \mathcal{MC} is ordered in increasing order of complexity of the MC schemes, the algorithm will stop faster. The transformation f between the channel symbol rate and the application layer rate is a known function, depending on the modulation coding scheme and the channel code rate used in each MC scheme. The packet loss probability p_k can be computed for each MC scheme, based on the user SNR level. Function $g(\cdot)$ can either be modeled from wireless channel models (Figure 3) factoring

in the mobility and the effect of temporal channel variations, or can be empirically discovered through field tests [15],[16]. The user channel conditions SNR_j can be computed at the base station under a symmetric channel assumption, or can be observed either through the feedback control channel, or through separate feedback mechanisms.

B. Selection of $MC(k)$ for each video layer k

Algorithm 1 determines the user subset $S_{MC(k)}$ which can decode a given modulation and coding scheme $MC(k)$. Keeping this in mind, we next present our algorithm for assigning an appropriate MC scheme for the transmission of each video layer, given the total channel capacity constraint. Our algorithm starts by assigning the lowest MC scheme (supported by all users) to all video layers until the channel capacity is filled. Then we increase the MC scheme sequentially for each layer, and we assess the benefit of this action by looking at the trade-off between the extra video quality achieved by saving channel capacity using higher MC schemes, and the number of users that are able to decode the video information. Algorithm 2 presents the sketch of our proposal.

Algorithm 2 Selection of $MC(k)$ for each video layer k .

Input: Video layer rates $\rho_k, \forall k \leq L$, available MC schemes \mathcal{MC} , ordered user subsets $S_{MC(k)}$, channel capacity C ;

Procedure: MC selection \forall video layers.

Initialization: video layer $k = 1$, MC index $i = 1$;

Assign $MC(u) = i$ to all video layers $u \geq k$ up to channel capacity;

Update l based on the assignment of the $MC(u)$;

Compute $D(l) = \sum_{j \in S_i} D_j(l)$; $D_{opt} = D_l$;

while video layer $k \leq L$ **do**

while $i \leq M$ **do**

$i := i + 1$;

 Test $MC(u) = i$ for all video layers $u \geq k$ up to channel capacity;

 Update l based on the assignment of $MC(u)$;

 Compute $D_l = \sum_{u=1}^l \sum_{j \in S_{MC(u)}} D_j(u)$;

if $D_l \leq D_{opt}$ **then**

$D_{opt} = D_l$; $MC(u) = i$ for all video layers $u \geq k$;

else

 Break;

end if

end while

$k = k + 1$;

end while

Output: $l, MC(u) \forall u \leq l$.

The algorithm attempts to assign ever higher MC schemes to all video layers starting with the base layer in order to save channel capacity, and fit more application layer data within the channel constraint. The optimal point lies in the trade-off between the number of users that are able to decode the higher modulation scheme and increase their received video quality, and the number of users that are only able to decode the lower video layers transmitted at a lower MC scheme. The algorithm

starts by computing the appropriate MC scheme for the base layer, and then for the subsequent enhancement layers. Once the MC scheme of one video layer is fixed, the MC scheme for all remaining higher layers is considered to be at least as high. Since the distortion metric while computing the appropriate MC scheme for a give video layer k is a concave function (the number of users receiving an increased video quality is decreasing, while the number of users unable to decode is increasing), the algorithm will converge quickly to the final solution. The worst case complexity is $O(LM)$.

Depending on the system constraints, the algorithm can easily be adapted to fix the MC scheme for the video base layer such that all users in the system are able to decode, and obtain a minimum video quality. In this case, the algorithm is run only on the remaining video enhancement layers, taking into account the remaining channel capacity. Otherwise, the MC scheme of the base layer can be optimized as well, case in which an arbitrary penalty distortion value should be attributed to all users that cannot decode any of the video layers (users dropped).

