

Impact of the video slice size on the visual quality for H.264 over 3G UMTS services

Jose Oscar Fajardo, Fidel Liberal, Nagore Bilbao
Department of electronics and Telecommunications
University of the Basque Country (UPV/EHU)
Bilbao, Spain
{joseoscar.fajardo, fidel.liberal, nagore.bilbao}@ehu.es

Abstract—The provision of video streaming services over mobile data accesses is emerging as a promising service for the current and future Internet. This paper focuses on the study of H.264-based video contents over a typical 3G UMTS connectivity. The performance of the service is finally determined by the combined effects of the encoding features and the radio access characteristics. Therefore, based on the UMTS channel modeling, this paper tries to find out the optimal encoding configuration that provides the best service performance, in terms of Quality of Experience. In addition to other well-known encoding parameters, such as the source target bitrate, frame rate or resolution, this paper studies the impact of the maximum video slice size as a critical parameter for the performance of the considered service.

H.264; mobile video streaming; 3G UMTS; video slice size

I. INTRODUCTION

The convergence of multimedia applications and the mobile world is becoming reality. On one hand, the possibility to access the Internet through mobile connections is currently in a quite mature state. Among all the technical alternatives, the most prominent solution for providing wide-area data connections is the Universal Mobile Telecommunications System (UMTS). Although the QoS-enabled management is not commercially extended today, the quality achievable by a typical best effort data connection seems to be fair enough to provide some kind of multimedia services, such as Video On Demand (VOD) or mobile IPTV.

At the same time, the technical advances concerning the video encoding schemes have suitably evolved in different aspects. First, the currently developed codecs are able to offer good visual quality levels at low data rates, especially if we consider mobile resolutions. In addition, these kind of encoding schemes have been developed taking into account the specific error-prone characteristics of the transmission systems. Therefore, they incorporate different mechanisms for protecting the data, which results on a better visual quality when losses occur during the transmission.

As a result, many research studies have been recently devoted to the study of the performance of different video transmission schemes over different mobile data networks. In particular, this paper focuses on the study of the performance of a H.264-based video streaming service over a 3G UMTS

mobile access. This combination has been selected as an exponent of the currently deployed service platforms.

Thus, this paper tries to offer 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 one. For that purpose, we review the most relevant requirements imposed by the 3GPP for this type of mobile video services, and we take also into consideration the latest research results in the area that can be found in the scientific bibliography.

As a particular contribution of this paper, we present the results of using different maximum video slice sizes during the encoding process. Since different alternative configurations can be found in the bibliography, we have carried out several simulations in order to study the effect of different sizes on the visual quality finally perceived by the end users.

The reminder of this paper is structured as follows. In Section II we describe the case study and address the technical specifications imposed by the requirements of the considered type of service: first, we overview the main features of the H.264 codec, as proposed by the 3GPP for mobile environments.; afterwards, we briefly describe the characteristics of the considered UMTS access; finally we study the different requirements that this type of UMTS transmission impose over the video encoding process. In Section III we introduce the concept of Quality of Experience (QoE) for video streaming services, in order to identify the most relevant technical parameters that impact the performance of the video streaming. Based on this discussion, Section IV and Section V present the proposed encoding set and describe the simulations that have been carried out to analyze the effects of the UMTS transmission on the performance of the service. Specifically, we focus the study on the analysis of the impact of the maximum video slice size into the visual quality. Section IV focuses on the impact of the UMTS transport characteristics in an error-free environment, addressing the effects of the limited resources and the slotted access to the radio medium. Section V extends this study by taking into account the error-prone behavior of the UTRAN, and studies the different visual outcomes in function of the maximum video slice size allowed in the splitting of the pictures. Finally, Section VI provides the conclusions to the work presented in this paper as well as some future research directions.

The research leading to these results has received funding from the European Community's Seventh Framework Programme FP7/2007-2013 under grant agreement n° 214751/ /ICT-ADAMANTIUM/

II. MOBILE VIDEO STREAMING: CONTEXT OF STUDY

This case study focuses on the provision of video streaming services over a UMTS mobile access. Considering a typical 3G UMTS access network, without any HSPA enhancements enabled, the maximum downlink achievable bitrate in a wide-area context is 384kbps. As said before, our interest focuses on the study of multimedia services over a typical general-purpose mobile connection. Thus, from now on we will consider a UMTS service provided over an “Interactive or Background / UL:64 DL:384 kbps / PS RAB”, as defined in 3GPP TR 25.993 [1]. This type of Radio Access Bearer (RAB) is usually offered with no specific QoS requirements to be fulfilled in the provided Internet access.

