# Cross–layer cross–domain adaptation of mobile video services

Jose Oscar Fajardo*, Ianire Taboada, Fidel Liberal

NQaS Research Group, Department of Electronics and Telecommunications, University of the Basque Country (UPV/EHU), ETSI Bilbao, 48013 Bilbao, Bizkaia, Spain

## Abstract

This paper deals with the analysis of user perceived visual quality for mobile multimedia services. H.264/AVC is selected for low resolution video encoding, and a mobile UMTS network is considered for the access to services. First, from subjective tests results, the combined impact of different service– and radio network–level parameters is inferred. As a result, different cross–layer adaptation alternatives are proposed to maximize the perceived quality level under different service conditions. Afterwards, the backhaul segment is considered taking into account possible congestion effects after the aggregation of multi–user media stream. The proposed dynamic adaptation process determines the best combination of media bitrates in order to optimise the overall quality, by making use of information related to network performance both at the radio and backhaul segments.

## 1. Introduction

Multimedia applications have become increasingly popular over Internet. Furthermore, the wide–spreading of enhanced mobile Internet access and the continuous evolution of multimedia encoding techniques allow provisioning video services over mobile data connections at acceptable quality levels. For instance, the mobile version of the YouTube video sharing platform currently offers H.264/AVC–encoded video clips, which can be accessed from a mobile handset via RTSP with a standard video client. Typically, two versions of the video clips are available. On one hand, the normal version is based on the low spatial resolution (SR) of QCIF (176x144) and low encoding bitrate (about 80 kbps). On the other one, the High Quality (HQ) versions offer a higher SR at QVGA –a.k.a. square pixel SIF– (320x240) at higher encoding bitrates (about 250 kbps).

The choice of the most suitable version depends on different parameters. The screen resolution of the mobile device imposes a first requirement. Nowadays, typical screen resolutions are QCIF, QVGA and VGA. As well, the performance variability associated to mobile data services must be considered. Taking into account a typical 3G Universal Mobile Telecommunications System (UMTS) service, the 384kbps bearer supports the requirements of HQ versions under perfect reception conditions. However, the variable radio conditions may introduce degradations in the transmission, and the normal version could be preferred.

As widely studied, an adaptive approach could provide the best quality level all over the service time by introducing real–time modifications in the service provision configuration. Particularly, combined service– and network–level actions are of paramount interest in order to dynamically perform cross–layer adaptations in the service provision.

In [1] we addressed this topic, focusing on the provision of H.264/AVC–based video services over 3G UMTS data connections. Specifically, this work presents a consolidated performance study by taking into account the combined effects of the specific characteristics of the UMTS Terrestrial Radio Access Network (UTRAN) on one hand, and the specific characteristics of the H.264 encoding on the other

*Corresponding author. Email: joseoscar.fajardo@ehu.es

one. In order to perform the most suitable adaptation actions, the proposed decision making process is driven by the expected visual quality level as perceived by end users. Thus, instead of using classical objective visual quality assessment metrics such as PSNR, VQM and SSIM [2], subjective video quality perception approach is considered in this work, which is essential for network providers in order to achieve an acceptable user satisfaction.

As a step forward, in this paper we take into consideration the possible effects of the backhaul segment in the mobile infrastructure. In this case, the decision making process is not only driven by the service configuration and the experienced performance in the radio link, but it also takes into account the most suitable adaptation actions in order to overcome possible congestion situations in the mobile backhaul. This new functionality is stated in this paper as a cross–domain feature, assuming that the radio and wired segments of a mobile network are considered as two different domains.

## 1.1. Background

The analysis of user perceived visual quality has been subject of many studies in recent years. With regard to H.264/AVC at mobile resolutions, several studies present their subjective tests results for service contexts similar to the considered in this paper.

From the subjective tests results presented in [3, 4], the H.264/AVC codec can be assumed as the best performing codec for QCIF mobile video services. As well, the most suitable audio–video (A/V) bitrate ratio is inferred as a function of the content type (CT). In [5] authors study the combined effects of the source bitrate (SBR) and frame rate (FR) for mobile resolutions. The main outcome is a regression–based expression which relates the Mean Opinion Score (MOS) to the SBR and FR values for different CT. The FR is considered in the range of 5 fps to 15 fps, while the target SBR varies from 24 kbps to 105 kbps. In this case, all the subjective test sequences are presented at QCIF resolution screen, which justifies the low values considered for SBR. This range of SBR values is considered insufficient for the aims of this paper, where higher SR values are considered. As well, in these studies only the encoding parameters are considered, assuming ideal transmission conditions.

Concerning higher SR, e.g. results in [6] present quality evaluations of CIF (352x288) video sequences in terms of blocking, blurring and flickering artifacts. The SBR is considered from 100 kbps to 300 kbps for this SR at 25 fps. However, only the encoding effects are considered for the quality analysis as well.

Both H.264/AVC CIF video sequences and transmission effects are considered in the analyses of perceived

video quality presented in [7, 8]. In both cases, a 2–state Markov model is used to implement the bursty packet losses. However, in both cases the loss model is implemented at IP level. In our case, the bursty error pattern is implemented at UMTS Radio Link Control (RLC) level following the results shown in [9], where the error pattern is obtained from live UMTS network traces. This feature entails a better emulation of the combined effects of the service–level settings and the experienced UMTS performance conditions.

As a result, although all the reviewed studies are close to the case study, none of them covers all the objectives proposed in this paper and hence new subjective tests have been performed.

## 1.2. Scope and objectives

The main objective of the study is to in–depth study the combined impact of different service– and network–level parameters into the experienced visual quality from a consolidated standpoint, in order to analyze the most suitable cross–layer adaptations that maximize the Quality of Experience (QoE). In this sense, the main contributions of this paper are twofold:

(i) A thorough analysis of the visual quality which could be expected in a mobile multimedia service, as currently being provided in real–world services.

(ii) A dynamic cross–layer cross–domain adaptation mechanism, aimed at maximizing the QoE under variable conditions in the mobile network, including both the radio and the backhaul segments.