C. Computing the AL-FEC for the transmitted video layers

Once we have established the appropriate MC scheme for each video layer, we need to establish the final transmission scenarios in which we protect the video information with application-layer FEC. To this end, we explore the trade-off between sending additional video layers, or better protecting the already scheduled layers, given the total available application rate. The MC scheme used for each video layer defines the channel capacity needed for transmitting the respective video layer. We associate to each MC_k chosen for layer transmission the packet loss probability p_k of the worst user assigned to the given subset $S_{MC(k)}$. Within this framework, we present a fast algorithm, which explores at each optimization step the trade-off between adding another video layer for transmission, or increasing the FEC protection of the previously scheduled layers. The decision of the algorithm is taken based on a utility function which assesses the decrease in overall video distortion of these two actions. The output of the algorithm consists in the number of video layers scheduled for transmission and their associated rate, plus the amount of additional FEC protection to be scheduled for each layer.¹

V. ARCHITECTURE

In this section we present the delivery architecture. Starting from the left side of Fig. 1, we consider a video encoded in H.264/SVC format at the streaming server. The quantization points, hence the encoding rates, and the number of quality layers, the GOP (Group of Picture) size and the frame rate of the video are selected a priori inside the SVC and FEC coding module. The additional packets provided by the FEC(n,s) redundancy inside the coding module increases the robustness of the video layers against channel erasures and is set high enough to allow the users to recover from losses in the worst

¹Due to lack of space, we do not formalize this algorithm in this paper, however, for its complete description, we refer the interested reader to our previous work [14].

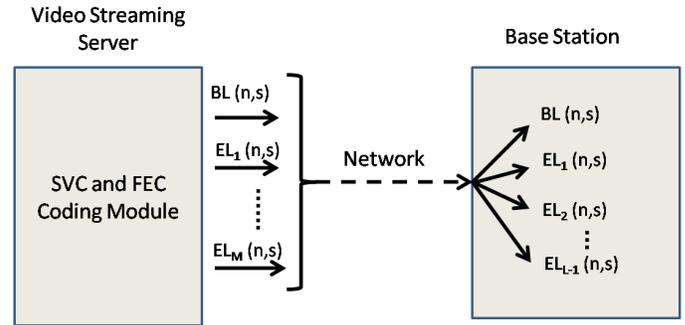


Fig. 1. Delivery architecture: from the video server to a base station.

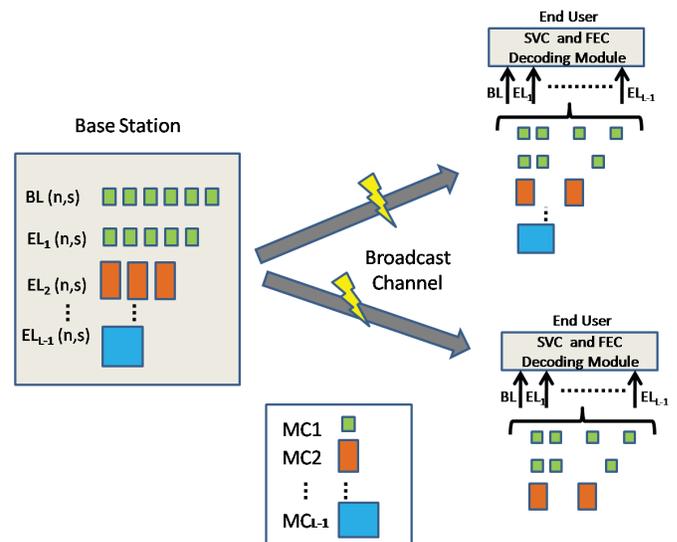


Fig. 2. Delivery architecture: from a base station to a set of end users.