With regard to the video encoding process, we select the H.264 codec as the most prominent alternative for low bandwidth connections. From an overview of the 3GPP recommendations, we find out that several codification schemes are considered for different service scenarios. For instance, the 3GPP TS 26.235 [2] lays the foundations for UMTS packet-switched conversational multimedia applications, such as a video-enabled peer-to-peer VoIP session. Concerning the functional requirements of the video codification, the specification states that a conversational client should support H.264 (AVC) Baseline Profile at Level 1.1. For video streaming services, such as VOD or unicast IPTV services, the H.264 (AVC) support is extended up to the Level 1.2 in its Baseline Profile in the 3GPP TS 26.234 [3]. Finally, the 3GPP TS 26.346 [4] focuses on the services to be provisioned over MBMS-enabled UMTS access networks, such as multicast IPTV, and states the H.264 Baseline Profile Level 1.2 as the only video codec recommended for MBMS.

A. H.264 Video Streaming

In this section, we review the most significant technical implications related to the cited encoding scheme. On one hand, the use of the H.264 Baseline Profile (BP) is mandatory. This profile is primarily designed for applications with limited computing resources such as mobile devices, which enforces several limitations as depicted in Table I. [5]

As well, the requirement of specifically making use of a level configuration up to the Level 1.2 determines that the video sequence will behave following a set of characteristics. Table II shows, for the considered levels, the maximum number of macroblocks per second (MB/sec), the maximum number of macroblocks (MBs) per picture and the maximum bitrate at the Video Coding Layer (VCL) allowed for each configuration. Moreover, the last column shows different examples of spatial resolutions (SR) at different frame rates (FR) for each level. The last value, namely the maximum stored frames, identifies the maximum number of frames that can be saved in the device memory at the same time and, therefore, the maximum number of frames that can be referred for the decodification process. Considering low capacity devices such as mobile phones, this number of frames has to be kept low in order not to saturate the resources of the terminal.

For the purposes of this study, we will take into consideration a typical mobile phone, and thus a QCIF video resolution (176x144).

TABLE I. CHARACTERISTICS OF THE H.264 BASELINE PROFILE

Encoding Features	Value
I and P Slices	Yes
B Slices	No
SI and SP Slices	No
Multiple Reference Frames	Yes
Entropy Coding (CAVLC, CABAC)	CAVLC only
Flexible Macroblock Ordering (FMO)	Yes
Arbitrary Slice Ordering (ASO)	Yes
Redundant Slices (RS)	Yes
Data Partitioning	No
Croma Format	4:2:0 only
Sample Depth	8 bit only

TABLE II. EXAMPLE CONFIGURATIONS FOR THE H.264 LEVELS IN BP

Level	Encoding characteristics			
	Max MBs/sec	Max frame size	Max VCL bitrate for BP	SP @ FR (max stored frames)
1.1	3000	396 MBs	192 kbps	176x144@30.3 (9) 320x240@10.0 (3)
1.2	6000	396 MBs	384 kbps	320x240@20.0 (7) 352x288@15.2 (6)

The H.264 encoding process differentiates two layers: the Video Coding Layer (VCL) and the Network Adaptation Layer (NAL). Once the video is encoded accordingly, each video frame is sent throughout the network as an RTP packet, adding extra RTP, UDP and IP headers to each video Network Adaptation Layer Unit (NALU). The impact of the packet-switched video streaming is analyzed in 3GPP TR 26.937. [6]

The video encoded data stream may be encapsulated following different approaches:

- One video slice of a target size per RTP packet.
- One Group of Blocks (GOB) containing a row of macroblocks per RTP packet.
- One video frame per RTP packet.

The selection of the most suitable scheme will depend on the expected performance. For instance, including a whole video frame per RTP packet results on a better usage of the available resources than slicing the pictures, since the packet overhead exhibits lower ratios. Figure 1 illustrates the overhead introduced at different video NALU sizes.

However, bigger IP packets get a higher packet loss probability at the same radio Bit Error Rate (BER). Thus, it can be stated that bigger packet sizes would introduce higher degradation in the performance of the streaming service in the same network conditions. This effect has to be carefully taken into account, especially when the multimedia stream traverses error-prone segments such as the considered UMTS access.

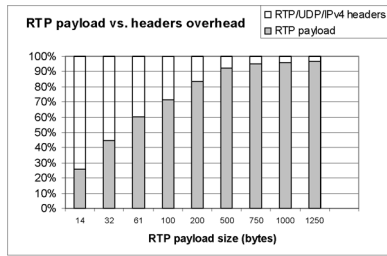


Figure 1. Bandwidth overhead for different video packet sizes [6]

B. UMTS Characteristics

The specific characteristics of the Packet Switched (PS) data transmission over UMTS networks have a great impact over the performance of the considered video streaming service. The UMTS PS data service is built over a protocol stack which includes different layers. The Packet Data Convergence Protocol (PDCP) service is configured to transport upper layer SDUs from the UE to the GGSN. The Radio Link Control (RLC) supports the data transmission service in the UTRAN. The MAC layer offers its data transfer service as Logical Channels, including control and traffic channels. Finally, the Physical Channel offers its data transfer service as Transport Channels, which can be common or dedicated channels. This service is available to the upper layers at configured slotted time intervals, namely Transmission Time Intervals (TTI), and the upper layers usually send the data included in the SDUs in a number of Transport Blocks (TB) per TTI. The specific settings to be used in a connection are determined by the Transport Format Set (TFS), which is negotiated within the establishment of the Transport Channel.