One of the main novelties of the paper is the mobile multimedia service awareness. All the subjective tests have been performed resembling actual service conditions, including mobile–oriented media encoding and presentation to end users. As well, the used UMTS reception patterns related to the radio interface are based on real–world measurements under different mobility scenarios. Finally, we include in the analysis possible congestion situations in the mobile backhaul, taking into account the bitrate ranges related to mobile video streams.

The remainder of this paper is structured as follows. Section 2 presents the methodology followed for performing the subjective tests. Section 3 focuses on the analysis of perceived visual quality, in function of the encoding settings. From subjective tests results, the evolution of expected MOS with the SBR is inferred for each CT and SR. This way, the most suitable SR can be identified for the achievable SBR and per CT. Likewise, Section 4 focuses on the analysis of perceived visual quality in function of the UMTS performance. From subjective tests results, the evolution of expected

MOS with the experienced UMTS conditions is inferred for each CT. This way, the most suitable SBR can be identified for the experienced Block Error Rate (BLER) and per CT. Section 5 illustrates the considered cross–layer adaptation capabilities, which provides improvement in terms of QoE based on the knowledge of the service and radio network states. Finally, Section 6 introduces the effects of the mobile backhaul in the decision making process. Section 7 gathers the main conclusions to this paper.

## 2. Subjective testing methodology

The experimental design for the subjective video tests is mainly based on ITU recommendations and tutorials [10–12]. Different aspects are taken into account for planning the subjective tests for this kind of multimedia applications.

Concerning the viewing conditions, video sequences are presented to end users resembling a mobile UMTS service in order to enhance the accuracy of results [13]. All video sequences are displayed in a mobile handset and users are asked to hold it in their hands. The device used in the tests is a Nokia N95-8Gb, which provides a 320x240 screen resolution. The tests are carried out with RealPlayer for s60, which uses image re–scaling for presenting QCIF sequences at full screen.

The video sequences are a priori generated including both encoding and transmission effects, and they are stored in the mobile device for its presentation to subjects. So, an appropriate device and displaying format is used to achieve the proposed objectives in a fixed environment conditions for all the subjects. In order to capture the combined effects of the specific encoding and transmission techniques, long duration video sequences (about 2 minutes) have been used instead of the typical short (about 10s) reference video sequences.

Taking as reference the proposed test structure in [10–12], before starting with test sessions, written instructions were shown to subjects and a training phase was done in which some videos are presented and evaluated, without taking into account these results. Then this is followed by several test sessions. In each test session different types of test scenes are shown. These are presented in random order and some implicit replications are included to check coherence [10]. Due to fatigue issues, break periods between sessions are introduced [12]. For carrying out the different experiments proposed in this study, each subject participates in the experiments for three different days.

In order to reproduce viewing conditions that are as close as possible to real–world contexts, Single Stimulus (SS) tests are used and the audio track is also included in the multimedia stream. The evaluations are based on Absolute Category Rating (ACR). Thus, after each test

**Table 1.** Complexity–motion metrics and snapshot of videos used.

| CT | Complexity–motion metrics | Snapshot |
|----|---------------------------|----------|
| LM | C=0.46, M=0.28 | |
| MM | C=0.49, M=0.42 | |
| HM | C=0.62, M=0.96 | |



sequence presentation the subjects are asked to evaluate the quality of the presented sequence in the MOS scale of 1 to 5.

## 3. Service–level parameters

All the video sequences have been encoded with the H.264/AVC Joint Model Reference Software. Three different CT have been considered: low–motion (LM), medium–motion (MM), and high–motion (HM) video sequences. The considered content is described as follows:

(i) LM sequence is made up of news video clip with two people, including change of planes and a commercial in middle of the sequence.

(ii) MM sequence is a typical TV series scene, featuring different people and including change of planes.

(iii) HM sequence is taken from a basketball top 10 best plays video sequence, with changes of planes from field to face planes.

In order to evaluate spatio-temporal activity level of video sequences the parameters proposed in [14] are used. This approach consists on associating an scene complexity (C) and level of motion (M) value to each video sequence based on the average bits per frame and the average quantization parameter (QP) for I and P frames respectively. Table 1 shows the associated complexity-motion metrics for each CT as well as a snapshot of the video signals used.

**Table 2.** Encoding settings for subjective tests.

| CT | SBR (kbps) | SR | FR (fps) |
|----|-----------|-----|---------|
| LM | {80, 130, 200} | 320x240 | 10 |
|    | {48, 88, 128} | 176x144 | 10 |
| MM | {80, 130, 200} | 320x240 | 12.5 |
|    | {48, 88, 128} | 176x144 | 12.5 |
| HM | {80, 200, 256} | 320x240 | 15 |
|    | {48, 88, 128} | 176x144 | 15 |

**Table 3.** Subjective tests results for QVGA.

| SBR (kbps) | LM | MM | HM |
|-----------|------|------|------|
| 80 | 3.49 | 2.90 | 1.98 |
| 130 | 4.26 | 3.92 | – |
| 200 | 4.53 | 4.49 | 4.26 |
| 256 | – | – | 4.53 |

Instead of analyzing the visual quality as a function of the combined setting of FR and SBR, in this study only the SBR is modified. The FR values are set up in function of the content dynamics: 10, 12.5 and 15 fps are established for LM, MM and HM respectively. Table 2 summarizes the considered values.

The H.264/AVC encoder is set up to its Baseline Profile at level 1.2. The frame structure is IPPP, with a Group of Pictures (GOP) size of 10 seconds. Thus, each 10 seconds the video sequence is refreshed with a new I frame.

## 3.1. Encoding quality

First, subjective tests were performed for evaluating the impact of SBR into the encoding quality of QVGA sequences. A total of 20 people evaluated each sequence, for a total of 180 tests. The obtained results are presented in Figure 1, which gathers the box plots of the subjective testing results for the whole set of considered CT–SBR conditions. For each combination of CT and SBR considered, the median, Q1 and Q3 percentiles and minimum/maximum values are illustrated. As can be observed, there is no too much variability in the quality scores provided among users. As well, those values which are considered outliers in the data sample are individually plotted. Outlier values are determined by the Chauvenet's Criterion and they are not taken into account in the computation of the MOS. Table 3 shows the average quality values derived from the subjective tests for QVGA resolutions.