channel conditions. At this stage the encoded video to be streamed is made of as many sub-streams as the number of quality video layers specified at the server, with additional FEC redundancy to be sent alongside with the original video packets. The video stream then is transmitted through the core network to the base station(s). A base station, when receiving the stream, extracts and buffers the sub-streams (quality video layers) with the corresponding FEC redundancy. Based on the broadcast channel feedback messages, i.e. packet erasure rates experienced by the users and available broadcast channel capacity, the base station performs the selection of the MC scheme and the amount of AL-FEC protection for each sub-stream that is scheduled for transmission (Fig. 2). Terminals with bad channel conditions (low SNR values) can successfully receive only packets sent with low order MCs, while users with good channel conditions can decode packets sent also with higher MCs. Low channel rates require more time to transmit the video packets, thus the channel capacity is eventually reached before the transmission of all the video layers. High channel rates allow to transmit the video layers and the FEC redundancy, requiring lower channel capacity. However, the drawback stands in the fact that users with bad channel condition cannot benefit from such high order

transmission schemes. Thus, the base station is in charge of selecting the most suitable combination of MCs to be used for sending the video layers so that to minimize the overall video distortion perceived by the set of end users participating in the video streaming session.

VI. SIMULATION RESULTS

A. Simulation Setup

As shown in Fig. 1, the *news_qcif* video sequence is encoded at the server in H.264/SVC format, using the JSVM software [13]. The quantization points are selected so that the base layer (BL) is sent with rate (at the application layer) of 80 kbps, the first and the second enhancement layer are sent with rate 100 (EL1) and 160 kbps (EL2), respectively. The amount of FEC redundancy to be provided at each video layer is computed on a per GOP-basis, where the GOP size is set to 16 and a GOP is composed of one I frame and 15 P frames. The frame rate is set to 30 frames per second. λ , the video sequence dependent constant to be used to compute the video distortion (MSE) in our experiments, is set to 354.8 [14]. Table I presents the set \mathcal{MC} of MCs that can be used to transmit the video layers at the base station, along with their channel capacity mapping function $f(\cdot)$. The function is expressed as a multiplication factor from channel capacity to application rate.

TABLE I
MCs AVAILABLE AT THE BASE STATION.

MC	Scheme	Mapping Function $f(\cdot)$ (b/symbol)
1	QPSK 1/8	0.25
2	QPSK 1/4	0.50
3	QPSK 1/2	1.00
4	16QAM 1/2	2.00
5	16QAM 1/3	2.67
6	64QAM 3/5	3.60

The set \mathcal{MC} in Table I is a representative subset picked from Figure 3, which presents the user PER for multiple MC schemes, at various SNR levels, in a LTE static scenario with an AWGN channel. In order to take user mobility and channel time variations into account, starting from the data in Figure 3, we derive error probabilities for the users for their operational MC scheme as discussed in Section III.

We map each video layer rate to the channel rate of the selected MCs in Table II. Hence, we present the amount of symbols per second required by each transmission scheme to send the base layer and the two enhancement layer into the channel; the conversion from application layer rate (kbps) to channel rate (ksymbols/s) for each MC is computed based on the bits-to-symbol factor in Table I.

B. Impact of user distribution and channel capacity

In this section we investigate the impact of the user distribution when varying the maximum channel capacity according to the video rates in (Table II). We consider a base station broadcasting the video stream to $N = 100$ users, with user distributions as reported in Table III. In our experiments we use 3 user distributions: *left*, i.e. most of the users can

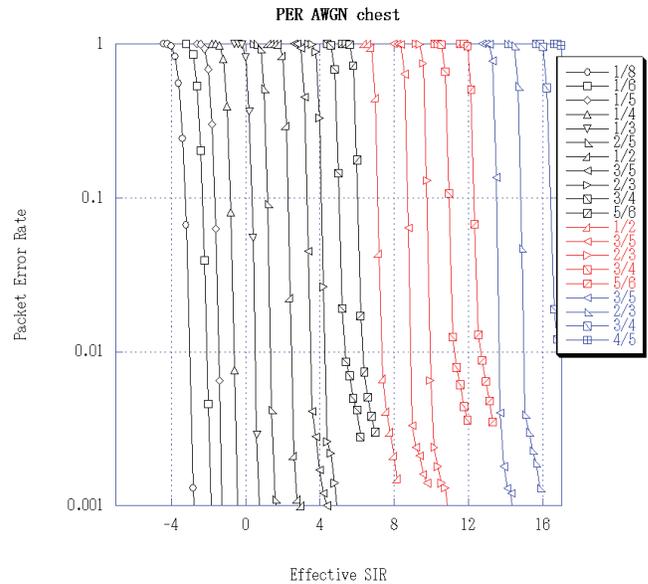


Fig. 3. User PER as a function of SNR for different MCs from simulations for an LTE system. Black curves are QPSK, red 16-QAM and blue 64-QAM.