Table III provides an overview of the most significant characteristics associated to the considered “Interactive or Background / UL:64 DL:384 kbps / PS RAB”. Each of the cited parameters will have an impact on the performance of the different multimedia services. For instance, the RLC layer is able to send at most 12* 320 bits of payload at exactly each 10 ms. This means that bursty traffic behaviors result on additional delays, which may be unbearable for multimedia services. A similar conclusion can be reached for the recovery functions provided by the RLC AM. This discussion is extended in [7] for the case of VoIP services.

TABLE III. CHARACTERISTICS OF THE H.264 BASELINE PROFILE

Encoding Features	Value
Max. bitrate at RLC level	384 kbps
RLC PDU size	320 bits
RLC Mode	Acknowledged Mode (AM)
Allowed Transport Format Set (TFS)	Six possible TFs 0-1-2-4-8-12 TB/TBS
Transport Block (TB) size	336 bits
Transmission Time Interval (TTI)	10 ms
Transmission Channel (TrCH) type	Dedicated Channel (DCH)

C. H.264 over UMTS

Taking into account the described mobile video streaming approach, we now analyze some characteristics related to the encoding process, and the effect of different encoding parameters in a UMTS environment as the considered.

When using a codec such as H.264/AVC in an error-prone and resource-limited system such as UMTS, the accurate use of the different error resiliency features is crucial. Therefore, the selection of the error resiliency method is crucial. A review of the scientific bibliography yields different performance results related to different alternative schemes, namely: slicing structure of the frames, flexible macroblock ordering (FMO) with different ordering schemes, data partitioning, intra placement of IDR and I frames, use of redundant slices and intra-block refreshing.

Among all the error resilient features offered by the H.264/AVC codec, some of them are not recommended or are banned by the 3GPP specifications for mobile video streaming. For instance, data partitioning is a common technique when H.264 video transmission is studied. [8] This feature provides the ability to separate syntax elements by its relevance, allowing an unequal error protection (UEP) and other types of improvement for error/loss robustness. Unfortunately, this feature is not allowed in the Baseline Profile.

In addition, the use of IDR and I frames results in a quite bitrate-greedy behavior for limited-resource environments, and introduces extra delay in the communication. I frames contain all the necessary data to refresh the video sequence with a completely new picture. As we will see, the size of this type of frames may increase the instantaneous required bitrate. Thus, I frames require several TTIs to transmit the whole information, which can make that the frame is finally received with a significant delay. Therefore, the frequency of I frames refreshing should be configured to low values, or even no I frame refreshing can be adopted. As a result, jointly to the fact that no B or SP/SI slices are allowed in the Baseline Profile, the frame structure proposed for the case study is the IPPP structure. Likewise, the use of redundant slices may be restricted due to the limited bitrate in mobile environments.

The splitting of video frames into slices proves to be one of the most useful features. The reception of a corrupted slice limits the visual artifact to that frame slice (and possibly to the next frames slices referring it in the GOP). In order to keep this isolation of errors, a 1:1:1 encapsulation scheme can be used: each video slice makes up one single NALU, which at the same time is transmitted in one RTP packet. As a result, the size of the slice determines the size of RTP packet and consequently has an impact on the video loss rate.

The most used slicing methods are: FMO, fixed number of MBs per slice and fixed number of bytes per slice. The use of FMO may introduce additional complexity in the codification process. Moreover, in [9] authors study the efficiency of different error resiliency features in 3G environments, obtaining that no FMO and extra intra block refreshing achieves the best error correction vs. bitrate result.

In the literature concerning H.264 streaming over UMTS access networks, different studies propose a fixed maximum

size for the video slices. In [10] the slice size is studied as a function of the Content Type, the BLER and the achievable bitrate. From the resulting Packet Loss Ratio and Video Quality for three types of sequences, values below 300 bytes are proven to achieve better performances. Fixed size slices can be used as proposed for example in [11, 12]. In these studies authors propose to use fixed slice sizes of about 680 bytes as a good trade-off between slice error probability and packet overhead. Finally, in [13] the encoder is configured to generate video slices of 110 bytes as maximum size.