Similarly, new tests were performed for the QCIF versions of the same videos. In order to resemble actual service conditions, QCIF video sequences are presented at full screen in the mobile handset with RealPlayer, so image scaling is required. A total of 20 people evaluated
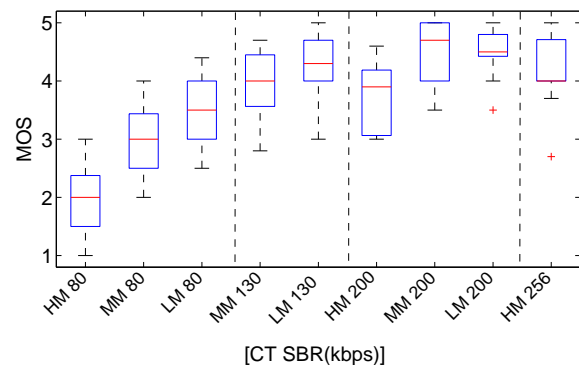


**Figure 1.** Boxplot of subjective testing results for encoding of QVGA sequences.

**Table 4.** Subjective tests results for QCIF.

| SBR (kbps) | LM | MM | HM |
|-----------|------|------|------|
| 48 | 3.49 | 2.75 | 1.83 |
| 88 | 3.52 | 3.46 | 2.65 |
| 128 | 3.66 | 3.60 | 2.91 |

**Table 5.** Experimental coefficients for the fitting function.

| | LM QVGA | MM QVGA | HM QVGA | LM QCIF | MM QCIF | HM QCIF |
|---|--------|--------|--------|--------|--------|--------|
| *a* | 1.141 | 1.739 | 2.274 | 0.159 | 0.608 | 1.124 |
| *b* | –1.442 | –4.665 | –7.954 | 2.858 | 0.556 | –2.48 |

each sequence, for a total of 180 tests. In this case only the averaged MOS values are provided in Table 4.

The QCIF versions, being half the size than the QVGA sequences, require less SBR to achieve an acceptable visual quality. However, the increase of SBR does not entail a proportional increase in the perceived quality for different SR values. As described in [8], the video encoding process may reach the visual quality threshold and above this threshold a higher SBR is not captured by the human visual system.

## 3.2. Impact on service–level adaptation

In order to evaluate the evolution of both alternatives, we applied fitting techniques to the subjective results and obtained an approximation of the relationship between MOS and SBR values for both SR. The fitting function is given in Equation (1) as proposed in [8].

$$MOS = a \cdot log(SBR) + b \qquad (1)$$

Parameters *a* and *b* are related to the activity level of the content and spatial resolution, and the obtained values are illustrated in Table 5.
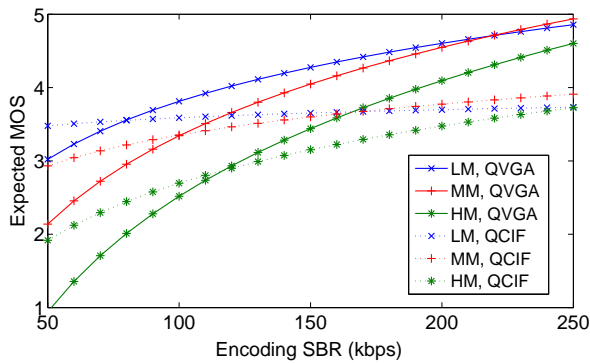
**Figure 2.** Evolution of expected MOS vs. SBR per CT and SR.

By using these results, Figure 2 presents the expected evolution of the subjective visual quality in terms of MOS for each pair of considered CT–SR values. For each CT, the two resulting curves cross at a specific point ($SBR_{th}$) and two differentiated regions are identified. For SBR values lower than $SBR_{th}$, it is preferable to switch to a lower SR in order to control the impact of the limited SBR. The $SBR_{th}$ value is higher for more dynamic sequences. For the three CT considered in this study, we find that this threshold is around 80, 100 and 115 kbps for LM, MM and HM respectively.

## 4. UMTS radio access

The scope of this paper regards to the provision of mobile video streaming services over a typical wide–area 3G UMTS data service, as defined in 3GPP TR 25.993 [15] for the *Interactive or Background/UL : 64 DL : 384 kbps/PSRAB*. This kind of bearer service provides a maximum downlink bitrate (DLBR) of 384 kbps. A detailed description of the considered service provision and the impact of the UTRAN is given in [16].

From the total bitrate, the final amount available for the video encoding process is reduced by several factors. First of all, the RTP/UDP/IP packetization introduces an overhead in the data transmission. Second, we must consider the effect of the audio stream within the multimedia transmission. The analysis of the audio quality and the integral quality (as shown e.g. in [17, 18]) is out of the scope of this paper. Yet, the impact of its transmission is simulated by adding a 64 kbps stream to the considered video stream. Finally, part of the DLBR is used by the RLC AM functions for the local recovery of lost MAC PDUs. In this sense, the performance of the video streaming service highly depends on the radio error pattern.

As cited in Sect. 1.1, one of the novelties of this paper is that the UMTS error model is implemented at RLC level from real–world measurement results, instead of using a typical 2–state Markov model for simulating the IP–level loss events. For the simulation of different

network conditions, the implemented error model is a 2–state Markov model with variable Block Error Rate (BLER) values.

Two characteristics are adopted from the results presented in [9]:

(i) For mobile users, the radio errors can be grouped at Transmission Time Interval (TTI) level.

(ii) The Mean Burst Length (MBL) of erroneous TTIs can be approximated to 1.75.

The error model, as well as the simulation methodology, is further detailed in [19]. The RLC–level error model, in combination with the application–level settings, determines the performance of the service. The RLC errors may derive to additional delays if the RLC is able to recover the lost PDU, or to video frame losses otherwise.

Taking into account the relevance of the different frame types, a content–aware scheduling is implemented in a similar way to the concepts proposed in [20] for 3G UMTS, in [21] for HSDPA and in [22] for LTE. In this case, the priority of different RLC retransmissions is modified in order to implement an enhanced protection for I frames. This way, we prevent severe degradations in the initial picture of each GOP and its propagation all over the 10 seconds period.