TABLE II
APPLICATION AND CHANNEL RATES FOR THE H.264/SVC VIDEO SENT WITH MCs IN TABLE I.

	BL	EL1	EL2
AL rate (kbps)	80	100	160
MC1 (ksymb/s)	320	400	640
MC2 (ksymb/s)	160	200	320
MC3 (ksymb/s)	80	100	160
MC4 (ksymb/s)	40	50	80
MC5 (ksymb/s)	30	37.5	60
MC6 (ksymb/s)	22.2	27.7	44.4

successfully decode packets only if sent with low order MCs (e.g., most users are considered to be far from the base station, hence they experience a low SNR), *middle*, i.e. most of the users can decode packets only if sent with MCs in the middle of the set or lower, *right*, i.e. most of the users can decode packets sent with high order MCs (e.g., most users are close to the base station and experience good SNR levels). The cumulative distribution of users that benefit from the use of any MC is reported in Table III.

TABLE III
USER DISTRIBUTIONS.

MC	Left	Middle	Right
1	100	100	100
2	60	95	99
3	35	85	98
4	20	50	95
5	10	15	80
6	5	5	50

Next, we show the performance of our heuristic algorithms, for each user distribution. We explore the cases when the base station optimizes the transmission of the base layer at the expense of dropping some of the users (Fig. 4 (a),(b),(c)), and when it decides to serve all users with minimum video quality at least, e.g., fix the MC of the base layer to the

lowest MC in \mathcal{MC} (Fig. 4 (d),(e),(f)). In the plots we show the system average video distortion (MSE), i.e. the average of all individual distortions perceived by the clients, while varying the channel capacity. Given the channel rates in Table II, we span the channel capacity region of [100-800] ksymb/s when the base station might take the decision to drop users with weak channel conditions, and the region [500-1500] ksymb/s when all users are served. We compare our proposal to a worst-user based solution, when the base station prioritizes the worst user in the cell, and to an ideal case where each user can be served with a single unicast stream with the same capacity of the broadcast channel. Thus, the ideal case can be seen as a multiple unicast streaming session, where each single unicast channel has the same channel capacity of the broadcast channel, without being shared among users.

Starting with the analysis of Figure 4 (a),(b),(c), dropping users belonging to lower order MCs is clearly beneficial in terms of overall video distortion (normalized on the number of effective receivers), with striking gains when most of the users can decode higher MCs (Figure 4(c)). The decrease in distortion is smoother when users are distributed uniformly (Figure 4(b)) and when a large set of users receives at lower order MCs (Figure 4(a)). Each curve starts from a channel capacity that depends on the channel rates of each MC, as specified in Table II. E.g., 400 ksymb/s is the minimum channel capacity for serving all users so that to meet the requirement of sending the whole base layer at MC1 with additional FEC packets. However, it is important to consider both the impact in terms of MSE gained and the number of users dropped when the base station might decide to discard users, and where the trade-off strictly depends on the requirements of the streaming application. Thus, we report in the legend of the plots the number of users dropped by the base station, in order to emphasize that the action taken to decrease the overall distortion comes with a cost in terms of number of users discarded. A more detailed prospect is provided in Table IV, where the channel capacity is fixed at 400 ksymb/s, and the average video distortion achieved when dropping users is reported with the amount of users discarded. Moreover, to complete the picture, we report the FEC required per GOP for each combination of MCs.