As previously discussed, the selection of the packet size becomes of great relevance. Using big-sized slices (and thus greater RTP payload sizes) decreases the header overhead and increases the efficiency of the available bitrate. However, bigger packets get a higher loss probability since the probability of errors in the radio channel increase with the packet size. On the other hand, the use of small packet sizes entails some drawbacks. On one hand, small slice sizes introduce higher overheads due to the RTP/UDP/IP headers. Thus, in limited resource environments, the actual bitrate available for video coding becomes lower resulting on lower media quality. On the other one, small slices may result on lower efficiency of the codification process, so more bitrate would be needed for the same visual quality.

Therefore, for the performance study proposed, we need to understand the combined effect of these encoding configuration parameters, jointly to the exhibited performance of the UTRAN, into the experienced Quality of Service (QoS). Therefore, during this paper we will focus on the study of the combined effect of different video slice sizes and the UTRAN transmission characteristics.

III. QOE IN MOBILE VIDEO STREAMING

In order to foresee the impact of the transmission effects, first of off we identify which the main sources of service degradation are in a scenario such as the proposed. The QoE for UMTS-based video services can be determined by two main factors:

1) *Interactivity*. This factor is related to the delay from the generation of the media at the source until it is finally played out. This aspect is especially relevant in conversational services, where the interactivity can be perceived as the total mouth-to-ear delay in a conversation. For streaming services, the interactivity becomes less relevant. However, it could be kept as a QoE requirement in certain cases, when a media service is composed by different media flows or to avoid annoying delays in the transmission of live events.

2) *Media Quality*. The visual quality is mainly determined by two factors.

a) *Encoding*. It is well-known that the configuration of the source encoding is a key factor in the perceived quality of the service, since it determines the maximum level for the perceived visual quality of the video sequence. Some basic parameters that affect the resulting video quality are the target bitrate (BR), the frame rate (FR), the spatial resolution (SR) or the quantization index (Q). When selecting the encoding

parameters, the combined effect of those parameters must be taken into account in order to maximize the QoE.

b) *Packet losses and packetization*. The packet losses in the transmission result on video frame (or slice) losses, and consequently on visual impairments. Even more, the impact of the same packet loss pattern on the visual quality may be perceived in a different way in function of the set of configured encoding values. All the possible combinations between those parameters may lead to a different experienced PQoS level, even in ideal network conditions. As a result, the most optimal combination of the encoding parameters should be determined for each particular case, analyzing the specific network conditions and the client characteristics.

Regarding the encoding parameters of H.264/AVC for mobile 3G video streaming communications, one of the most complete studies is [14], where authors provide a simple expression that allows estimating the visual quality based on the target frame rate and bitrate:

$$MOS_{CC} = A + B.BR + C/BR + D.FR + E/FR \quad (1)$$

This expression is based on experimental subjective tests with H.264/AVC Baseline Profile Level 1b in the ranges of frame rate of {5, 7.5, 10, 15 fps} and bitrates of {24, 50, 56, 60, 70, 80, 105 kbps}. Additionally, since the characteristics of the video stream have a great impact on the evaluation of the different configurations, different coefficients are provided for each of the different content classes authors analyze (news, soccer, cartoon, panorama, rest).

Finally, we must consider the possible effects of the RLC AM recovery functions. This operating mode provides the capability to recover RLC PDUs lost in the radio access, at the cost of adding an extra delay to the whole RLC SDU. Thus, an increase in the de jittering buffer size at the client could help to mitigate the effects of delay variation. However, we must also take into consideration the capacity requirements: mobile devices may have very limited storage resources which could make that few frames can be buffered.

IV. EFFECTS OF THE SLOTTED TRANSMISSION

In order to study the impact of the different encoding parameters, jointly with the UMTS transmission characteristics, we use the methodology illustrated in Figure 2. From a reference video clip, we generate different H.264 versions. Once the video clip is encoded, we reproduce the packet generation process in the OPNET Modeler® simulator, in order to analyze the impact of the UMTS access network into the specific service provision.

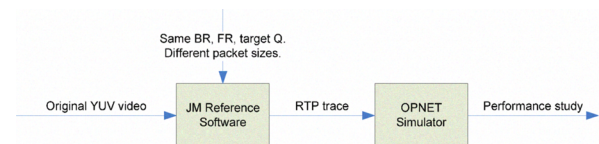


Figure 2. Methodology for video performance study

A. Generation of the RTP trace files

The selected original YUV file for the study is the ‘news’ video clip, at a QCIF resolution.¹ The H.264 encoded file is obtained using the H.264/AVC Joint Model Reference Software². The encoding parameters related to the video quality, which are presented in Table IV, are similar to those determined in [14]. Based on (1), we configure the output H.264 to a target bitrate of 105 kbps and a frame rate of 15 fps with a QCIF resolution. Thus, although the original file is sampled at 30 fps, we modify the video clip in order to reduce the frame rate to 15 fps, resulting on a total of 150 video frames for the whole video clip. With regard to the error resilience, the slice mode is set to consider a maximum size in bytes. For comparison purposes, two values have been tested: 650 bytes and 110 bytes.