### 4.1. 3G UMTS transmission quality

Considering the mentioned characteristics for 3G UMTS mobile multimedia services, different combinations of service and network conditions are simulated. All the simulations are run with OPNET Modeler, where both the specific UMTS error pattern and a H.264/AVC RTP trace injector have been implemented. QVGA video traces corresponding to the three CT have been used in the simulations with several SBR values. 80, 130 and 200 kbps have been considered for LM and MM video sequences, while an additional 256 kbps version has been used for HM traces. All the traces have been transmitted several times from a video server to the mobile endpoint, traversing the UMTS network segment. The downlink BLER value in the radio part is set up from 1% to 30% at 5% steps. At 30% of BLER, all the sequences experience high degradations except for the LM 80 kbps versions.

From the whole set of results obtained, a mapping between different BLER values and experienced IP Loss Ratio (IPLR) patterns is established. For those points with negligible IPLR values (under 0.1%), the service performance is considered accurate. Similarly, high IPLR values (over 5%) indicate unaffordable service conditions.

The rest of intermediate points are considered for subjective evaluation of the visual quality perceived by

**Table 6.** SBR and BLER settings for subjective tests.

| CT | SBR (kbps) | BLER(%) |
|----|-----------|---------|
| LM | 130 | 15, 20, 25 |
|    | 200 | 5, 10 |
| MM | 130 | 20, 25 |
|    | 200 | 5, 10 |
| HM | 256 | 5 |
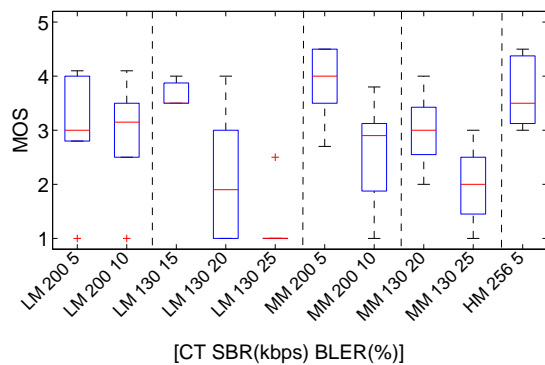


**Figure 3.** Boxplot of subjective testing results for UMTS transmission of QVGA sequences.



**Figure 4.** Evolution of expected MOS vs. BLER per CT and SBR for QVGA sequences: (a) low motion, (b) medium motion, (c) high motion.

users. Table 6 presents all the service– and network–level conditions that have been included in this set of subjective tests. The application–level performance (in terms of IPLR) is not only determined by the experienced BLER values, but also by the different traffic patterns associated to the different CT. Thus, different conditions require subjective evaluation in function of the CT. For the aims of this study, a total of 114 subjective tests were performed by 20 people.

Figure 3 illustrates the results obtained from subjective tests for the considered encoding/transmission conditions. In this plot, the same blox plot parameters as in Figure 1 are depicted. As can be observed, especially two conditions (namely LM–130–20 and MM–200–10) show a high variability in the quality scores provided by users. In those cases, the associated MOS are 2.15 and 2.58 respectively.

## 4.2. Impact on cross–layer service adaptation

The extensive results obtained both from the analysis of simulation results and from subjective tests allow us depicting a mapping between the application and transmission conditions to the expected video quality in terms of MOS, as shown in Figure 4 for each considered CT.

From the behavior illustrated in Figure 4, each combination of CT, SBR and BLER determines the expected visual quality value. Thus, for a specific BLER
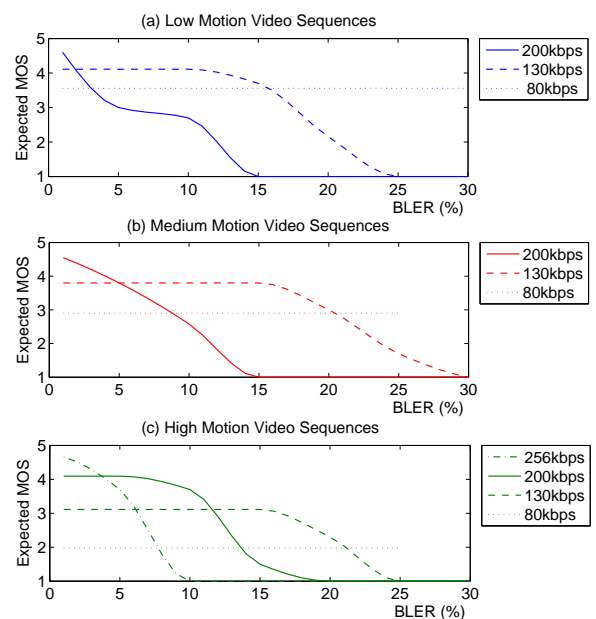
value, the video streaming session can be set up to the new SBR value that maximizes the expected MOS. For the aims of this paper, this adaptation is considered as a standalone decision making process.

The BLER value is not modified (e.g. by modifying the power control functions of the link layer) so the impact of the adaptation of a video stream on the performance experienced by other users in the same cell is limited in this case.

If power or rate control mechanisms are considered in multi–user environments, where several users are contending for the access to limited cell resources at the same time, the optimization problem can be studied as shown in [23, 24].

Another alternative in the standalone management of mobile services is the capability of modifying the Radio Bearer settings in function of the experienced conditions. If for the experienced BLER condition none of the highest considered SBR values is suitable, the UMTS Radio Bearer can be switched to a DLBR value of 128 kbps, which exhibits a better resilience to noise and interference at the same transmission power levels. Thus, low BLER values can be expected in order to guarantee no further impairments than the encoding process itself.

At the same time, the multimedia streams are switched to a lower SBR in order to cope with the new DLBR requirements. As a result, this approach entails a combined encoding/network service adaptation. As in

the previous case, the impact on other users is limited and each user can be adapted by itself.