We next discuss Figure 4(d),(e),(f), i.e. when all users are served by the base station. As expected, when most of the users can be served with low order MCs (Figure 4(d)), our algorithm sticks to the performance of the worst-user based solution. In fact, among all the possible combinations of MCs, the algorithm selects constantly MC1 for sending all the video layers, as shown in Table V. This is due to the fact that only a small set of users can benefit from the use of higher order MCs, thus the overall amount of distortion strictly depends on the majority, i.e. users successfully receiving with MC1. However, as shown in Figure 4(a), the base station can choose to drop a percentage of users so that to serve only the best users with a significantly improvement in terms of video distortion, but at the cost of discarding a high percentage of users. As soon as users are more uniformly distributed in terms of MCs (Figure 4(e)), by using our algorithm the overall video distortion drops towards the ideal case, sticking to it when

TABLE IV
BEST COMBINATIONS OF MCs, MSE AND AMOUNT OF FEC WHEN DROPPING USERS, CHANNEL CAPACITY 400 KSYMB/S.

Users Dropped	MCs BL:EL1:EL2	Distortion MSE	FEC BL:EL1:EL2
<i>Left Distribution</i>			
0	1:1:-	16.1509	2:0:0
40	2:2:2	8.5979	3:0:0
65	3:3:3	4.5145	1:0:0
80	4:4:4	4.5145	3:0:0
90	5:5:5	4.5145	3:1:0
95	6:6:6	4.5145	1:0:0
<i>Middle Distribution</i>			
0	1:3:3	14.0936	2:0:0
5	2:3:3	6.7342	3:0:0
15	3:3:3	4.5145	1:0:0
50	4:4:4	4.5145	3:0:0
85	5:5:5	4.5145	3:1:0
95	6:6:6	4.5145	1:0:0
<i>Right Distribution</i>			
0	1:4:4	11.7199	2:0:0
1	2:3:4	4.4583	3:0:0
2	3:3:3	4.5145	1:0:0
5	4:4:4	4.5145	3:0:0
20	5:5:5	4.5145	3:1:0
50	6:6:6	4.5145	1:0:0

users are mostly using higher order MCs (Figure 4(f)), from channel capacity 600 ksymb/s. The overall progression of our algorithm with respect to the selection of the MCs to use for sending the video layers for each user distribution, when all users are served, is also reported in Table V. With user distribution *left* the algorithm keeps using MC1 for streaming the video, while with distribution *middle* and *right* the algorithm keeps MC1 for sending the base layer, so that to serve the users with worst channel condition with the base video quality. Higher order MCs are selected for sending the enhancement layers, so that to provide higher video quality to the users with good channel condition. At a channel capacity above 1500 ksymb/s, according to our scenario, all the video layers can be sent with MC1, without being truncating, together with the additional FEC packets. Regarding the amount of FEC required when selecting the combinations of MCs in Table V, only the base layer is protected with 2 redundant packets per GOP (FEC(18,16)). This is also due to the relatively low packet error rate that we consider for the users, and is in line with our previous results reported in [14].

TABLE V
BEST COMBINATIONS OF MCs FOR EACH USER DISTRIBUTION, ALL USERS SERVED.

Channel (ksymb/s)	Left BL:EL1:EL2	Middle BL:EL1:EL2	Right BL:EL1:EL2
500	1:1:-	1:3:3	1:4:4
600	1:1:-	1:3:3	1:3:4
700	1:1:-	1:3:3	1:3:3
800	1:1:1	1:3:3	1:3:3
900	1:1:1	1:1:3	1:3:3
1000	1:1:1	1:1:3	1:3:3
1200	1:1:1	1:1:2	1:3:3
1500	1:1:1	1:1:1	1:1:1