Since the output is configured to RTP, the result of the encoding process is a .264 file that includes the sequence of RTP packets created by the *lencod* tool to transport all the media frames that make up the video sequence. Once the encoding process ends, the result cannot be played out with a conventional video player. In order to check the result of the codification, the JM Reference Software includes the *rtpdump* utility that provides information about the generated RTP packets. The output of the encoding process throws the following results: for the case of 650 bytes as maximum slice size, the RTP stream is made up of 302 RTP packets, resulting on a bitrate of 104.99 kbps at 15 Hz; for the 110 bytes case, we found a total of 1502 RTP packets, resulting on a total of 105.34 kbps at VCL.

TABLE IV. SELECTED ENCODING SET

Encoding Parameter	Value
Profile/Level IDC	(66,11) Baseline Profile, level 1.1.
Target BR	105 kbps
Frame Skip (resulting FR)	1 (15 fps)
SR	176x144
Sequence type	IPPP
Quantization parameter	(QP: I 28, P 28)
Entropy coding method	CAVLC
RD-optimized mode decision	Enabled
Data Partitioning Mode	1 partition
FMO	No FMO
Slice Mode	Fixed size, with 650 bytes and 110 bytes
Total number of reference frames	5
Output Type	RTP Packet File Format

¹ Available at <http://www.cipr.rpi.edu/resource/sequences/sequences/qcif/yuv/news.qcif.gz>

² Available for download at <http://iphome.hhi.de/suehring/tml/>

TABLE V. ENCODING RESULTS

Video Space	{ PSNR (dB), cSNR (dB), MSE }	
	650 bytes of max. slice size	110 bytes of max. slice size
Y	{ 40.85, 40.78, 5.43 }	{ 40.48, 40.41, 5.91 }
U	{ 42.91, 42.87, 3.36 }	{ 42.63, 42.59, 3.58 }
V	{ 43.33, 43.29, 3.05 }	{ 43.04, 43.00, 3.26 }

Table V shows the output of the encoding process for the two considered cases. As a first result, we can check the impact of the slice size on the codec efficiency. For the presented case study, the difference between encoding the same video clip at the same bitrate with different slice sizes is 0.37 dB in the PSNR and 0.48 in the MSE scale.

B. Simulation of video streaming over UMTS

In order to simulate the impact of the UTRAN transmission effects over the streamed video sequence, we make use of the UMTS model included in OPNET Modeler® in conjunction with some extensions implemented to support the delivery of the generated H.264 streams. The proposed network scenario, illustrated in Figure 3, allows the establishment of a basic H.264 video session between the mobile node (UE) and the Video Server over a Background UMTS service.

In order to support the transmission of H.264 video flows with different encoding configurations as required, a new application process model has been developed. This new application is based on two main pieces of code. The *gna_packet_generator_snd_mgr* process model, implemented as a child process of the general *gna_profile_mgr* process in the video server, provides the capability of reading from an RTP trace file and generating the corresponding packets toward the video destination.

The *gna_packet_generator_rcv_mgr* process model, implemented as a child process of the general *gna_clsvr_mgr* process in the mobile video client, is able to receive the video packets and collects each RTP packet reception time to a results file.

From the collected information from both the sending and receiving processes, we can construct the relationship between the video frames generation and playing functions. From the collected data, we can analyze the evolution of the required bitrate over time. Figure 4 and Figure 5 present the resulting bitrate at different levels