## 5. Evaluation of cross–layer adaptations

Based on the results obtained from the subjective tests, different adaptation approaches can be evaluated. Four alternative approaches are considered, each option including a new adaptation capability from zero to three configurable parameters.

(i) *No Adaptation (NA)*. In this case, no adaptations actions are considered. Thus, the achieved MOS values are those corresponding to the initial service conditions. In order to offer accurate quality levels, for LM and MM sequences the initial SBR is set up to 200 kbps, while HM sequences are configured to 256 kbps.

(ii) *Network–Aware Bitrate Adaptation (NABA)*. This alternative considers the adaptation of SBR to the value that maximizes the expected MOS for the specific CT and experienced BLER, driven by the estimation curves shown in Figure 4. For the purposes of this paper, the allowed SBR values are 256, 200, 130 and 80 kbps.

(iii) *Network–Aware Bitrate and Spatial Adaptation (NABSA)*. This approach takes into account the two configurable service–level parameters considered in this paper: SBR and SR. Thus, based on the resulting SBR adaptation, the SR is also switched from QVGA to QCIF if the optimal SBR is lower than the $SBR_{th}$ for the specific CT.

(iv) *Network–Aware Cross–Layer Adaptation (NACLA)*. In this case, a cross–layer adaptation is adopted by taking into account the three parameters considered. Besides the application–level adaptation (combined SBR/SR), the DLBR can be decreased to overcome severe degradation conditions in the UMTS data connection.

Figure 5 shows the obtained results for each dynamic adaptation approach, taking into account the different CT considered.

For LM sequences, only two curves are differentiated. If no dynamic adaptation is applied, the mobile video service gets completely degraded around the 15% of BLER. However, the service can be kept at suitable quality levels (MOS=3.5) just with bitrate adaptation under the analysed UMTS conditions. The LM sequences exhibit no degradations at BLER=30%, and both SR versions provide similar quality levels at SBR = 80 kbps. Hence, the three dynamic adaptation approaches offer analogous behaviours at the whole range of the BLER conditions studied. Thus, just SBR adaptation provides the maximum achievable quality under different conditions in this case.
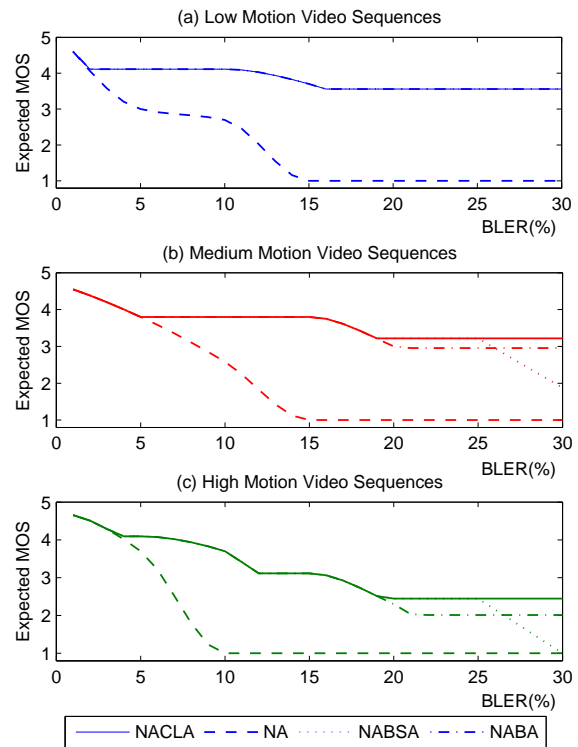


**Figure 5.** Quality achieved by the different adaptation approaches for (a) low motion, (b) medium motion, (c) high motion sequences.

As can be observed, this is not case for MM and HM sequences. On one hand, the $SBR_{th}$ is located above 80 kbps in both cases, and thus the SR should be switched to QCIF when SBR is set up to 80 kbps. On the other hand, the QCIF versions exhibit frame losses when the experienced BLER is above 25%. As a result, the three dynamic adaptation approaches provide different quality levels under severe UMTS degradations and different type of adaptations are required in order to maximize the QoE.

From the analysis of results, a dynamic network–aware cross–layer adaptation mechanism can be proposed, as illustrated in Figure 6. Following the depicted logic, a mobile endpoint could be capable of launching the required adaptation procedures to keep the mobile video service in the maximum achievable quality level along the service time.

## 6. Flow aggregation at the mobile backhaul

In the previous sections, we have analysed the behaviour of each mobile user in a standalone fashion. Based on the performance experienced in the radio access network, the system is capable of adapting the service– and network–level parameters regardless the rest of the users.
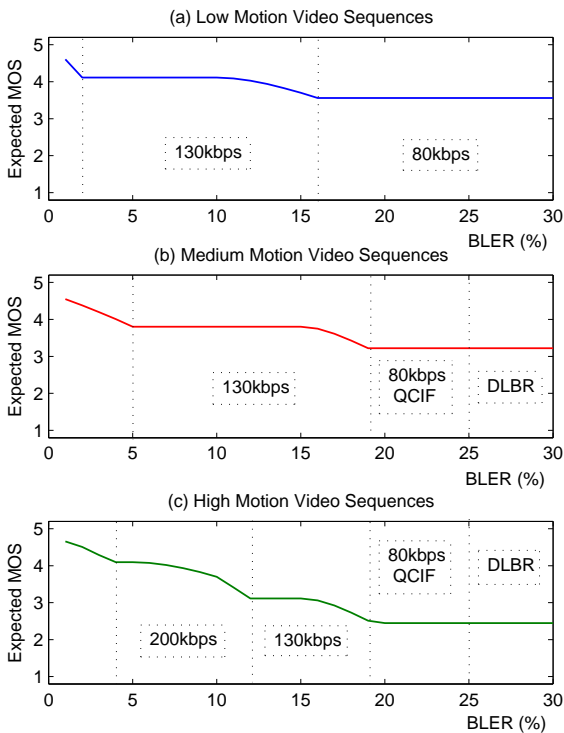
**Figure 6.** Dynamic configuration and maximum achievable quality for (a) low motion, (b) medium motion, (c) high motion sequences.