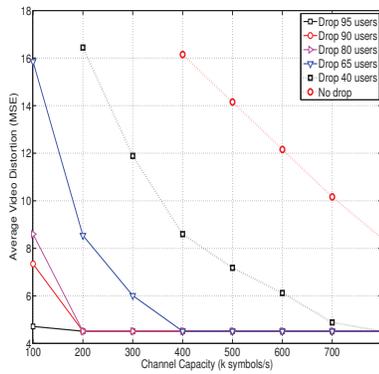
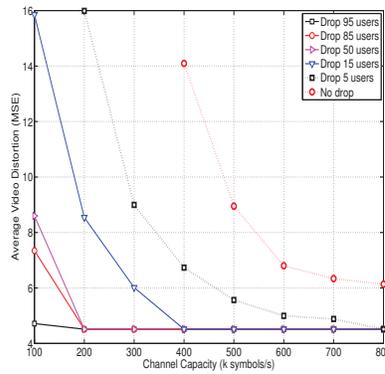
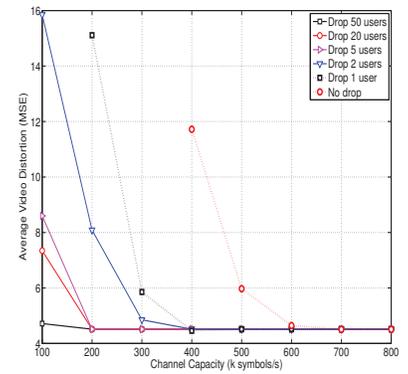
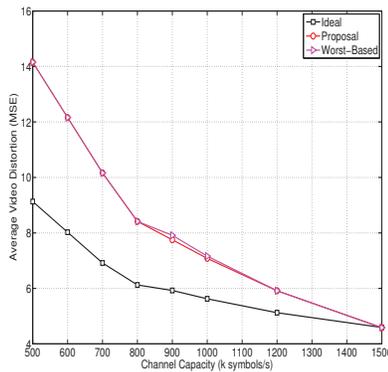
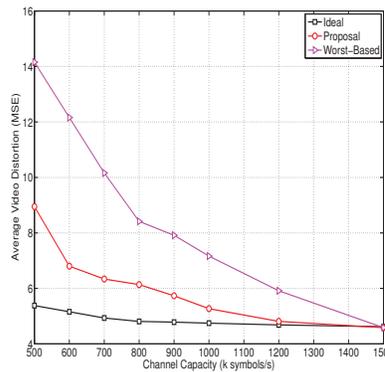
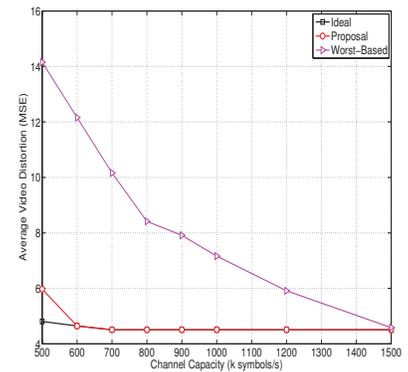
(a) User distribution *left*, dropping users.(b) User distribution *middle*, dropping users.(c) User distribution *right*, dropping users.(d) User distribution *left*, performance of our algorithm vs. the ideal and the worst-user based solution.(e) User distribution *middle*, performance of our algorithm vs. the ideal and the worst-user based solution.(f) User distribution *right*, performance of our algorithm vs. the ideal and the worst-user based solution.

Fig. 4. Impact of user distribution and channel capacity. Average video distortion (normalized on the number of effective receivers) vs. channel capacity.

C. Comparison with AVC-based solutions

Finally, we compare the performance of our SVC-based algorithm with a AVC-based streaming solution, when the base station broadcasts different independent video streams for users with different channel qualities. Three copies of the same video stream are built based on H.264/AVC from the same QCIF video sequence considered above, so that the encoding rates are 80, 240 and 340 kbps, respectively. We set the channel capacity to 500 ksymb/s, and we let the base station select which MC to use to broadcast the 3 bitstreams so that to meet the channel constraint and to provide the minimum overall video distortion. Making sure that all the users at least can get the first bitstream (80 kbps), the remaining channel capacity can be allocated to transmit in parallel a second bitstream (240 kbps) and eventually a third bitstream. This means that the base station will store multiple copies of the same video stream (additional storage and capacity cost) and each user will play only the best stream that fits its channel condition. In Table VI we report three bitstream-switching solutions based on H.264/AVC and one solution based on H.264/SVC for the mentioned setting, varying the user distribution.