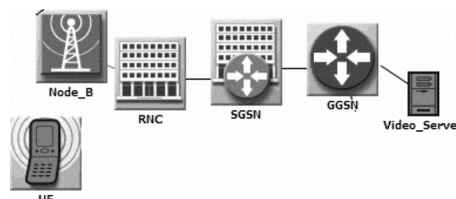


Figure 3. OPNET scenario

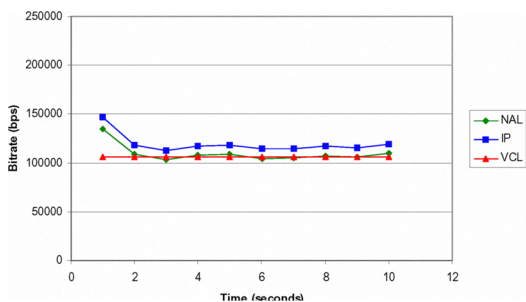


Figure 4. Bitrate at different levels over time @ 650 bytes max. per slice

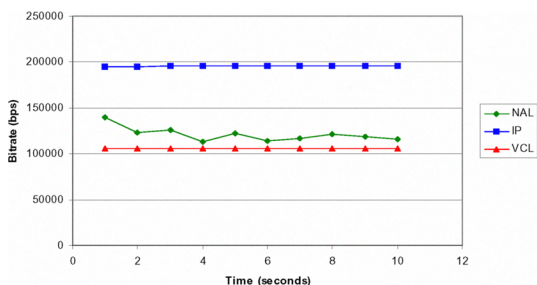


Figure 5. Bitrate at different levels over time @ 110 bytes max. per slice

The VCL trace represents the bitrate at the output of the H.264/AVC Video Coding Layer, which was preconfigured in the encoding parameters set to 105 kbps; the NAL trace represents the bitrate at the output of the Network Adaptation Layer, which includes the additional headers added to each individual video slice to make up the RTP-ready sequence of NALUs; the IP trace represents the bitrate required at the IP layer, which includes the extra RTP/UDP/IP header overhead.

As can be shown, the observed data confirms the hypothesis concerning the usage of I frames in limited resource environments. The transmission of the IDR frame (the first frame in the sequence) requires several RTP packets to be transmitted, namely up to 8 and 64 RTP packets respectively for the cases of maximum slice size of 650 bytes and 110 bytes.

The reception of the first video frame can be considered finalized when the RTP packet containing the last slice is received at the client. Thus, from a simple TTI inspection, the delay between the start of the stream and the reception of the whole IDR is 12 TTIs (120 ms) and 22 TTIs (220 ms) respectively for the cases of maximum slice size of 650 bytes and 110 bytes. Due to this effect, the play out of the video stream gets initially delayed because of the usage of I frames over a limited bitrate access such as the UTRAN. Moreover, the RTP packets corresponding to the subsequent frames have to be buffered in downlink path (in the RNC in this case) until the whole I frame is transmitted.

After the initial burst of data, the source bitrate gets inferior values to the configured available bitrate (384 kbps). Hence,

the media stream can make use of the extra bitrate to recover the synchrony and RTP packets can be transmitted without additional delays. For the two proposed cases, this effect happens at the 3rd and 5th video frames. However, even when the new video frames are received without additional delays, the playing function may be configured to keep the temporal resolution to the established frame rate. As a result, from that time on, we may consider an effect similar to a dejittering buffer, and the effects of the instantaneous jitter would be mitigated.

V. EFFECTS OF THE ERROR-PRONE TRANSMISSION

Following with the study of the performance of H.264 video streaming over UMTS networks, we focus now on the effect of the errors introduced by the radio channel. More specifically, the study aims at identifying the combined impact of the radio channel error pattern and the different encoding parameters into the PQoS.

Since the video transmission in this case is directed from the video server towards the mobile device, the interest falls in the characteristics and performance of the downlink UMTS DCH channel. From a review of the scientific literature, we can highlight the results of a recent study lead by Wolfgang Karner at the Vienna University of Technology. [15, 16] The authors characterize the behavior of the downlink direction of a UMTS DCH based on the measurement of live networks.

Therefore, in order to obtain accurate simulation results, as close as possible to the performance experienced in currently deployed UMTS networks, we have implemented the DL DCH error model as presented in [16] In this work, from the analysis of the number of erroneous TBs per TTI, authors conclude that the error pattern for a mobile user can be modeled at TTI level.

In order to study the impact of the packet size in the video transmission, we have developed a simulation toolset (Figure 6) that allows us to generate a video trace with an RTP packet loss that follows the UMTS DL DCH BLER pattern, and thus to predict the impact on the users' perceived quality.

As described in the previous section, we first generate an H.264 RTP trace from the original reference 'news' video clip, and taking into account the encoding data set depicted in Table IV. Then, after running the corresponding simulations, we obtain a data series where the RTP packet sequence and payload size are associated to the expected time of reception, and thus the corresponding TTI slot(s) used to transport each RTP packet.

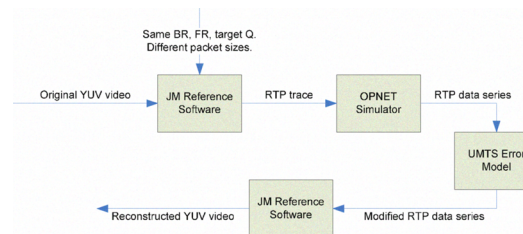


Figure 6. Methodology for video performance study (ii)

Once we obtain the corresponding data series, including the characteristics of each RTP packet and the TTIs used to transport them to the user, we make use of an ‘UMTS Error Model’ implemented in MATLAB. Figure 7 and Figure 8 show the CDF of the gap periods and bursts lengths of the artificially generated error traces. The gap length identifies a period of time where a number of consecutive TTIs are found without errors. As well, the burst length provides information about how many TTIs on a row can include bit errors.