In this section we introduce a new variable by taking into account the backhaul link, where the traffic of all users is aggregated. In such context, the analysis of the backhaul capacity is critical for the study of the QoE. In the simplest formulation, traffic losses will appear whenever the aggregated traffic results in a higher bitrate than the actual link capacity. Thus, we introduce a new relevant system parameter: the traffic loss ratio (TLR):

$$TLR = \frac{\sum_{n=1}^{N} SBR_n}{R} \quad (2)$$

being N the total number of mobile multimedia users in the system and R the capacity of the aggregation link.

To overcome this problem, the system could decide to implement some kind of admission control based on users' required bitrate. In a more advanced strategy, the system could propose service–level adaptations in order to decrease the bitrates of the different sessions to suitable levels. The key issue is to identify which users are more suitable for adaptation and how much bitrate can be decreased, in order to maximise the general offered QoE.

Therefore, the decision making process becomes a QoE-driven multi-user resource allocation problem,

where the QoE is also subject to the service and radio network conditions specific to each user.

Instead of the typical MOS averaging between all users, the proposed system is aimed at maximising the overall MOS while keeping the worst–case users at the maximum possible quality level. Thus, the proposed objective function is expressed as:

$$F(\tilde{x}) = \frac{1}{N} \sum_{n=1}^{N} (\frac{1}{MOS_n(\tilde{x})})^2 \quad (3)$$

where the parameter tuple $\tilde{x} \in \tilde{\chi}$, being $\tilde{\chi}$ the set of all accepted value tuples in accordance to the definitions of the parameter set {CT, SBR, SR, BLER, DLBR, TLR}.

The decision making process can be expressed as:

$$\tilde{x}_{opt} = \arg\min_{\tilde{x} \in \tilde{\chi}} F(\tilde{x}) \quad (4)$$

This is, $\tilde{x}_{opt}$ is the parameter tuple which maximizes the proposed objective function and therefore the system performance.

In the case study here presented, the parameters CT and BLER are non–variable. From results in Section 5, the tupla CT–BLER determines the maximum value for SBR, and consequently the optimal values for SR and DLBR parameters. As a result, the parameters to study are the individual SBR values and the experienced TLR, being TLR a function of the sum of individual SBR values. In the context of this paper, the decision making process is aimed at keeping TLR=0. As a result, the objective parameter becomes the vector of individual SBR values.

We can state the problem as:
Minimise

$$\frac{1}{N} \sum_{n=1}^{N} (\frac{1}{MOS(SBR_n)})^2 \quad (5)$$

subject to

$$\sum_{n=1}^{N} (SBR_n) \leq R \quad (6)$$

### 6.1. Adaptation decision making algorithm

The optimization problem formulated in Equations (5) and (6) is solved by means of a genetic algorithm [25]. This way, in order to efficiently solve the proposed adaptation problem in real–time, this approach allows us to obtain optimal or suboptimal solutions in lower execution time than exact optimization methods.

Figure 7 summarizes the implemented genetic algorithm based adaptation logic.

Each individual represents a possible system state that can be reached by performing SBR adaptation over the existing multimedia sessions. The initial population is generated at the moment of the invocation of the adaptation.
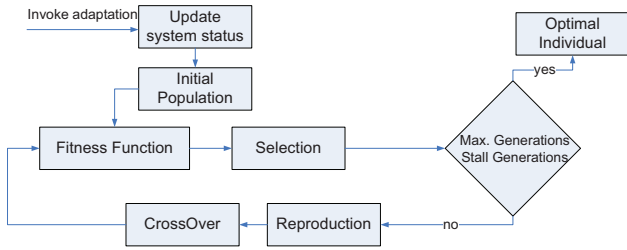
**Figure 7.** Genetic algorithm based adaptation logic.

**Table 7.** Initial service and radio access conditions.

| Session ID | CT | SBR (kbps) | BLER(%) |
|---|---|---|---|
| 1,2,3 | LM | 200 | 0, 10, 25 |
| 4,5,6 | LM | 130 | 0, 10, 25 |
| 7,8,9 | MM | 200 | 0, 10, 25 |
| 10,11,12 | MM | 130 | 0, 10, 25 |
| 13,14 | HM | 256 | 0, 10 |
| 15,16,17 | HM | 200 | 0, 10, 25 |
| 18,19,20 | HM | 130 | 0, 10, 25 |

The fitness function provides an assessment of how accurate the evaluated solution is. The implemented fitness function receives two parameters: the SBR for each media flow and all the information required for the computation of the expected MOS of each flow. The output of the fitness function is determined by all the MOS values associated to each multimedia session, and the fitness score is computed based on the average and minimum MOS values obtain for the whole set of multimedia sessions.

In the selection module two outcomes are possible. If the algorithm reaches the Maximum number of established generations (5) or when the best score has not changed over the last Stall generations (2), the best scoring individual from the current generation is returned. Otherwise, the algorithm goes on the reproduction phase which determines how the next generations of the individuals are generated. On this case study, the three best scoring individuals survive in each generation, and the rest of individuals are generated by the cross–over function between the selected parents in a per flow basis.

## 6.2. Evaluation of cross–domain adaptations

In order to analyze the performance of the proposed cross–domain management system, we propose an scenario where 20 mobile users experience different service and network conditions, as shown in Table 7.

As well, we evaluate the performance of the system in different conditions of the backhaul segment by introducing several background traffic loads. Taking

**Table 8.** Background load conditions and backhaul traffic loss.

| Background load | Total load | Traffic loss |
|---|---|---|
| 0 Mbps | 3.482 Mbps | 0% |
| 0.5 Mbps | 3.982 Mbps | 3.97% |
| 1 Mbps | 4.482 Mbps | 17.01% |
| 1.5 Mbps | 4.982 Mbps | 30.07% |

into account SBR values in Table 7, the traffic load due to media flows becomes 3.482 Mbps. We assume that the backhaul capacity is overdimensioned 10% over the aggregation of initial service bitrates (thus, $R = 3.83 Mbps$). Following Equation (2), we compute the TLR associated to different background traffic loads as shown in Table 8.