The bitstreams (AVC) or quality layers (SVC) are sent with MCs reported in the second column of the table as first, second and third video stream (for the SVC solution, EL1 and EL2 are sent with MCs 1 for user distribution *left*, 3 for the *middle*

and 4 for *right* one). From the table we draw the conclusion that sending one scalable stream provides in average a better video quality, due to the more efficient usage of the channel capacity. Sending multiple AVC streams reduces the number of solutions that can be used to meet the channel constraint, which moreover leads to further increasing the number of users receiving a stream with lower video quality than expected. Similar results were obtained for different settings, and are omitted here due to space constraints.

TABLE VI
COMPARISON WITH AVC, CHANNEL CAPACITY 500 KSYMB/S, AVERAGE MSE.

	MCs 1st:2nd:3rd	MSE <i>left</i>	MSE <i>middle</i>	MSE <i>right</i>
AVC	1:4:-	15.81	12.52	7.58
AVC	1:5:6	16.78	16.23	7.96
AVC	1:6:6	17.33	17.33	11.25
SVC	1:(1,3,4):-(-,3,4)	14.10	8.95	6.12

VII. DISCUSSION ON PRACTICAL ISSUES

In the previous sections we detail our proposal for a wireless broadcast system for scalable video streaming. Our evaluation concentrates on emphasizing the theoretical characteristics of this system, and simulate its behavior. Our conclusions

open the discussion for more practical issues which must be evaluated in the case of a real deployment of our proposal:

- Coverage area of the broadcast system: When choosing an appropriate MC scheme for the video broadcast, we implicitly shape the maximum size of the cell in which the broadcast can be successfully received. Field information on the placement and coverage area of the base stations of a cellular operator offering such a service could lead to a more informed analysis of the trade-offs involved.
- Number of users served/number of base stations involved: The broadcast decisions taken by our algorithm influence the total number of users that can decode the broadcast stream. For higher MC schemes, only the users with the better SNR levels, hence closer to the base station, are able to decode the video information. Hence, for coverage preservation, a cellular operator would have to increase the density of base stations offering this service. Insights on the implementation and operational costs of such a deployment would bring an additional dimension to our proposal.
- Realistic user density models and SNR levels: In our simulations we present results for three different user spatial distribution scenarios, when a majority of users have good SNR levels, e.g., are close to the base station, have medium channel quality, or poor channel quality, e.g., are further away from the base station. For each of these cases, our mechanisms adapt and choose an appropriate transmission strategy, such that the overall system is optimized. Real data regarding user spatial distribution and user SNR levels would increase the value of our experiments from the quantitative point of view.
- Different video scalability dimensions: In our analysis we only consider the SNR scalability feature of the H.264/SVC codec. However, the spatial and temporal scalability features of the same codec could bring different dimensions and more flexibility to our framework. E.g., broadcasting a video stream encoded at different spatial resolutions would address the problem of the heterogeneity of mobile devices in terms of screen size and resolution capabilities.
- H.264/AVC vs. SVC: Finally, a full understanding of a practical broadcast system would imply a thorough investigation of the differences between solutions based on SVC, as the one presented here, and simpler bitstream switching mechanisms based on the better known AVC standard. Full comparisons would imply codec complexity measures, coding performance differences, availability of low-complexity decoders in the mobile hand-sets, and backward compatibility issues.