As a result of this process, we obtain a modified version of the RTP data series, where the corrupted RTP packets are marked as lost. In summary, the process of identifying corrupted packets is as follows: first, each RTP packet is associated to the TTI(s) used to its transport; if any of those TTIs is marked as erroneous by the UMTS error model, the whole RTP packet is marked as lost, and the video information contained within that RTP packet could not be used in the decodification process.

Therefore, taking into account the new RTP data series, we make use of a modified version of the *rtp_loss* utility, which in its original version is included within the JM Reference SW. This new utility deletes the estimated lost RTP packets from the RTP trace file, taking as input a text file that includes the sequence numbers of the erroneous RTP packets obtained with the implemented MATLAB functions.

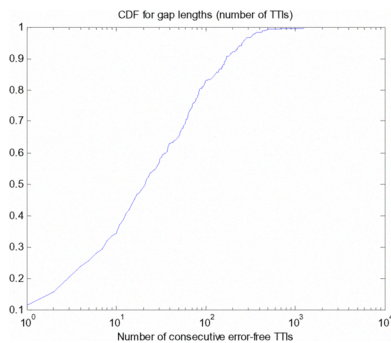


Figure 7. CDF of generated error-free gap lengths

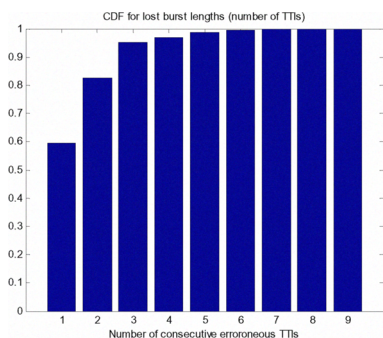


Figure 8. CDF of generated error burst lengths

TABLE VI. QUALITY OF THE DECODED VIDEO CLIPS

Video Space	SNR (dB)	
	650 bytes of max. slice size	110 bytes of max. slice size
Y	28.53	34.17
U	39.78	41.60
V	40.64	42.72

Finally, the YUV video file is regenerated with the *ldecod* utility, so the modified video clip can be played out. In order to cope with the RTP packet losses, the error concealment function of the decoder is configured to copy the precedent frame or slice. As well, in order to compare the video sequence estimated at the UMTS client with the source clip, we set up the *ldecod* utility to be able to access the original YUV file, and thus we obtain the estimated SNR values.

For comparative purposes, Table VI shows the results obtained for the same source video clip, encoded with the same H.264 parameters, but varying the maximum slice size from 650 bytes to 110 bytes.

From the results of the *rtp_loss* utility, we can identify the number of packet losses experienced in each case. In the case of 650 bytes of maximum slice size, we find 5 lost RTP packets from a total of 302 RTP packets in the video stream. In the case of 110 bytes, 31 of 1502 RTP packets get lost.

As can be observed, the video clip configured with maximum slice sizes of 110 bytes experiences a much bigger number of RTP packet losses. Even the RTP packet loss ratio is higher for the 110 bytes case: 1.33% vs. 2.06%. However, the SNR values obtained for the 110 bytes case are better than the 650 bytes case.



Figure 9. Impact of a lost slice for 110 bytes and 650 bytes of max. slice



Figure 10. Impairment propagation for 110 bytes and 650 bytes of max. slice

Figure 9 shows the impact of an RTP packet loss into the overall image quality. The slice structure used for the H.264 encoding makes that the visual impact of an RTP packet loss is more limited for the global quality of the picture for the case of lower slice sizes.

In addition to this individual effect, the composed effect over the sequence of pictures is even more significant. Figure 10 shows how the cumulative effects of RTP packet losses affect to the video sequence. The decoder is configured to copy the previous slice when an error occurs. Therefore, severe visual impairments are more likely propagated to the subsequent frames as the slice size gets increased.

VI. CONCLUSIONS

This paper focuses on the performance study of H.264-based video streaming services over typical mobile 3G UMTS accesses. The performance of this type of services becomes determined by the combined effects of the video encoding parameters and the inherent characteristics to the UTRAN transmission.

From a review of the performance parameters that affect the visual quality, we can state two main factors as the most relevant. On one hand, the source encoding set includes different parameters, such as the target bitrate, the frame rate, the spatial resolution, the quantization indexes or even the selected video slice size at the VCL. The combination of all this parameters, jointly with the content type itself, determines the upper threshold for the visual quality. On the other hand, the effect of frame losses determines the possible visual impairments due to the transmission over the mobile network.

At the same time, the frame losses pattern is determined by the combined effects of the UTRAN transmission, in terms of limited bitrate and error-prone medium, and other parameters that affect the transmission, such as the used packet sizes. However, we must take also into account the higher RTP/UDP/IP overhead due to low video packet sizes.