Under these conditions and the adaptation procedure stated in Section 6.1, we compare the performance of three alternative approaches:

(i) *No adaptation*, thus the additional backhaul load makes the TLR increase until the overall quality is totally degraded for the multimedia sessions.

(ii) *NACLA*, the network–aware cross–layer adaptation approach as described in Section 5.

(iii) The new *Cross–Layer Cross-Domain Adaptation (CLCDA)*, which implements the multi–user adaptation actions as presented in Section 6 based on the performance of both radio and backhaul segments.

Figure 8 gathers the MOS values experienced by every multimedia session with the different approaches, each subplot representing a different backhaul load condition. As well, Figure 9 shows the resulting average and minimum MOS statistics.

As can be observed, the two adaptation approaches proposed highly outperform the no adaptation case. In the first case, there are no backhaul losses and thus both adaptation approaches exhibit similar performance. As the additional backhaul load increases, it is proven that cross–domain adaptation entails a better performance. While the average values are quite similar for both optimisation processes, the cross–domain approach always provides a better performance with regard to the minimum value, which means that the system will offer enhanced protection to severe degradations.

## 7. Conclusions

This paper deals with possible dynamic adaptations for H.264/AVC based video services over 3G UMTS mobile connections. In order to get an optimal configuration, different cross–layer adaptations are proposed and evaluated. Thus, both service– and
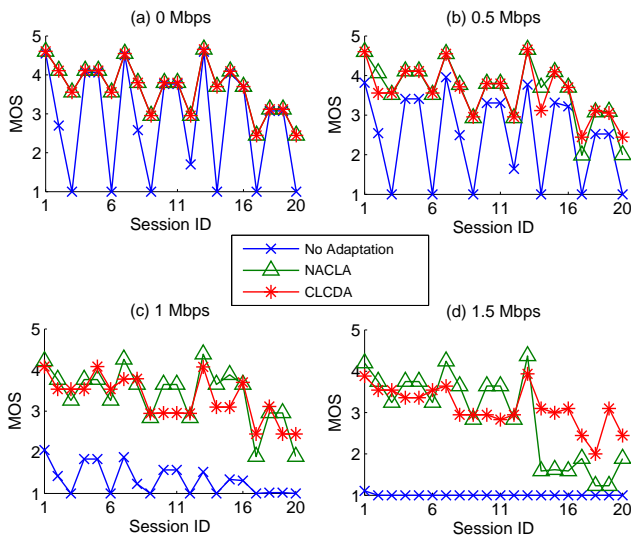
**Figure 8.** Evolution of MOS scores for different backhaul background traffic loads: (a) 0 Mbps, (b) 0.5 Mbps, (c) 1 Mbps, (d) 1.5 Mbps.
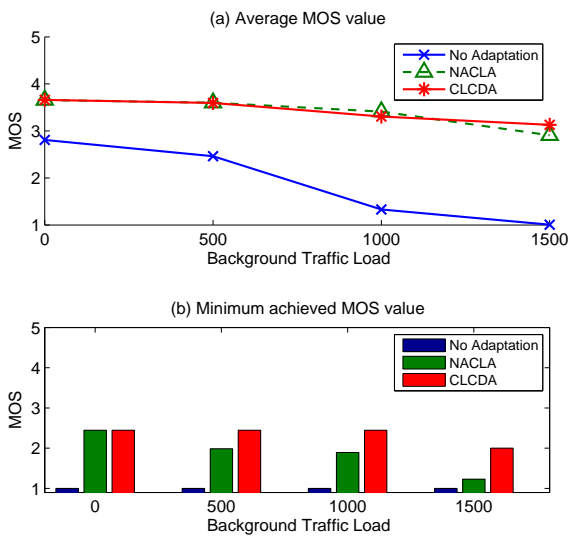


**Figure 9.** QoE statistics for each adaptation scheme: (a) average MOS, (b) minimum MOS.

network–level adaptations are considered to cooperate in a coordinated way in order to maximize the QoE.

Since the proposed adaptations are driven by the expected QoE, specific subjective tests have been performed to cope with the specific service context. Thus, both specific H.264/AVC parameters and specific error models for commercial UMTS networks have been used. The outcomes of the subjective testing phase are later considered as inputs for the dynamic service adaptation logic.

Different adaptation approaches have been evaluated, considering different number of variable parameters:

(i) On the one hand, we can state the enhancements of the three–parameter adaptation approach, based on the combination of encoding bitrate, spatial resolution and UMTS bearer bitrate. As described in Section 5, the considered adaptations do not have an impact on other users, so it can be implemented in a per–user basis.

(ii) On the other hand, once the service provision is optimised in terms of service and radio network, the multi–user problem is analysed. The aggregation of traffic at the mobile backhaul may introduce further degradations in the media streams. Thus, we propose an enhanced automated decision process which takes into account the performance of the different segments of the network (i.e., radio and backhaul segments) and maximises the overall QoE of the whole set of users. For this aim, the decision making process, driven by a genetic algorithm, considers the effect of bitrate reduction for each considered user, in terms of QoE, and determines the most suitable adaptation actions.

In order to obtain optimal results, the overall decision making process must be aware of different service– and network–level parameters. Thus, it could be deployed as a central management element which dynamically drives the provision of mobile media services and reacts to variations in the network performance. Regarding just the radio-aware adaptation, the adaptation logic could be implemented at the mobile endpoint if the mobile device requires access to low–level parameters. Currently, several Android–based commercial mobile devices are capable of providing BLER statistics in real–time. As a work in progress, we have developed the software to capture these statistics from the chipset in order to make them available for the applications.

## References

[1] Fajardo, J.O., Taboada, I. and Liberal, F. (2011) Cross-Layer Adaptation of H.264/AVC over 3G UMTS Mobile Video Services. In *Proceedings of the 3rd International ICST Conference on Mobile Lightweight Wireless Systems, MOBILIGHT 2011* (ICST). doi:10.1007/978-3-642-29479-2_8.

[2] Wang Y.(2006) Survey of objective video quality measurements. *EMC Corporation Hopkinton, MA* **1748**.