VIII. CONCLUSIONS

We address the problem of wireless broadcast of a scalable video bitstream in a cellular system. We formulate a cross-layer framework in which we attempt to optimize the number of streamed video layers, the modulation and coding scheme, and the application layer FEC used for each layer. Our

framework takes into account the total channel capacity and the individual channel conditions observed by the users. Our optimization metric is represented by the total video distortion observed by the system. In order to make the analysis of the proposed problem tractable, we decouple the influencing factors, and we provide heuristic algorithms for each of the individual sub-problems. We test our proposed method for different channel scenarios and user distributions, and we compare our scalable solution with baseline algorithms and with a traditional broadcast system based on AVC encoding. We observe the better results obtained by our proposed algorithms, due to the extra flexibility offered by the scalable application paradigm and due to the cross-layer approach. Finally we explore further issues and problems that influence a practical deployment of our solution by a mobile operator.

REFERENCES

- [1] 3GPP, "Introduction of the Multimedia Broadcast/Multicast Service (MBMS) in the Radio Access Network (RAN)," Mar. 2006.
- [2] "3GPP Specification Detail: MBMS." [Online]. Available: <http://www.3gpp.org/ftp/Specs/html-info/26346.htm>
- [3] H. Schwarz, D. Marpe, and T. Wiegand, "Overview of the scalable video coding extension of H.264/AVC," *IEEE Trans. Circuits Syst. Video Technology*, vol. 17, no. 9, pp. 560–576, 2003.
- [4] C. W. Chan, N. Bambos, S. Wee, and J. Apostolopoulos, "Wireless Video Broadcasting to Diverse Users," in *IEEE ICC 2008*, Beijing, China, May 2008, pp. 377–382.
- [5] A. Shokrollahi, "Raptor codes," *IEEE/ACM Transactions on Networking (TON)*, vol. 14, pp. 2551–2567, 2006.
- [6] "Dvb: the global standard for digital television." [Online]. Available: <http://www.dvb.org/>
- [7] A. Naghdinezhad, M. R. Hashemi, and O. Fatemi, "A novel adaptive unequal error protection method for scalable video over wireless networks," in *Proceedings of IEEE International Symposium on Consumer Electronics*, 2007.
- [8] H. Mansour, P. Nasiopoulos, and V. Krishnamurthy, "Joint media-channel aware unequal error protection for wireless scalable video streaming," in *Proceedings of IEEE ICASSP*, 2008.
- [9] M. Van Der Schaar, S. Krishnamachari, C. Sungyun, and X. Xiaofeng, "Adaptive cross-layer protection strategies for robust scalable video transmission over 802.11 WLANs," *IEEE Journal on Selected Areas in Communications*, vol. 21, pp. 1752–1763, 2003.
- [10] J. Villalon, P. Cuenca, L. Orozxo-Barbosa, Y. Seok, and T. Turletti, "Cross-Layer Architecture for Adaptive Video Multicast Streaming Over Multirate Wireless LANs," *IEEE Journal on Selected Areas in Communications*, vol. 25, no. 4, May 2007.
- [11] R. Cohen, G. Grebla, and L. Katzir, "Cross-Layer Hybrid FEC/ARQ Reliable Multicast with Adaptive Modulation and Coding Broadband Wireless Networks," in *IEEE Infocom 2009*, Rio de Janeiro, Brazil, Apr. 2009.
- [12] D. Jurca, W. Kellerer, E. Steinbach, S. Kahn, S. Thakolsri, and P. Frossard, "Joint network and rate allocation for video streaming over multiple wireless networks," in *IEEE International Symposium on Multimedia (ISM'07)*, 2007.
- [13] "Joint scalable video model - reference software." [Online]. Available: http://ip.hhi.de/imagecom_G1/savce/downloads/SVC-Reference-Software.htm
- [14] D. Munaretto, D. Jurca, and J. Widmer, "A Fast Rate-Adaptation Algorithm for Robust Wireless Scalable Streaming Applications," in *IEEE WiMob 2009*, Marrakech, Morocco, Oct. 2009.
- [15] J. Karlsson and M. Riback, "Initial Field Performance Measurements of LTE," Ericsson Review, Mar. 2008.
- [16] Nokia Siemens Networks, "LTE Performance for Initial Deployments," White Paper, 2009.