In addition, for the considered configuration, the packet size becomes determined by the video slice size. Therefore, the slice size has a twofold effect on the finally perceived visual quality, due to its impact in both the encoding and the transmission processes.

The results provided in this paper show that, for the considered mobile resolutions and encoding configurations, the impact of the maximum slice size is higher for the transmission process than for the encoding process. However, only two slice sizes have been considered for comparative purposes. As further work, we state the need for carrying out more simulations considering other intermediate slice sizes. As well, the effect of the content type or the specific spatial resolution should be evaluated.

Thus, the results here presented should be revised for other encoding configurations. Further simulations could extend the study, providing new results taking into account higher bitrates or new error concealment techniques, as proposed in [17].

REFERENCES

- [1] 3GPP TR 25.993 (2008-03): "Typical examples of Radio Access Bearers (RABs) and Radio Bearers (RBs) supported by Universal Terrestrial Radio Access (UTRA)".
- [2] 3GPP TS 26.235 (2008-03): "Packet switched conversational multimedia applications; Default codecs".
- [3] 3GPP TS 26.234 (2008-03): "Transparent end-to-end Packet-switched Streaming Service (PSS); Protocols and codecs".
- [4] 3GPP TS 26.346 (2008-03): "Multimedia Broadcast/Multicast Service (MBMS); Protocols and codecs".
- [5] ITU-T Recommendation H.264 (2003): "Advanced video coding for generic audiovisual services" | ISO/IEC 14496-10:2003: "Information technology – Coding of audio-visual objects – Part 10: Advanced Video Coding".
- [6] 3GPP TR 26.937 (2007-06): "Transparent end-to-end packet switched streaming service (PSS); Real-time Transport Protocol (RTP) usage model".
- [7] J. O. Fajardo, F. Liberal, N. Bilbao, "Study of the impact of UMTS Best Effort parameters on QoE of VoIP services", The Fifth International Conference on Autonomic and Autonomous Systems (ICAS 2009), in press.
- [8] X.-J. Zhang, X.-H. Peng, R. Haywood and T. Porter, "Robust Video Transmission over Lossy Network by Exploiting H.264/AVC Data Partitioning", in Proc. Fifth International Conference on Broadband Communications, Networks, and Systems, 2008.
- [9] L. Liu, S. Zhang, X. Ye, Y. Zhang, "Error Resilience Schemes of H.264/AVC for 3G Conversational Video Services," in Proc. of the Fifth International Conference on Computer and Information Technology (CIT'05), pp. 657 – 661, Sept. 2005.
- [10] A. T. Connie, P. Nasiopoulos, V. C. M. Leung and Y. P. Fallah, "Video Packetization Techniques for Enhancing H.264 Video Transmission over 3G Networks", in Proc. 5th IEEE Consumer Communications and Networking Conf. (CCNC) 2008.
- [11] W. Karner, O. Nemethova, P. Svoboda, M. Rupp, "Link Error Prediction Based Cross-Layer Scheduling for Video Streaming over UMTS", in Proc. 15th IST Mobile and Wireless Summit, Myconos, Greece, June, 2006.
- [12] L. Superiori, O. Nemethova, W. Karner, M. Rupp, "Cross-Layer Detection of Visual Impairments in H.264/AVC Video Sequences streamed over UMTS Networks", in Proc. IEEE 1st International Workshop on Cross Layer Design, Jinan, Shandong, China, September 2007.
- [13] S. Benayoune, N. Achir, K. Boussetta and K. Chen, "Content-aware ARQ for H.264 Streaming in UTRAN", in Proc. IEEE Wireless Communications and Networking Conference, 2008 (WCNC 2008), pp. 1956-1961, March 2008
- [14] M. Ries, O. Nemethova and M. Rupp, "Video Quality Estimation for Mobile H.264/AVC Video Streaming", Journal of Communications, vol. 3, no. 1, January 2008
- [15] W. Karner, P. Svoboda, M. Rupp, "A UMTS DL DCH Error Model Based on Measurements in Live Networks"; in Proc. 12th International Conference on Telecommunications (ICT), Capetown, South Africa, May, 2005.
- [16] W. Karner, O. Nemethova, P. Svoboda, and M. Rupp, "Link Error Analysis and Modeling for Video Streaming Cross-Layer Design in Mobile Communication Networks", ETRI Journal, vol. 29, no 5, October 2007.
- [17] T. Thaipanich, P.-H. Wu, C.-C. Jay Kuo, "Low-complexity Video Error Concealment For Mobile Applications Using OBMA," IEEE Transactions on Consumer Electronics, Vol. 54, Issue 2, May 2008, pp.753-761.