[3] Jumisko–Pyykkö, S. and Häkkinen, J. (2005) Evaluation of subjective video quality of mobile devices. In *Proceedings of the 13th Annual ACM International Conference on Multimedia, MULTIMEDIA '05* (New York: ACM), 535–538. doi:10.1145/1101149.1101270.

[4] Khan, A., Sun, L. and Ifeachor, E. (2011) QoE Prediction Model and its Application in Video Quality Adaptation over UMTS Networks . *Multimedia IEEE Transactions on*, **99**: 431–442.

[5] Ries, M., Nemethova, O. and Rupp, M. (2008) Video Quality Estimation for Mobile H.264/AVC Video Streaming. *Journal of Communications* **3**(1): 41–50 doi:10.4304/jcm.3.1.41-50.

[6] Ho, H.H., Wolff, T., Salatino, M., Foley, J.M., Mitra, S.K., Yamada, T. and Harasaki, H. (2007) An investigation on the subjective quality of H. 264 compressed/decompressed videos. In *Proceedings of the Third International Workshop on Video Processing and Quality Metrics for Consumer Electronics, VPQM 2007.*

[7] De Simone, F., Naccari, M., Tagliasacchi, M., Dufaux, F.D. Tubaro, S. and Ebrahimi, T. (2009) Subjective assessment of H.264/AVC video sequences transmitted over a noisy channel. In *Proceedings IEEE International Workshop on Quality of Multimedia Experience, QoMEX 2009* (IEEE), 204–209. doi:10.1109/QOMEX.2009.5246952.

[8] Koumaras, H., Lin, C.H., Shieh, C.K. and Kourtis, A. (2010) A Framework for End–to–End Video Quality Prediction of MPEG Video. *J. Visual Communication and Image Representation* **21**(2): 139–154. doi:10.1016/j.jvcir.2009.07.005.

[9] Karner, W., Nemethova, O. Svoboda, P. and Rupp, M. (2007) Link Error Analysis and Modeling for Video Streaming Cross–Layer Design in Mobile Communication Networks. *ETRI Journal* **29**(5): 569 – 595. doi:10.4218/etrij.07.0107.0102.

[10] ITU-T (2003) Rec. J.148: Requirements for an objective perceptual multimedia quality model.

[11] ITU-T (2001) P.862 Perceptual evaluation of speech quality (PESQ), an objective method for end–to–end speech quality assessment of narrowband telephone networks and speech codecs.

[12] Beerends, J.G. and Stemerdink, J.A.: A perceptual audio quality measure based on a psychoacoustic sound representation (1992) *Audio Engineering Society* **40**: 963–974.

[13] Jumisko–Pyykkö, S. and Hannuksela, M.M (2008) Does Context Matter in Quality Evaluation of Mobile Television? In: *Proceedings of the 10th international conference on Human computer interaction with mobile devices and services, MobileHCI '08* (New York: ACM), 63–72. doi:10.1145/1409240.1409248.

[14] Hu, J.and Wildfeuer, H.(2009) Use of content complexity factors in video over IP quality monitoring. In: *International Workshop on Quality of Multimedia Experience – QoMEX* IEEE: 216–221. doi: 10.1109/QOMEX.2009.5246950.

[15] 3GPP (2008) TR 25.993: Typical examples of Radio Access Bearers (RABs) and Radio Bearers (RBs) supported by Universal Terrestrial Radio Access (UTRA).

[16] Fajardo, J.O., Liberal, F. and Bilbao, N. (2009) Impact of the video slice size on the visual quality for H.264 over 3G UMTS services. In: *Proceedings of the Sixth International Conference on Broadband Communications, Networks, and Systems, BROADNETS 2009*, 1–8. doi:10.4108/ICST.BROADNETS2009.7022.

[17] Winkler, S. and Faller, C. (2006) Perceived audiovisual quality of low-bitrate multimedia content. *IEEE Trans. Multimedia* **8**(5): 973–980. doi:10.1109/TMM.2006.879871.

[18] Fajardo, J.O., Taboada, I and Liberal, F. (2011) QoE-driven and network-aware adaptation capabilities in mobile multimedia applications. *Multimedia Tools and Applications* , 1–22. doi:10.1007/s11042-011-0825-y.

[19] Khan, A., Sun, L., Ifeachor, E., Fajardo, J.O. and Liberal, F. (2010) Video Quality Prediction Model for H.264 Video over UMTS Networks and their Application in Mobile Video Streaming. In: *Proceedings of the 2010 IEEE International Conference on Communications, ICC* (IEEE), 1–5. doi:10.1109/ICC.2010.5502195.

[20] Benayoune, S., Achir, N., Boussetta, K., and Chen, K. (2008) Content–aware ARQ for H.264 Streaming in UTRAN. In: *Proceedings of the IEEE Wireless Communication and Networking Conference, WCNC 2008*, 1956–1961. doi:10.1109/WCNC.2008.348.

[21] Superiori, L., Wrulich, M., Svoboda, P. and Rupp, M. (2009) Cross–Layer Optimization of Video Services over HSDPA Networks. In: *First International ICST Conference on Mobile Lightweight Wireless Systems, MOBILIGHT 2009* (ICST), 135-146. doi:10.1007/978-3-642-03819-8-14.

[22] Li, L. and Song, J. (2012) Research of Scheduling for Video Communication in OFDM Systems. In: *Video Engineering.* doi:CNKI:SUN:DSSS.0.2012-05-020.

[23] Khan, S., Thakolsri, S., Steinbach, E. and Kellerer, W. (2008) QoE–based Cross-layer Optimization for Multiuser Wireless Systems. In: *Proceedings of the 18th ITC Specialist Seminar on Quality of Experience*, 63–72.

[24] Xie, L., Hu, C., Wu, W. and Shi, Z. (2011) QoE-aware Power Allocation Algorithm in Multiuser OFDM Systems. In: *Mobile Ad-hoc and Sensor Networks (MSN), 2011 Seventh International Conference on IEEE*, 418–422.

[25] Affenzeller, M.and Winkler, S.(2009) Genetic algorithms and genetic programming: modern concepts and practical applications. *Chapman & Hall/CRC* **6**.